CERN Yellow Reports: School Proceedings
Volume 1/2018

CERN-2018-001-SP

The CERN Accelerator School

Free Electron Lasers and Energy Recovery Linacs

Hamburg, Germany
31 May– 10 June 2016

Editor: R. Bailey

CERN

CERN Yellow Reports: School Proceedings
Published by CERN, CH-1211 Geneva 23, Switzerland

ISBN 978–92–9083–482–3 (paperback)
ISBN 978–92–9083–483–0 (PDF)
ISSN 2519-8041 (Print)
ISSN 2519-805X (Online)
DOI https://doi.org/10.23730/CYRSP-2018-001

Accepted for publication by the CERN Report Editorial Board (CREB) on 6 April 2018

Available online at https://e-publishing.cern.ch/ and https://cds.cern.ch/

Copyright © CERN, 2018, unless otherwise specified

Creative Commons Attribution 4.0

Knowledge transfer is an integral part of CERN's mission.
CERN publishes this volume Open Access under the Creative Commons Attribution 4.0 license (http://creativecommons.org/licenses/by/4.0/) in order to permit its wide dissemination and use. The submission of a contribution to a CERN Yellow Reports series shall be deemed to constitute the contributor's agreement to this copyright and license statement. Contributors are requested to obtain any clearances that may be necessary for this purpose.

This volume is indexed in: CERN Document Server (CDS), INSPIRE, Scopus.

This report should be cited as:

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018). https://doi.org/10.23730/CYRSP-2018-001

A contribution in this volume should be cited as:

[Author name(s)], in Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018), pp. [first page] – [lastpage], https://doi.org/10.23730/CYRSP-2018-001. [first page]

Abstract

These proceedings collate lectures given at the course on Free Electron Lasers and Energy Recovery Linacs (FELs and ERLs), organised by the CERN Accelerator School (CAS). The course was held at the Hotel Scandic Hamburg Emporio, Hamburg, Germany from 31 May to 10 June 2016, in collaboration with DESY. Following introductory lectures on radiation issues, the basic requirements on linear accelerators and ERLs are discussed. Undulators and the process of seeding and lasing are then treated in some detail, followed by lectures on various beam dynamics and controls issues.

Preface

The aim of the CERN Accelerator School (CAS) is to collect, preserve and disseminate the knowledge accumulated in the world's accelerator laboratories over the years. This applies not only to general accelerator physics, but also to related sub-systems and associated technologies, and to the development of novel, dedicated facilities. These wider aims are achieved by means of specialized courses currently held twice per year. The topic of the first 2016 specialized course was Free Electron Lasers and Energy Recovery Linacs (FELs and ERLs) and was held at the Hotel Scandic Hamburg Emporio, Hamburg, Germany from 31 May to 10 June 2016.

The course was made possible through the fruitful collaboration with DESY, in particular through the efforts of Ruth Mundt, Christel Oevermann and Kay Wittenburg.

A full day visit to DESY and the European XFEL in Hamburg Bahrenfeld provided a practical insight into the field. Participants also had the opportunity to work on realistic case studies as an integral part of the programme. For the organisation and execution of the latter we are indebted to Sven Reiche from PSI. The backing of the CERN management and the guidance of the CAS Advisory and Programme Committees enabled the course to take place, while the attention to detail of the Local Organising Committee and the management and staff of the Hotel Scandic Hamburg Emporio ensured that the school was held under optimum conditions.

Special thanks must go to the lecturers for the preparation and presentation of the lectures, even more so to those who have written a manuscript for these proceedings.

For the production of the proceedings we are indebted to the efforts of Barbara Strasser and to the CERN Publishing Service, especially Valeria Brancolini for her very positive and efficient collaboration.

These proceedings have been published in paper (black and white) and electronic form. The electronic version, with full colour figures, can be found at https://e-publishing.cern.ch/index.php/CYRSP/issue/view/47.

Roger Bailey, Editor
CERN Accelerator School

PROGRAMME

Free Electron Lasers and Energy Recovery Linacs (FELs and ERLs), 31 May – 10 June, 2016

Time	Tuesday 31 May	Wednesday 1 June	Thursday 2 June	Friday 3 June	Saturday 4 June	Sunday 5 June	Monday 6 June	Tuesday 7 June	Wednesday 8 June	Thursday 9 June	Friday 10 June
08:30 09:30		Opening Talks	Wakefields U. Van Rienen	Lasers in FEL Facilities M. Divall	Concept of ERL A. Jankowiak		Quantum FEL G. Robb	LLRF Controls and Feedbacks S. Pfeiffer	Advanced Undulator Concepts J. Rosenzweig	Undulator Tapering M. Yurkov	
09:30 10:30	**A**	Electro-magnetic Theory W. Herr	Linear Beam Optics I B. Holzer	Motion in an Undulator S. Reiche	High Gain Regime 1D I K.-J. Kim	**E** **X**	High Gain Regime 3D K.-J. Kim	Coherence Properties of the Radiation from SASE FEL M. Yurkov	Photon Beam Transport M. Yabashi	Crystallography and Molecular Imaging using X-Ray Lasers T. White	**D** **E**
		COFFEE	**COFFEE**	**COFFEE**	**COFFEE**		**COFFEE**	**COFFEE**	**COFFEE**	**COFFEE**	**P**
11:00 12:00	**R** **R** **I**	Special Relativity W. Herr	Linear Beam Optics II B. Holzer	Pendulum Equations and Low Gain Regime S. Reiche	High Gain Regime 1D II K.-J. Kim	**C** **U**	Temporal Coherence K.-J. Kim	The European XFEL H. Weise	Electron Beam Diagnostics R. Ischebeck	Machine Protection L. Froehlich	**A** **R**
12:00 13:00	**V** **A**	Synchrotron Radiation I L. Rivkin	Coherence in Beams J. Rossbach	Bunch Length Compressors S. Di Mitri	Seeding Schemes I L. Giannessi	**R** **S**	Experience from FLASH: FEL Theory versus Experiment M. Yurkov	Plasma Wake Driven FEL R. Assmann	Timing and Synchronisation M. Bellaveglia	Different ERL Applications K.Aulenbacher	**T** **U** **R**
	L	**LUNCH**	**LUNCH**	**LUNCH**	**LUNCH**		**LUNCH**	**LUNCH**	**LUNCH**	**LUNCH**	
14:30 15:30	**D**	Synchrotron Radiation II L. Rivkin	Space Charge Mitigation M. Ferrario	Coherent Synchrotron Radiation and Microbunching Instability S. Di Mitri	Seeding Schemes II L. Giannessi	**I** **O**	XFELO K.-J. Kim	**V** **I** **S** **I**	Case Study Work	Case Study Presentations	**E**
15:30 16:30	**A** **Y**	Undulators J. Pflueger	SC Cavities (High Q) J. Knobloch	Electron Sources and Injector Systems E. Chiadroni	Beam Dynamics of ERLs A. Jankowiak	**N**	Case Study Work	**T** **T** **O**	Case Study Work	Case Study Presentations	**D** **A**
		TEA	**TEA**	**TEA**	**TEA**		**TEA**	**D**	**TEA**	**TEA**	
17:00 18:00	**Registra tion**	Linear Accelerator Technology D. Alesini	Historical Survey of FELs M. Couprie	Energy Efficiency E. Jensen	Transverse Optics in the ERL Arcs A. Valloni		Case Study Work	**E** **S** **Y**	Case Study Work	**Seminar** Gravitational Waves R. Schnabel	**Y**
18:30		Case Study Introduction									
19:30		**Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	

Contents

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Classical Electrodynamics and Applications to Particle Accelerators

W. Herr
CERN, Geneva, Switzerland

Abstract
Classical electrodynamic theory is required theoretical training for physicists and engineers working with particle accelerators. Basic and empirical phenomena are reviewed and lead to Maxwell's equations, which form the framework for any calculations involving electromagnetic fields. Some necessary mathematical background is included in the appendices so the reader can follow the work and the conventions used in this text. Plane waves in vacuum and in different media, radio frequency cavities, and propagation in a waveguide are presented.

Keywords
Electrodynamics; Maxwell's equations; electromagnetic waves; cavities; polarization.

1 Introduction and motivation

Together with classical mechanics, quantum theory, and thermodynamics, the theory of classical electrodynamics forms the framework for the introduction to theoretical physics. Classical electrodynamics can be applied where the length scale does not required a treatment on the quantum level. Although both electricity and magnetism can exert forces on other objects, they were for a long time treated as distinct effects. The empirical laws were unified in a single theory by Maxwell and culminated in the prediction of electromagnetic waves. Although very successful in describing most phenomena, it is not possible to reconcile the theory with the concepts of classical mechanics. This was solved by the introduction of special relativity by Einstein, in studying the effects of moving charges. This reformulation not only explained the origin of such effects as the Lorentz force, but also showed that electricity and magnetism are two different aspects of the same underlying physics. Since in accelerator physics we are mainly concerned about moving charges, the topic of special relativity is treated in a separate lecture at this school [1].

This paper touches on many different areas of electromagnetic theory, with a strong focus on applications to accelerator physics [2]. It covers the field of electrostatics and the equations of Gauss and Poisson, magnetic fields generated by linear and circular currents, and electromagnetic effects in vacuum and different media, and leads to Maxwell's equations [3–5].

Electromagnetic waves and their behaviour at boundaries and in waveguides and cavity resonators are treated in some detail. Because of their importance, such phenomena as polarization and propagation in perfect and resistive conductors are presented.

The paper is intended as a recapitulation for physicists and engineers and mathematical subtleties are avoided where it is acceptable.

This paper cannot replace a full course on electromagnetic theory. This is, in particular, true for students less familiar with this subject. Although they will not be able to understand everything in this lecture, it is attempted to provide access to the core material and the direct features relevant for accelerator physics.

The background required is a knowledge of calculus and differential equations; some more advanced concepts, such as vector calculus are summarized in the appendices.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.1

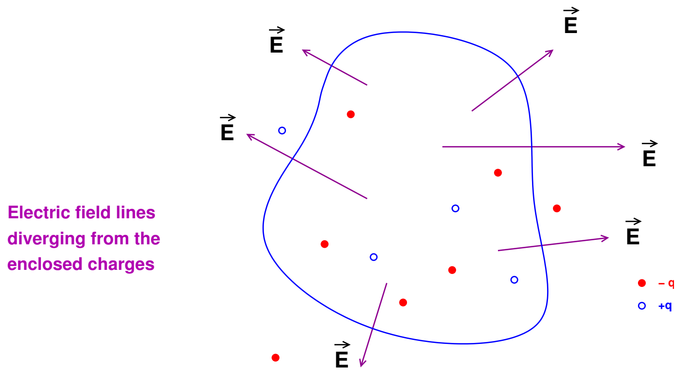


Fig. 1: Charges enclosed within a closed surface

2 Electrostatics

Electrostatics deals with phenomena related to time-independent charges. It was found empirically that charged bodies exert a force on each other, attracting in the case of unlike charges or repelling for charges of equal sign. This is described by the introduction of electric fields and the Coulomb force acting on the particles. Charges are the origin of electric fields, which form a vector field.

2.1 Gauss's theorem

The fields of a distribution of charges add to form the overall field and the latter can be computed when the distribution of charge is known. This treatment is based on the mathematical framework worked out by Gauss and others and is summarized in Gauss's theorem. Gauss's theorem in its simplest form is illustrated in Fig. 1.

We assume a surface S enclosing a volume V, within which are charges: $q_1, q \ldots$, producing electromagnetic fields $\vec{E}$ originating from the charges and passing through the surface (Fig. 1).

Summing the normal component of the fields passing through the surface, we obtain the flux ϕ:

$$\phi = \int_S \vec{E} \cdot \vec{n} \mathrm{d}A = \sum_i \frac{q_i}{\epsilon_0} = \frac{Q}{\epsilon_0}\,, \tag{1}$$

where $\vec{n}$ is the normal unit vector and $\vec{E}$ the electric field at an area element $\mathrm{d}A$ of the surface. The surface integral of $\vec{E}$ equals the total charges Q inside the enclosed volume.

This holds for *any* arbitrary (closed) surface S and is:

- independent of *how* the particles are distributed inside the volume;
- independent of *whether* the particles are moving or at rest;
- independent of *whether* the particles are in vacuum or material.

Using Gauss's formula (see Appendix B), we can formulate the theorem as:

$$\begin{aligned} &\textstyle\int_S \vec{E} \cdot \mathrm{d}\vec{A} = \int_V \frac{\rho}{\epsilon_0} \cdot \mathrm{d}V = \frac{Q}{\epsilon_0} = \phi_E \\ &\underbrace{\int_S \vec{E} \cdot \mathrm{d}\vec{A} = \int_V \nabla\vec{E} \cdot \mathrm{d}V = \int_V \mathsf{div}\vec{E} \cdot \mathrm{d}V}_{\text{Gauss's formula}} \qquad \text{(relates surface and volume integrals)} \end{aligned} \tag{2}$$

It follows that from Eq. (2):

$$\mathsf{div}\ \vec{E} = \nabla \cdot \vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho}{\epsilon_0}\,, \tag{3}$$

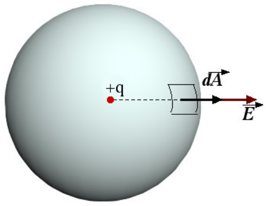


Fig. 2: Flux through a surface element created by a single point charge

which is Maxwell's *first* equation.

As a physical picture, the divergence 'measures' outward flux ϕ_E of the field. The simplest possible example is the flux from a single charge q, shown in Fig. 2. A charge q generates a field $\vec{E}$ according to (Coulombs law):

$$\vec{E} = \frac{q}{4\pi\epsilon_0}\frac{\vec{r}}{r^3}\,. \tag{4}$$

It is enclosed by a sphere and obviously $\vec{E} = \mathrm{const.}$ on a sphere (area, $4\pi \cdot r^2$):

$$\int\int_{\text{sphere}} \vec{E}\cdot \mathrm{d}\vec{A} = \frac{q}{4\pi\epsilon_0}\int\int_{\text{sphere}}\frac{\mathrm{d}A}{r^2} = \frac{q}{\epsilon_0}\,. \tag{5}$$

The surface integral through the sphere A equals the charge inside the sphere (for any radius of the sphere), consistent with Eq. (1).

2.2 Electrostatic potential and Poisson's equation

We can derive the field $\vec{E}$ from a scalar electrostatic potential $\phi(x, y, z)$, i.e.,

$$\vec{E} = -\mathsf{grad}\ \phi = -\nabla\phi = -\left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right)\,, \tag{6}$$

then we have

$$\nabla\vec{E} = -\nabla^2\phi = -\left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}\right) = \frac{\rho(x, y, z)}{\epsilon_0}$$

This is *Poisson's equation.*

Once we can compute ϕ for a given distribution of the charge density ρ, we can derive the fields. As an example, the simplest possible charge distribution is an isolated point charge with the potential:

$$\phi(r) = \frac{q}{4\pi\epsilon_0 r}$$
$$\vec{E} = -\nabla\phi(r) = \frac{q}{4\pi\epsilon_0}\cdot\frac{\vec{r}}{r^3}$$

As a realistic case, we assume a distribution $\rho(x, y, z)$ that is Gaussian in all three dimensions:

$$\rho(x, y, z) = \frac{Q}{\sigma_x\sigma_y\sigma_z\sqrt{2\pi^3}}\exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right)$$

($\sigma_x, \sigma_y, \sigma_z$ are the r.m.s. sizes).

The potential $\phi(x, y, z, \sigma_x, \sigma_y, \sigma_z)$ becomes (see e.g., Ref. [6]):

$$\phi(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{Q}{4\pi\epsilon_0}\int_0^\infty \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+t} - \frac{y^2}{2\sigma_y^2+t} - \frac{z^2}{2\sigma_z^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)(2\sigma_z^2+t)}}\mathrm{d}t\,. \tag{7}$$

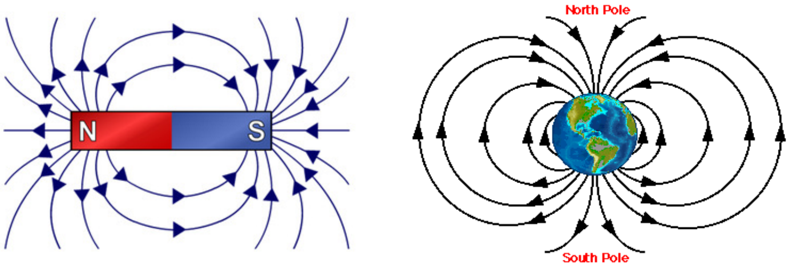


Fig. 3: Field lines between magnetic dipoles

In many realistic cases, the charge distribution shows a strong symmetry, Then we can rewrite the Poisson equation and obtain some very important formulae in practice.

Poisson's equation in polar co-ordinates (r, φ):

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \varphi^2} = -\frac{\rho}{\epsilon_0}\,; \tag{8}$$

Poisson's equation in cylindrical co-ordinates (r, φ, z):

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \varphi^2} + \frac{\partial^2 \phi}{\partial z^2} = -\frac{\rho}{\epsilon_0}\,; \tag{9}$$

Poisson's equation in spherical co-ordinates (r, θ, φ):

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \phi}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial \phi}{\partial \theta}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial^2 \phi}{\partial \varphi^2} = -\frac{\rho}{\epsilon_0}\,. \tag{10}$$

Examples for solutions of these equations are found in Ref. [3].

3 Magnetostatics

In the treatment of magnetostatic phenomena, we follow the strategy developed for electrostatics. The striking difference is the absence of magnetic charges, i.e., magnetic 'charges' occur only in combination with opposite 'charges', i.e., in the form of a magnetic dipole.

The field lines between magnetic poles for a magnet and the Earth's magnetic field are shown in Fig. 3.

We start with some basic definitions and properties.

- Magnetic field lines always run from north to south.
- They are described as vector fields by the magnetic flux density $\vec{B}$.
- All field lines are closed lines from the north to the south pole.

3.1 Gauss's theorem

We follow the same procedure as for electrostatic charges and enclose a magnetic dipole within a closed surface Fig. 4.

From very simple considerations, it is rather obvious that field lines passing outwards through the surface also return through this surface, i.e., the overall flux is zero. This is formally described by Gauss's second theorem, for magnetic fields:

$$\int_S \vec{B}\mathrm{d}\vec{A} = \int_V \nabla\vec{B}\mathrm{d}V = 0\,. \tag{11}$$

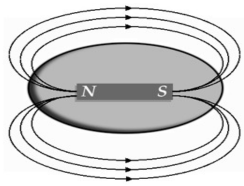


Fig. 4: Closed surface around magnetic dipole

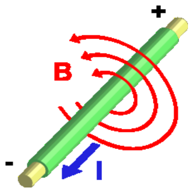


Fig. 5: Static electric current inducing an encircling (curling) magnetic field

This leads to Maxwell's second equation:

$$\nabla \vec{B} = 0 \,. \tag{12}$$

The physical significance of this equation is that magnetic charges (monopoles) do not exist (although Maxwell's equations could easily be modified if necessary).

3.2 Ampère's law

Static currents produce a magnetic field described by Ampère's law (Fig. 5).

Assuming a current density $\vec{j}$, we can compute the magnetic field:

$$\mathrm{curl}\vec{B} = \nabla \times \vec{B} = \mu_0 \vec{j} \,, \tag{13}$$

or in integral form, where the current density becomes the current I,

$$\int\int_A \nabla \times \vec{B} \mathrm{d}\vec{A} = \int\int_A \mu_0 \vec{j} \mathrm{d}\vec{A} = \mu_0 \vec{I} \,. \tag{14}$$

For a *static electric current* I in a *single wire* (Fig. 6), we get the Biot–Savart law (we have used Stoke's theorem and the area of a circle, $A = r^2 \cdot \pi$):

$$\begin{aligned} \vec{B} &= \frac{\mu_0}{4\pi} \oint \vec{I} \cdot \frac{\vec{r} \times \mathrm{d}\vec{s}}{r^3} \\ \vec{B} &= \frac{\mu_0}{2\pi} \frac{\vec{I}}{r} \end{aligned} \tag{15}$$

4 Time-varying electromagnetic fields

Extending the subject of static electric and magnetic fields opens a large range of new phenomena. Furthermore it shows a close connection between electricity and magnetism.

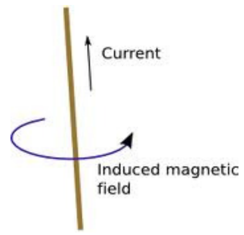


Fig. 6: Induced magnetic fields by static current

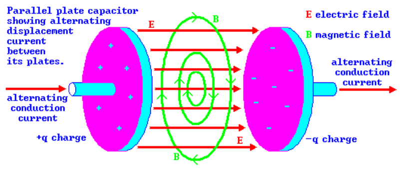


Fig. 7: Maxwell's displacement current, e.g., charging capacitor

4.1 Maxwell and time-varying electric fields

We need to address the question of whether we need an electric current to produce magnetic fields. This was addressed by Maxwell, which led him to the introduction of the displacement current $\vec{j}_\mathrm{d}$.

We define this displacement current by:

$$\vec{I}_d = \frac{\mathrm{d}q}{\mathrm{d}t} = \epsilon_0 \cdot \frac{\mathrm{d}\phi}{\mathrm{d}t} = \epsilon_0 \frac{\mathrm{d}}{\mathrm{d}t} \int\int_\mathrm{area} \vec{E} \cdot \mathrm{d}\vec{A} \,. \tag{16}$$

It must be understood that this is not a current from *moving charges* but from *time-varying electric fields*.

The displacement current I_d produces magnetic fields, just like 'actual currents' do. An example for a displacement current is a charging capacitor (Fig. 7).

Time-varying electric fields induce magnetic fields (using the current density $\vec{j}_\mathrm{d}$). We can formulate this as:

$$\nabla \times \vec{B} = \mu_0 \vec{j}_d = \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \,. \tag{17}$$

The bottom line of this result is that magnetic fields $\vec{B}$ can be generated in two ways:

$$\nabla \times \vec{B} = \mu_0 \vec{j} \tag{18}$$

are the magnetic fields produced by an electric current (Ampère), while

$$\nabla \times \vec{B} = \mu_0 \vec{j}_\mathrm{d} = \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \tag{19}$$

are the magnetic fields produced by a changing electric field (Maxwell).

Putting them together we obtain Maxwell's third law:

$$\nabla \times \vec{B} = \mu_0 (\vec{j} + \vec{j}_\mathrm{d}) = \mu_0 \vec{j} + \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \,. \tag{20}$$

Using Stoke's formula, this can be rewritten in integral form:

$$\underbrace{\oint_C \vec{B} \cdot \mathrm{d}\vec{s} = \int_A \nabla \times \vec{B} \cdot \mathrm{d}\vec{A}}_{\text{Stoke's formula}} = \int_A \left(\mu_0 \vec{j} + \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \right) \cdot \mathrm{d}\vec{A} \,. \tag{21}$$

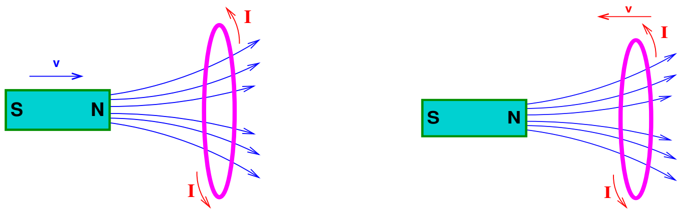


Fig. 8: Electromotive force (EMF) produced by changing magnetic flux Ω

4.2 Faraday's law and varying magnetic fields

Assuming a conducting coil in a static magnetic field $\vec{B}$ (Fig. 8). The area enclosed by the coil should be A. Changing the magnetic flux Ω through the area A produces an electromotive force (EMF) in the coil resulting in a current I:

$$\text{flux} = \Omega = \int_A \vec{B}\mathrm{d}\vec{A}, \qquad \text{EMF} = \oint_C \vec{E}\cdot\mathrm{d}\vec{s}, \tag{22}$$

$$-\frac{\partial\Omega}{\partial t} = -\frac{\partial}{\partial t}\underbrace{\int_A \vec{B}\mathrm{d}\vec{A}}_{\text{flux}\Omega} = \oint_C \vec{E}\cdot\mathrm{d}\vec{s}, \tag{23}$$

$$-\frac{\partial\Omega}{\partial t} = -\int_A \frac{\partial}{\partial t}\vec{B}\mathrm{d}\vec{A} = \oint_C \vec{E}\cdot\mathrm{d}\vec{s}. \tag{24}$$

The magnetic flux can be changed by:

- moving the magnet relative to the conducting coil;
- moving the coil relative to the magnet.

4.3 Ampère and Maxwell's law

In a more general form, this can be written using Stoke's formula, which relates line integrals and surface integrals. It is then rewritten as:

$$-\int_A \frac{\partial\vec{B}}{\partial t}\mathrm{d}\vec{A} = \underbrace{\int_A \nabla\times\vec{E}\mathrm{d}\vec{A} = \oint_C \vec{E}\cdot\mathrm{d}\vec{s},}_{\text{Stoke's formula}} \tag{25}$$

and we arrive at the well-known formulation:

$$\nabla\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}. \tag{26}$$

A changing magnetic field through *any* closed area induces electric fields in the (arbitrary) boundary. A sketch demonstrating Stoke's formula is shown in Fig. 9. This formulation is known as the Maxwell–Faraday law.

5 Maxwell's equations

The empirical concepts and experimental findings can be put together in a set of differential equations, usually referred to as Maxwell's equations.

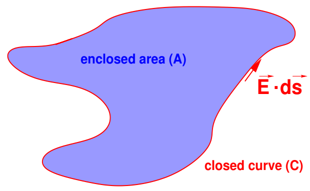


Fig. 9: Stoke's formula

5.1 Maxwell's equations in vacuum

Putting together Eqs. (3), (12), (20), and (26), Maxwell's equations in vacuum (so-called microscopic equations) read:

$$
\begin{aligned}
\nabla \vec{E} &= \frac{\rho}{\epsilon_0} = -\Delta\phi\,, && \text{(I)}\\
\nabla \vec{B} &= 0\,, && \text{(II)}\\
\nabla \times \vec{E} &= -\frac{\mathrm{d}\vec{B}}{\mathrm{d}t}\,, && \text{(III)}\\
\nabla \times \vec{B} &= \mu_0 \left(\vec{j} + \epsilon_0 \frac{\mathrm{d}\vec{E}}{\mathrm{d}t}\right)\,, && \text{(IV)}
\end{aligned} \tag{27}
$$

or, written in integral form (using Gauss's and Stoke's theorems):

$$
\begin{aligned}
\int_A \vec{E}\cdot \mathrm{d}\vec{A} &= \frac{Q}{\epsilon_0}\,,\\
\int_A \vec{B}\cdot \mathrm{d}\vec{A} &= 0\,,\\
\oint_C \vec{E}\cdot \mathrm{d}\vec{s} &= -\int_A \left(\frac{\mathrm{d}\vec{B}}{\mathrm{d}t}\right)\cdot \mathrm{d}\vec{A}\,,\\
\oint_C \vec{B}\cdot \mathrm{d}\vec{s} &= \mu_0 \int_A \left(\vec{j} + \epsilon_0 \frac{\mathrm{d}\vec{E}}{\mathrm{d}t}\right)\cdot \mathrm{d}\vec{A}\,.
\end{aligned} \tag{28}
$$

5.2 Maxwell's equations in material

In material, we have to modify the electromagnetic fields $\vec{E}$ and $\vec{H}$ and relate those to the magnetic induction $\vec{B}$ and electric displacement $\vec{D}$. In vacuum, we had:

$$\vec{D} = \epsilon_0 \cdot \vec{E}, \qquad \vec{B} = \mu_0 \cdot \vec{H}\,. \tag{29}$$

In a material, the relations read:

$$\vec{D} = \epsilon_{\mathrm{r}} \cdot \epsilon_0 \cdot \vec{E} = \epsilon_0 \vec{E} + \vec{P}\,, \tag{30}$$

$$\vec{B} = \mu_{\mathrm{r}} \cdot \mu_0 \cdot \vec{H} = \mu_0 \vec{H} + \vec{M}\,. \tag{31}$$

The origin of these additional contributions are $\vec{P}$olarization and $\vec{M}$agnetization in material.

We can summarize:

$$\epsilon_{\mathrm{r}}(\vec{E}, \vec{r}, \omega) \Rightarrow \epsilon_{\mathrm{r}} \text{ is } \textit{relative} \text{ permittivity} \approx [1\text{–}10^5]\,;$$

$$\mu_{\rm r}(\vec{H}, \vec{r}, \omega) \Rightarrow \mu_{\rm r} \text{ is } \textit{relative} \text{ permeability} \approx [0\text{–}10^6]\,.$$

If $\vec{D}$ and $\vec{B}$ do not depend on the fields $\vec{E}$ and $\vec{H}$, they are linear; if they do not depend on the direction ($\vec{r}$) or frequency (ω), they are isotropic and non-dispersive.

The so-called macroscopic Maxwell's equations become:

$$\begin{aligned}
\nabla\vec{D} &= \rho\,,\\
\nabla\vec{B} &= 0\,,\\
\nabla\times\vec{E} &= -\frac{\mathrm{d}\vec{B}}{\mathrm{d}t}\,,\\
\nabla\times\vec{H} &= \vec{j} + \frac{\mathrm{d}\vec{D}}{\mathrm{d}t}\,.
\end{aligned} \tag{32}$$

6 Electromagnetic potentials

It was shown that electric fields can be derived from a scalar potential ϕ:

$$\vec{E} = -\vec{\nabla}\phi\,. \tag{33}$$

Since div $\vec{B} = 0$, we can write $\vec{B}$ using a vector potential $\vec{A}$:

$$\vec{B} = \vec{\nabla}\times\vec{A} = \mathsf{curl}\vec{A}\,, \tag{34}$$

combining Maxwell (I) and Maxwell (III):

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t}\,. \tag{35}$$

Fields can be written as derivatives of scalar and vector potentials $\phi(x, y, z)$ and $\vec{A}(x, y, z)$. Knowledge of the potentials allows computation of the fields.

6.1 Gauge invariance

The equations for the potentials can be directly derived from Maxwell's equations:

$$\Delta\phi = \frac{1}{c}\frac{\partial(\nabla\cdot\vec{A})}{\partial t} = -4\pi\rho\,, \tag{36}$$

and

$$\Delta\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial^2 t} - \nabla\left(\nabla\cdot\vec{A} + \frac{1}{c}\frac{\partial\phi}{\partial t}\right) = \frac{4\pi}{c}\vec{j}\,. \tag{37}$$

We have two coupled differential equations for the potentials, which may be difficult to solve for general charge densities and current densities. We shall try to decouple these equations using a particular property of the potentials. While the absolute values of the electric and magnetic fields can be measured, the absolute values of the potentials are not defined. The electromagnetic potentials are merely auxiliary 'constructions', although very important ones, in particular, for the relativistic formulation of the electromagnetic theory.

Without going into the details, the theory should be invariant under a change of scale ('gauge'). The most commonly used is the Lorentz gauge, which yields a condition between the potentials:

$$\vec{A}_{\rm g} = \vec{A} + \nabla f\,, \tag{38}$$

$$\phi_{\rm g} = \phi + \frac{1}{c}\frac{\partial f}{\partial t}\,, \tag{39}$$

$$\frac{1}{c}\frac{\partial \phi_{\rm g}}{\partial t} + \nabla \vec{A}_{\rm g} = 0\,, \tag{40}$$

where f is an arbitrary function of position and time. These equations lead to the same (measurable) fields and do therefore satisfy Maxwell's equations. This 'gauge' transformation decouples Eq. (36) and Eq. (37) and leads to:

$$\Delta\phi(\vec{r},t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi(\vec{r},t) = -4\pi\cdot\rho(\vec{r},t)\,, \tag{41}$$

$$\Delta\vec{A}(\vec{r},t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\vec{A}(\vec{r},t) = -\frac{4\pi}{c}\cdot\vec{j}(\vec{r},t)\,. \tag{42}$$

We observe two consequences: first, the equations for the potentials are decoupled and depend only on the charge density and current density. Second, without charges or current, the equations have the form of a wave equation. The relevance becomes clear later, in particular, when Maxwell's equations are written in a relativistically invariant form [1].

Another very useful gauge is the Coulomb gauge:

$$\nabla\cdot\vec{A} = 0\,. \tag{43}$$

This leads us to a particularly simple expression for the electric potential:

$$\Delta\phi(\vec{r},t) = -4\pi\rho(\vec{r},t)\,. \tag{44}$$

The name 'Coulomb gauge' becomes obvious.

A formal solution can now be written as:

$$\phi = \int \frac{\rho(\vec{r'},t)}{\mid\vec{r}-\vec{r'}\mid}{\rm d}V\,. \tag{45}$$

6.2 Example: Coulomb potential

Equation (45) can immediately be applied to compute the Coulomb potential of a static charge q:

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0}\cdot\frac{q}{|\vec{r}-\vec{r}_q|}\,, \tag{46}$$

where $\vec{r}$ is the observation point and $\vec{r}_q$ the location of the charge.

7 Powering and self-induction

There are also induction effects in a single coil. A varying current (e.g., in a transformer or power line) produces a varying magnetic field inside itself and the flux of this field is continually changing, leading to a self-induced electromotive force (Fig.10). This electromotive force (EMF) is acting on any current when it is building up a magnetic field or when the field is changing in any way. This effect is called self-inductance. According to Lenz's rule, this EMF is opposing any flux change. The direction of an induced EMF is always such that it produces a flux of $\vec{B}$ that opposes the change of the flux that produces the EMF. It tries to keep the current constant: it is opposite to the current when the current is increasing and in the direction of the current when it is decreasing.

This effect is particularly important for particle accelerators. A large electromagnet will have a large self-inductance. To change the current I in such a magnet requires a minimum voltage U to overcome this effect. This voltage is computed as:

$$U = -L\frac{\partial I}{\partial t}\,. \tag{47}$$

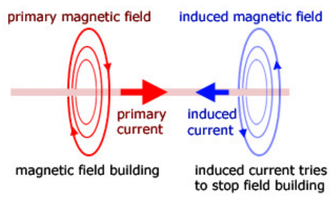


Fig. 10: Self-induction by a changing electric current

The self-inductance L is measured in henrys (H).

The necessary voltage is determined by this self-inductance and the rate of change of the current (Eq. (47)).

As a numerical example, we use the Large Hadron Collider parameters:

- required ramp rate, 10 A/s;
- self-inductance, L = 15.1 H per powering sector;
- required voltage to ramp at this rate, ≈150 V.

7.1 Lorentz force

A charge experiences forces in the presence of electromagnetic fields. This force depends not only on where it is (which determines the electromagnetic fields), but also on how it is moving. Moving ($\vec{v}$) charged (q) particles in electric ($\vec{E}$) and magnetic ($\vec{B}$) fields experience the force $\vec{f}$ (Lorentz force):

$$\vec{f} = q \cdot (\vec{E} + \vec{v} \times \vec{B})\,. \tag{48}$$

The electric force $q \cdot \vec{E}$ is always in the direction of the field $\vec{E}$ and proportional to the magnitude of the field and the charge.

The magnitude of the magnetic force $q \cdot \vec{v} \times \vec{B}$ is proportional to the velocity perpendicular to the direction of the field $\vec{B}$.

The Lorentz force is often treated as an ad-hoc plug-in to Maxwell's equation, but it is a relativistic effect (shown in Ref. [1]).

8 Electromagnetic waves in vacuum

A remarkable success of Maxwell's equations was the prediction of electromagnetic waves. Their existence was proven experimentally for very different wavelengths; in all cases, they were found to satisfy Maxwell's equations.

Starting from $\nabla \times \vec{E} = -\partial\vec{B}/\partial t$, we can apply several mathematical transformations in steps:

$$\begin{aligned} &\Longrightarrow \nabla \times (\nabla \times \vec{E}) = -\nabla \times \left(\frac{\partial \vec{B}}{\partial t}\right) \\ &\Longrightarrow -(\nabla^2 \vec{E}) = -\frac{\partial}{\partial t}(\nabla \times \vec{B}) \\ &\Longrightarrow -(\nabla^2 \vec{E}) = -\mu_0\epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} \end{aligned} \tag{49}$$

The last equation has the form of a plane wave.

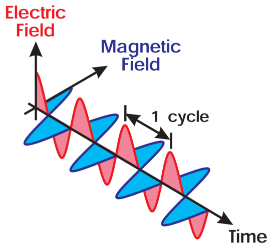


Fig. 11: Propagating electric and magnetic fields

This wave happens to be

$$\mu_0 \cdot \epsilon_0 = \frac{1}{c^2}\,,$$

and we rewrite:

$$\nabla^2 \vec{E} = \frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} = \mu_0 \cdot \epsilon_0 \cdot \frac{\partial^2 \vec{E}}{\partial t^2}$$

and

$$\nabla^2 \vec{B} = \frac{1}{c^2}\frac{\partial^2 \vec{B}}{\partial t^2} = \mu_0 \cdot \epsilon_0 \cdot \frac{\partial^2 \vec{B}}{\partial t^2} \tag{50}$$

This is a general form of a wave equation.

As a solution of these equations, we can write:

$$\begin{aligned} \vec{E} &= \vec{E}_0 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)}\,, \\ \vec{B} &= \vec{B}_0 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)}\,, \end{aligned} \tag{51}$$

where we use the following definitions:

$$\begin{array}{ll} \text{propagation vector}: & |\vec{k}| = \dfrac{2\pi}{\lambda} \;=\; \dfrac{\omega}{c}\,, \\ \text{wavelength, 1 cycle}: & \lambda\,, \\ \text{frequency} \cdot 2\pi: & \omega\,, \\ \text{wave velocity}: & c \;=\; \dfrac{\omega}{k}\,. \end{array} \tag{52}$$

Magnetic and electric fields are transverse to the direction of propagation (Fig. 11):

$$\vec{E} \perp \vec{B} \perp \vec{k} \quad \Rightarrow \quad \vec{k} \times \vec{E}_0 = \omega \vec{B}_0\,.$$

The speed of the wave in vacuum is exactly the speed of light: $c = 299792458$ m/s. Examples of the spectrum of electromagnetic waves are shown in Fig. 12 and Table 1.

The frequencies and, therefore, energies of existing waves span about 20 orders of magnitude.

9 Polarization of electromagnetic waves

9.1 General features

The solutions of the wave equations imply monochromatic plane waves. The solutions for electric and magnetic fields are:

$$\vec{E} = \vec{E}_0 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)}\,, \tag{53}$$

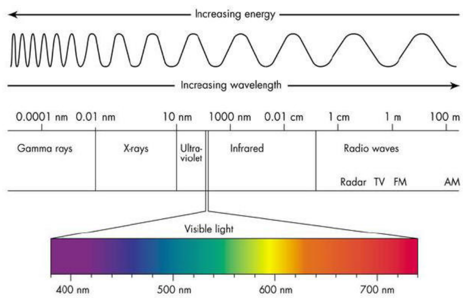


Fig. 12: Electromagnetic spectrum

Table 1: Properties of parts of the electromagnetic spectrum

Type	Frequency	Energy per photon
Radio	as low as 40 Hz	($\lesssim 10^{-13}$ eV)
Cosmic Microwave Background	$\lesssim 3 \cdot 10^{11}$ Hz	($\lesssim 10^{-3}$ eV)
Yellow light	$\approx 5 \cdot 10^{14}$ Hz	(≈ 2 eV)
X rays	$\leq 1 \cdot 10^{18}$ Hz	(≈ 4 keV)
γ rays	$\leq 3 \cdot 10^{21}$ Hz	(≤ 12 MeV)
$\pi^0 \; \to \gamma\gamma$	$\geq 2 \cdot 10^{22}$ Hz	(≥ 70 MeV)

$$\vec{B} = \vec{B_0} \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)} \,. \tag{54}$$

These equations can be rewritten using unit vectors in the plane transverse to propagation. For example, for the electric component:

$$\vec{\epsilon_1} \perp \vec{\epsilon_2} \perp \vec{k} \,.$$

The two orthogonal components are:

$$\vec{E_1} = \vec{\epsilon_1} E_1 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)} \,,$$
$$\vec{E_2} = \vec{\epsilon_2} E_2 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)} \,.$$

The general field is a superposition of the two components:

$$\Rightarrow \vec{E} = (\vec{E_1} + \vec{E_2}) = (\vec{\epsilon_1} E_1 + \vec{\epsilon_2} E_2) \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)} \,. \tag{55}$$

For the propagation, we can allow for a phase shift ϕ between the two directions as well as different amplitudes:

$$\vec{E} = \vec{\epsilon_1} E_1 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t)} + \vec{\epsilon_2} E_2 \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r}-\omega t+\phi)} \,.$$

Depending on the amplitudes E_1 and E_2 and the relative phase ϕ, we get different types of polarized light:

$$\begin{array}{ll} \phi = 0 : & \text{linearly polarized light}\,; \\ \phi \neq 0 \text{ and } E_1 \neq E_2 : & \text{elliptically polarized light}\,; \\ \phi = \pm\dfrac{\pi}{2} \text{ and } E_1 = E_2 : & \text{circularly polarized light}\,. \end{array}$$

9.2 Polarized light in accelerators

Polarized light is rather important in accelerators and is produced (amongst others) in synchrotron light machines (linearly and circularly polarized light, adjustable).

Typical applications and phenomena of polarized light are:

- polarized light reacts differently with charged particles;
- material science;
- beam diagnostics, medical diagnostics (blood sugar, ...);
- inverse free electron lasers;
- 3-D motion pictures, LCD display, outdoor activities, cameras (glare), ...

10 Energy of electromagnetic waves

We define the *Poynting vector* (SI units):

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B} \,. \tag{56}$$

The vector $\vec{S}$ points in the direction of propagation and describes the 'energy flux', i.e., energy crossing a unit area, per second [J / m^2 s].

In free space, the energy in a plane is shared between the electric and magnetic fields The energy density $\mathcal{H}$ would be:

$$\mathcal{H} = \frac{1}{2} \left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2 \right) \,. \tag{57}$$

11 Electromagnetic waves in material

We start now with the macroscopic Maxwell's equations (Eq. (32)), using $\mu_0 \vec{H} = \vec{B}$ and $\epsilon_0 \vec{E} = \vec{D}$:

$$\begin{aligned} \nabla \times \vec{E} &= -\mu_0 \frac{\mathrm{d}\vec{H}}{\mathrm{d}t} \,, \\ \nabla \times \vec{H} &= \vec{j} \, + \, \epsilon_0 \frac{\mathrm{d}\vec{E}}{\mathrm{d}t} \,. \end{aligned} \tag{58}$$

We assume a material with relative permittivity ϵ and permeability μ, as well as a finite conductivity σ, and get:

$$\begin{aligned} \nabla \times \vec{E} &= -\mu \cdot \mu_0 \cdot \frac{\mathrm{d}\vec{H}}{\mathrm{d}t} \,, \\ \nabla \times \vec{H} &= \sigma \vec{E} + \epsilon \cdot \epsilon_0 \cdot \frac{\mathrm{d}\vec{E}}{\mathrm{d}t} \,, \end{aligned} \tag{59}$$

where the current density $\vec{j}$ is replaced by $\sigma \vec{E}$ (Ohm's law). Following the same procedure as before, we obtain for the wave equation (electric field only):

$$\nabla^2 \vec{E} = \sigma \cdot \epsilon \cdot \epsilon_0 \cdot \frac{\partial \vec{E}}{\partial t} + \mu \cdot \mu_0 \cdot \epsilon \cdot \epsilon_0 \cdot \frac{\partial^2 \vec{E}}{\partial t^2} \tag{60}$$

For non-conducting media, we can set $\sigma \; = \; 0$ and obtain the previous equations.

As a direct consequence of Eq. (60) we see that the speed of this wave in the medium is now:

$$v = \frac{1}{\sqrt{\epsilon_0 \cdot \mu_0 \cdot \epsilon \cdot \mu}} \,, \tag{61}$$

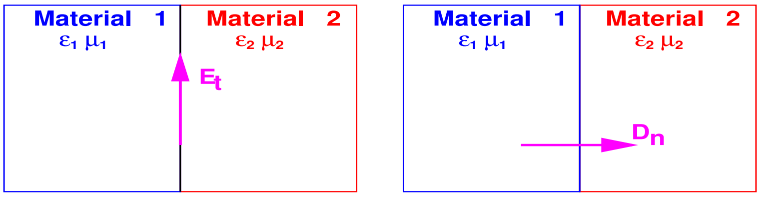


Fig. 13: Boundary conditions for electric fields

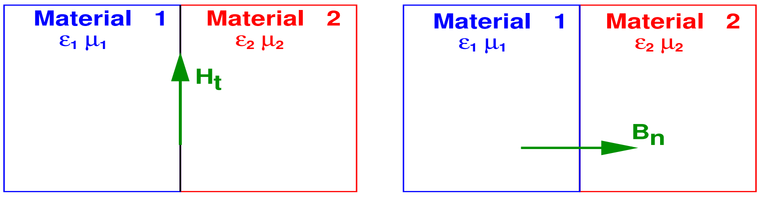


Fig. 14: Boundary conditions for magnetic fields

or, if rewritten using $n = \sqrt{\epsilon \cdot \mu}$,

$$v = \frac{c}{n}\,. \tag{62}$$

The speed of electromagnetic waves in vacuum is c, but reduced by the factor n in a medium with relative permittivity ϵ and permeability μ.

11.1 Boundary conditions

When electromagnetic waves pass through the boundary between two media with different ϵ and μ, we must fulfil some boundary conditions. The results are presented here without proof. For details see Refs. [3, 7]. Assuming *no* surface charges and, from curl$\vec{E} = 0$ we can derive that the *tangential* $\vec{E}$*-field* is continuous across a boundary ($E_t^1 = E_t^2$) (shown schematically in Fig. 13). Similarly, since we have div$\vec{D} = \rho$, the *normal* $\vec{D}$*-field* must be continuous across the boundary ($D_n^1 = D_n^2$) (shown schematically in Fig.13).

We follow the same line of reasoning for the boundary conditions for magnetic fields. Assuming *no* surface currents (for a proof, see, e.g., Refs. [3, 7]), we find (see Fig. 14):

From curl$\vec{H} = \vec{j}$,
$\Rightarrow$ *tangential* $\vec{H}$*-field* continuous across boundary $(H_t^1 = H_t^2)$.
From div $\vec{B} = 0$,
$\Rightarrow$ *normal* $\vec{B}$*-field* continuous across boundary $(B_n^1 = B_n^2)$.

A short summary for the electromagnetic fields at boundaries between different materials with different permittivity and permeability ($\epsilon_1, \epsilon_2, \mu_1, \mu_2$) is:

$$\begin{aligned} &(E_t^1 = E_t^2) && (E_n^1 \neq E_n^2)\,, \\ &(D_t^1 \neq D_t^2) && (D_n^1 = D_n^2)\,, \\ &(H_t^1 = H_t^2) && (H_n^1 \neq H_n^2)\,, \\ &(B_t^1 \neq B_t^2) && (B_n^1 = B_n^2)\,. \end{aligned} \tag{63}$$

These conditions lead to reflection and refraction of the waves at the surface; the angles are related to the refraction index $n = \sqrt{\epsilon_1 \mu_1}$ and $n' = \sqrt{\epsilon_2 \mu_2}$.

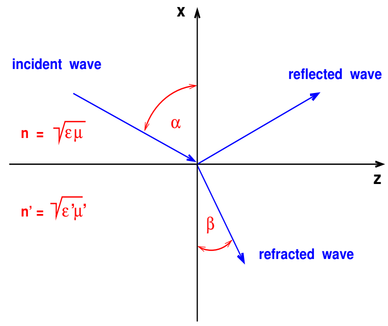


Fig. 15: Reflected and refracted components of an incident wave

The connection between the refraction indices and the scattering and refraction angles shown in Fig. 15 are:

$$\frac{\sin\alpha}{\sin\beta} = \frac{n'}{n} = \tan\alpha_{\mathrm{B}}\,. \tag{64}$$

If ϵ and μ depend on the wave frequency ω, the medium is dispersive and we have to write:

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} \neq 0\,, \tag{65}$$

i.e., the refraction index and therefore the angles depend on the wavelength.

If light is incident under the special angle α_{B} (the Brewster angle) [3], the reflected light is linearly polarized perpendicular to the plane of incidence.

A popular application is used when fishing, since air–water gives a comfortable angle $\alpha_{\mathrm{B}} \approx 53°$ and reflections can be avoided using polarization glasses.

12 Cavities and waveguides

Of particular interest in accelerator physics and technology is the behaviour and propagation of electromagnetic waves in cavities and waveguides. This behaviour is determined by the boundary conditions and we have to distinguish between material with infinite and finite conductivity. The case of perfectly conducting cavities and wave guides is treated first.

12.1 Rectangular cavities and waveguides

Cavities can be seen as a three-dimensional storage for electromagnetic waves, i.e., photons. The wave functions are contained inside and therefore the dimensions determine the maximum wavelength that can fit inside. This is due to the boundary conditions at the cavity walls.

If the fields are only constrained in two dimensions and allowed to move freely in the third dimension, the fields propagate as waves. The waves are guided through these 'wave guides'. Both are sketched in Fig. 16.

12.2 Cavities and modes

We assume a rectangular RF cavity with dimensions (a, b, c), and as an ideal conductor.

Without derivations (e.g., Refs. [3, 7, 8]), the components of the electric fields are:

$$E_x = E_{x0} \cdot \cos(k_x x) \cdot \sin(k_y y) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,,$$
$$E_y = E_{y0} \cdot \sin(k_x x) \cdot \cos(k_y y) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,,$$

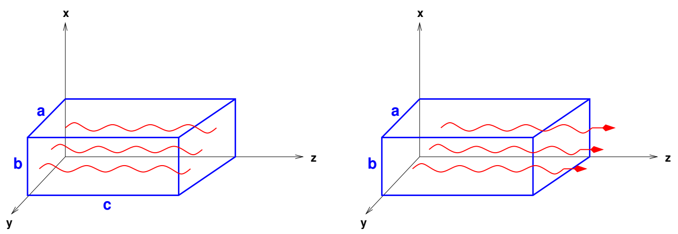


Fig. 16: Boundary conditions for electromagnetic fields. Fields are fully enclosed in a cavity (left-hand side) and can move freely in one dimension in waveguides (right-hand side).

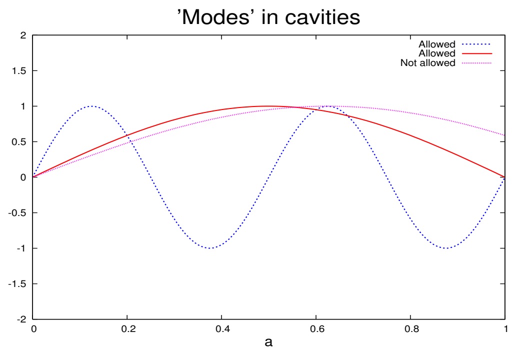


Fig. 17: Boundary conditions for electromagnetic fields

$$E_z = E_{z0} \cdot \sin(k_x x) \cdot \sin(k_y y) \cdot \cos(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,. \tag{66}$$

For the magnetic fields we get immediately, with $\nabla \times \vec{E} = -\partial \vec{B}/\partial t$:

$$\begin{aligned}
B_x &= \frac{\mathrm{i}}{\omega}(E_{y0}k_z - E_{z0}k_y) \cdot \sin(k_x x) \cdot \cos(k_y y) \cdot \cos(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_y &= \frac{\mathrm{i}}{\omega}(E_{z0}k_x - E_{x0}k_z) \cdot \cos(k_x x) \cdot \sin(k_y y) \cdot \cos(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_z &= \frac{\mathrm{i}}{\omega}(E_{x0}k_y - E_{y0}k_x) \cdot \cos(k_x x) \cdot \cos(k_y y) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,.
\end{aligned} \tag{67}$$

12.3 Consequences for cavities

The fields must be zero at the conductor boundary, as shown before. This is possible only with the condition:

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}\,, \tag{68}$$

and for k_x, k_y, k_z we can write:

$$k_x = \frac{m_x \pi}{a}\,, \qquad k_y = \frac{m_y \pi}{b}\,, \qquad k_z = \frac{m_z \pi}{c}\,. \tag{69}$$

The integer numbers m_x, m_y, m_z are called the mode numbers of the wave and are directly related to the dimensions of the cavity.

Equations (68) and (69) imply that a half wavelength $\lambda/2$ must always fit exactly the size of the cavity. This is shown in Fig. 17 for different wavelengths compared with the cavity dimensions. Only modes that 'fit' into the cavity are allowed.

We can examine three cases:

$$\frac{\lambda}{2} = \frac{a}{4}, \qquad \frac{\lambda}{2} = \frac{a}{1}, \qquad \frac{\lambda}{2} = \frac{a}{0.8}.$$

No electric field at the boundaries requires that the wave must have 'nodes' at the boundaries. Only the first two wavelengths fulfil this condition; the third form cannot exist.

12.4 Waveguide modes

Similar considerations lead to (propagating) solutions in (rectangular) waveguides:

$$\begin{aligned}
E_x &= E_{x0} \cdot \cos(k_x x) \cdot \sin(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}, \\
E_y &= E_{y0} \cdot \sin(k_x x) \cdot \cos(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}, \\
E_z &= \mathrm{i} \cdot E_{z0} \cdot \sin(k_x x) \cdot \sin(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}, \qquad (70) \\
B_x &= \frac{1}{\omega}(E_{y0} k_z - E_{z0} k_y) \cdot \sin(k_x x) \cdot \cos(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}, \\
B_y &= \frac{1}{\omega}(E_{z0} k_x - E_{x0} k_z) \cdot \cos(k_x x) \cdot \sin(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}, \\
B_z &= \frac{1}{\mathrm{i} \cdot \omega}(E_{x0} k_y - E_{y0} k_x) \cdot \cos(k_x x) \cdot \cos(k_y y) \cdot \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}. \qquad (71)
\end{aligned}$$

12.5 Consequences for waveguides

To have no field at the boundary, we must again satisfy the condition:

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}. \qquad (72)$$

This leads to modes like (no boundaries in direction of propagation z):

$$k_x = \frac{m_x \pi}{a}, \qquad k_y = \frac{m_y \pi}{b}, \qquad (73)$$

The numbers m_x, m_y are called the mode numbers for planar waves in waveguides.

12.6 Cut-off frequency

One can rewrite Eq. (72) as:

$$k_z^2 = \frac{\omega^2}{c^2} - k_x^2 - k_y^2 \qquad (74)$$

and

$$k_z = \sqrt{\frac{\omega^2}{c^2} - k_x^2 - k_y^2}. \qquad (75)$$

A propagation without losses requires k_z to be real, i.e.,

$$\frac{\omega^2}{c^2} > k_x^2 + k_y^2 = \left(\frac{m_x \pi}{a}\right)^2 + \left(\frac{m_y \pi}{b}\right)^2. \qquad (76)$$

This defines a cut-off frequency ω_c:

$$\omega_\mathrm{c} = \frac{\pi \cdot c}{a}. \qquad (77)$$

For frequencies above this cut-off frequency, we have propagation without losses. At the cut-off frequency, we obtain a standing wave and an attenuated wave for lower frequencies, i.e., k_z becomes complex.

The cut-off frequencies are different for different modes and no modes can propagate below the lowest frequency. The mode of Eq. (77) is assumed to be this lowest frequency mode.

12.7 Circular cavities

Waveguides and cavities used in accelerators are more likely to be circular.

Derivation involves using the Laplace equation in cylindrical co-ordinates; for the derivation see e.g., Refs. [7, 8]:

$$
\begin{aligned}
E_r &= E_0 \frac{k_z}{k_r} J_n'(k_r) \cdot \cos(n\theta) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
E_\theta &= E_0 \frac{n k_z}{k_r^2 r} J_n(k_r) \cdot \sin(n\theta) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
E_z &= E_0 J_n(k_r r) \cdot \cos(n\theta) \cdot \sin(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_r &= \mathrm{i} E_0 \frac{\omega}{c^2 k_r^2 r} J_n(k_r r) \cdot \sin(n\theta) \cdot \cos(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_\theta &= \mathrm{i} E_0 \frac{\omega}{c^2 k_r r} J_n'(k_r r) \cdot \cos(n\theta) \cdot \cos(k_z z) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_z &= 0\,.
\end{aligned} \tag{78}
$$

12.8 Accelerating circular cavities

For accelerating cavities, we need a longitudinal electric field component $E_z \neq 0$ and purely transverse magnetic fields:

$$
\begin{aligned}
E_r &= 0\,, \\
E_\theta &= 0\,, \\
E_z &= E_0 J_0 \left(p_{01} \frac{r}{R}\right) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_r &= 0\,, \\
B_\theta &= -\mathrm{i} \frac{E_0}{c} J_1 \left(p_{01} \frac{r}{R}\right) \cdot \mathrm{e}^{-\mathrm{i}\omega t}\,, \\
B_z &= 0\,.
\end{aligned} \tag{79}
$$

(p_{nm} is the mth zero of J_n, e.g., $p_{01} \approx 2.405$.)

This would be a cavity with a TM_{010} mode: $\omega_{010} = p_{01} \cdot c/R$.

13 Case of finite conductivity

Starting from Maxwell's equation,

$$
\nabla \times \vec{H} = \vec{j} + \frac{\mathrm{d}\vec{D}}{\mathrm{d}t} = \underbrace{\overbrace{\sigma \cdot \vec{E}}^{\vec{j}}}_{\text{Ohm's law}} + \epsilon \frac{\mathrm{d}\vec{E}}{\mathrm{d}t}\,, \tag{80}
$$

and the solutions of the wave equations,

$$
\vec{E} = \vec{E_0} \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r} - \omega t)}, \qquad \vec{H} = \vec{H_0} \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{r} - \omega t)}\,, \tag{81}
$$

we want to know k; applying the calculus to the wave equations we have:

$$
\frac{\mathrm{d}\vec{E}}{\mathrm{d}t} = -\mathrm{i}\omega \cdot \vec{E}, \qquad \frac{\mathrm{d}\vec{H}}{\mathrm{d}t} = -\mathrm{i}\omega \cdot \vec{H}, \qquad \nabla \times \vec{E} = \mathrm{i}\vec{k} \times \vec{E}, \qquad \nabla \times \vec{H} = \mathrm{i}\vec{k} \times \vec{H}\,. \tag{82}
$$

Put these together, using Eqs. (80) and (82):

$$
\vec{k} \times \vec{H} = \mathrm{i}\sigma \cdot \vec{E} - \omega\epsilon \cdot \vec{E} = (-\mathrm{i}\sigma + \omega\epsilon) \cdot \vec{E}\,. \tag{83}
$$

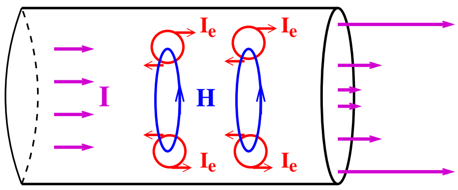


Fig. 18: Flow of current and induced magnetic fields and eddy currents

With $\vec{B} = \mu\vec{H}$:

$$\nabla \times \vec{E} = \mathrm{i}\vec{k} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -\mu\frac{\partial \vec{H}}{\partial t} = \mathrm{i}\omega\mu\vec{H}\,. \tag{84}$$

Multiplication with $\vec{k}$ and using Eq. (83):

$$\vec{k} \times (\vec{k} \times \vec{E}) = \omega\mu(\vec{k} \times \vec{H}) = \omega\mu(-\mathrm{i}\sigma + \omega\epsilon) \cdot \vec{E}\,. \tag{85}$$

After some calculus and using the property $\vec{E} \perp \vec{H} \perp \vec{k}$:

$$k^2 = \omega\mu(\mathrm{i}\sigma + \omega\epsilon)\,. \tag{86}$$

The propagation vector k now differs from the equation in vacuum by the contributions from the medium and the finite conductivity. This has consequences for the propagation and penetration of waves in material.

13.1 Skin and penetration depth

For a good conductor, $\sigma \gg \omega\epsilon$ (e.g., for Cu we have $\sigma \approx 5.7 \cdot 10^7$ S/m, this value for Cu is also valid for for very high ω):

$$k^2 \approx -\mathrm{i}\omega\mu\sigma \qquad \Rightarrow \qquad k \approx \sqrt{\frac{\omega\mu\sigma}{2}}(1-\mathrm{i}) = \frac{1}{\delta}(1-\mathrm{i})\,. \tag{87}$$

We define the parameter δ:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}\,. \tag{88}$$

The parameter δ is called the *skin depth*.

From Eq. (88), we deduce that high frequency waves 'avoid' penetrating a conductor, and mainly flow near the surface. One can understand this effect using Fig. 18.

A changing $\vec{H}$-field induces eddy currents in the conductor. These cancel the current flow in the centre of the conductor but enforce current flow at the skin (surface).

13.2 Attenuated waves

Waves incident on conducting material are attenuated. It is basically skin depth, (attenuation to 1/e). The wave form becomes:

$$\mathrm{e}^{\mathrm{i}(kz-\omega t)} = \mathrm{e}^{\mathrm{i}((1+\mathrm{i})z/\delta-\omega t)} = \mathrm{e}^{\frac{-z}{\delta}} \cdot \mathrm{e}^{\mathrm{i}(\frac{z}{\delta}-\omega t)}\,. \tag{89}$$

Some numerical examples:

– Skin depth of copper:
  1 GHz : $\delta \approx 2.1$ µm; 1 kHz : $\delta \approx 2.1$ mm, 50 Hz : $\delta \approx 10$ mm.
  This has important consequences for the design of conducting cables since the high frequency currents propagate at a very thin layer at the surface of the conductor.

- Penetration depth into seawater (σ typically 4 S/m):
  To get $\delta \approx 25$ m, one needs $\approx$76 Hz.
  Because of the long wavelength and low frequency, communication is very inefficient and has a very low bandwidth (0.03 bps).

14 Summary

Without any attempt to be rigorous or complete, electromagnetic effects most important for the design and operation of particle accelerators have been presented, such as:

- basic concepts;
- Maxwell's equations;
- fields and potentials from charge and current distributions;
- electromagnetic waves in vacuum and media;
- electromagnetic waves in waveguides and cavities;
- polarization of electromagnetic waves and skin effects.

References

[1] W. Herr, Short theory of special relativity and invariant formulation of electrodynamics, these proceedings.

[2] Proceedings of the CAS-CERN Accelerator School on RF for Accelerators, Ebeltoft, Denmark, 8–17 June 2010, edited by R. Bailey, CERN-2011-007 (CERN, Geneva, 2011), https://doi.org/10.5170/CERN-2011-007

[3] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998).

[4] L. Landau and E. Lifschitz, *The Classical Theory of Fields* (Butterworth-Heinemann, Oxford, 1975).

[5] J. Slater and N. Frank, *Electromagnetism* (McGraw-Hill, New York, 1947).

[6] W. Herr, in Proc. of the CAS-CERN Accelerator School: Intermediate Course on Accelerator Physics, Zeuthen, Germany, 15–26 September 2003, edited by D. Brandt, CERN-2006-002 (CERN, Geneva, 2006), pp. 379-410. https://doi.org/10.5170/CERN-2006-002.379

[7] R.P. Feynman, *Feynman Lectures on Physics*, (Basic Books, New York, 2001), Vol. 2.

[8] A. Chao *et al.*, *Handbook of Accelerator Physics and Engineering*, World Scientific Publishing, New Jersey (1998).

Appendices

A Electromagnetic units

Formulae use *SI units* throughout.

$\vec{E}(\vec{r},t)$	electric field [V/m]
$\vec{H}(\vec{r},t)$	magnetic field [A/m]
$\vec{D}(\vec{r},t)$	electric displacement [C/m^2]
$\vec{B}(\vec{r},t)$	magnetic flux density [T]
q	electric charge [C]
$\rho(\vec{r},t)$	electric charge density [C/m^3]
$\vec{I}, \vec{j}(\vec{r},t)$	current [A], current density [A/m^2]
μ_0	permeability of vacuum, $4\,\pi \cdot 10^{-7}$ [H/m or N/A^2]
ϵ_0	permittivity of vacuum, $8.854 \cdot 10^{-12}$ [F/m]

To save typing and space where possible (e.g., equal arguments):
$\vec{E}(\vec{r},t) \Longrightarrow \vec{E}$ and the same for other variables.

B Refresher on vector calculus

B.1 Vector operators

We can define a special vector ∇ (sometimes written as $\vec{\nabla}$):

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right). \tag{B.1}$$

It is called the 'gradient' and invokes 'partial derivatives'.

It can operate on a scalar function $\phi(x,y,z)$:

$$\nabla\phi = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right) = \vec{G} = (G_x, G_y, G_z)\,, \tag{B.2}$$

and we get a vector $\vec{G}$. It is a kind of 'slope' (steepness) in the three directions.
Example:

$$\phi(x,y,z) = C \cdot \mathbf{ln}(r^2) \qquad \text{with} \qquad r = \sqrt{x^2+y^2+z^2}$$
$$\nabla\phi = (G_x, G_y, G_z) = \left(\frac{2C\cdot x}{r^2}, \frac{2C\cdot y}{r^2}, \frac{2C\cdot z}{r^2}\right)$$

B.1.1 Physical interpretation of the gradient operator

The gradient applied to a scalar field measures the local slope, as shown in Fig. B.1:

- lines of pressure (isobars);
- gradient is large (steep) where lines are close (fast change of pressure).

B.2 Operation on vectors and scalar fields

The gradient ∇ can be used in a scalar or a vector product with a vector $\vec{F}$, sometimes written as $\vec{\nabla}$ and these are used as:

$$\nabla \cdot \vec{F} \qquad \text{or} \qquad \nabla \times \vec{F}\,. \tag{B.3}$$

The definition for products is the same as before; ∇ is treated like a 'normal' vector, but the results depend on how they are applied:

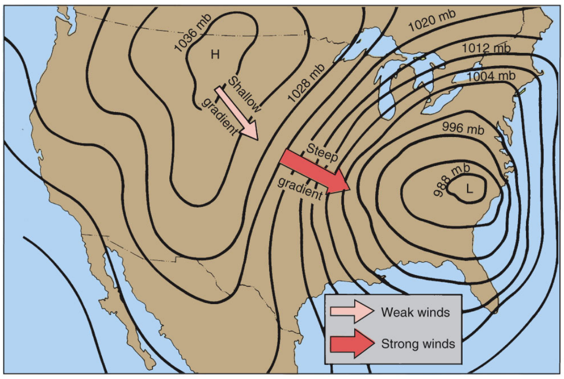


Fig. B.1: Gradient of a scalar field (here air pressure)

- $\nabla\phi$ is a vector;
- $\nabla \cdot \vec{F}$ is a scalar;
- $\nabla \times \vec{F}$ is a pseudo-vector.

B.3 Divergence and curl

Two operations of ∇ have special names.

B.3.1 Divergence (scalar product of gradient with a vector):

$$\mathsf{div}(\vec{F}) = \nabla \cdot \vec{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\,. \tag{B.4}$$

Physical significance: ‘amount of density’ (see later).

B.3.2 Curl (vector product of gradient with a vector):

$$\mathsf{curl}(\vec{F}) = \nabla \times \vec{F} = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z},\, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x},\, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\,. \tag{B.5}$$

Physical significance: ‘amount of rotation’ (see later).

B.3.3 Meaning of divergence

Figure B.2 shows the field lines of a vector field $\vec{F}$ seen from some origin.

The divergence (scalar, a single number) characterizes what comes from (or goes to) the origin. How much comes out is measured by the surface integral. For the integrated field vectors passing (perpendicularly) through a surface area A, we obtain the flux:

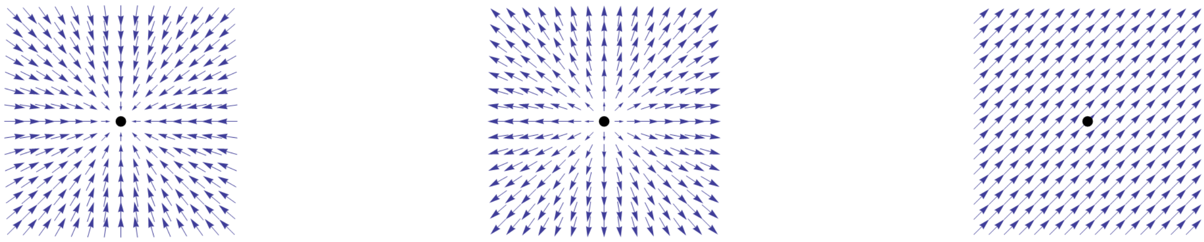
$$\int\!\!\int_A \vec{F} \cdot \mathrm{d}\vec{A}\,. \tag{B.6}$$

It has the meaning of the density of field lines through the surface (Fig. B.3).

For closed surfaces, we can rewrite the integral using Gauss’s theorem: the integral through a closed surface (flux) is the integral of the divergence *in the enclosed volume*

$$\int\!\!\int_A \vec{F} \cdot \mathrm{d}\vec{A} = \int\!\!\int\!\!\int_V \nabla \cdot \vec{F} \cdot \mathrm{d}V\,. \tag{B.7}$$

This relates surface integral to volume integrals (Fig. B.4) and is often easier to evaluate.



$\nabla\vec{F} < 0$ (sink) $\nabla\vec{F} > 0$ (source) $\nabla\vec{F} = 0$ (fluid)

Fig. B.2: Field lines of a vector field $\vec{F}$ seen from some origin

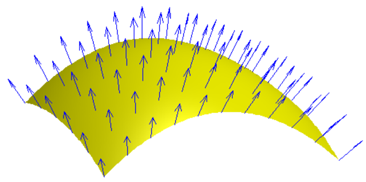


Fig. B.3: Flux

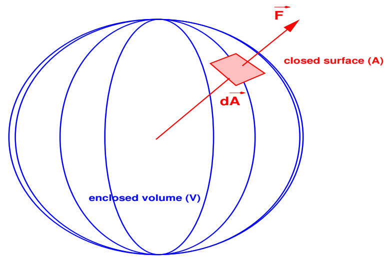


Fig. B.4: Gauss's theorem relates surface integrals to volume integrals

B.3.4 Meaning of curl

The *curl* quantifies a rotation of vectors: it is the integration of (vector-) fields. For two vector fields, we perform a *line integral* along a (closed) line C:

$$\oint_C \vec{F} \cdot \mathrm{d}\vec{r} = \int\int_A \nabla \times \vec{F} \cdot \mathrm{d}\vec{A} \tag{B.8}$$

i.e., we 'sum up' vectors (length) in the *direction* of the line C

The line integral for the second vector field in Fig. B.5 vanishes because the field lines are orthogonal to the direction of the integration path along the curve C. The physical significance of this line integral is the work performed along a path.

We can formulate this integral more generally:
Stokes' theorem: *Integral along a closed line is integral of curl in the enclosed area.*

$$\oint_C \vec{F} \cdot \mathrm{d}\vec{s} = \int\int_A \nabla \times \vec{F} \cdot \mathrm{d}\vec{A}\,. \tag{B.9}$$

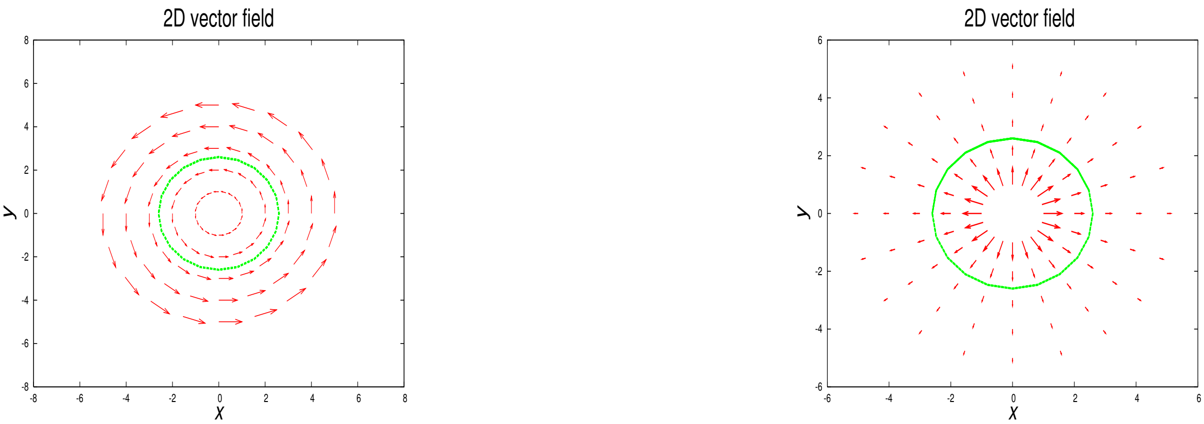


Fig. B.5: Two types of vector field, arbitrary units. For the left field we have: $\nabla \vec{F} = 0 \quad \nabla \times \vec{F} \neq 0$. For the right field: $\nabla \vec{F} \neq 0 \quad \nabla \times \vec{F} = 0$.

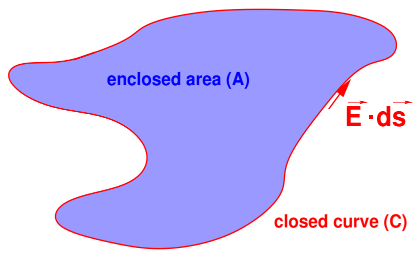


Fig. B.6: Stoke's theorem

B.4 Scalar product

We define a scalar product for (usual) vectors as: $\vec{a} \cdot \vec{b}$,

$$\vec{a} = (x_a, y_a, z_a)\,,$$
$$\vec{b} = (x_b, y_b, z_b)\,,$$
$$\vec{a} \cdot \vec{b} = (x_a, y_a, z_a) \;\cdot\; (x_b, y_b, z_b) = (x_a \cdot x_b \;+ y_a \cdot y_b \;+\; z_a \cdot z_b)\,.$$

This product of two vectors is a *scalar* (number), not a vector.

Example:

$$(-2, 2, 1) \cdot (2, 4, 3) = -2 \cdot 2 + 2 \cdot 4 + 1 \cdot 3 = 7\,.$$

B.5 Vector product (sometimes referred to as cross product)

Define a vector product for (usual) vectors as: $\vec{a} \times \vec{b}$,

$$\vec{a} = (x_a, y_a, z_a)\,,$$
$$\vec{b} = (x_b, y_b, z_b)\,,$$
$$\vec{a} \times \vec{b} = (x_a, y_a, z_a) \;\times\; (x_b, y_b, z_b)$$
$$= (\underbrace{y_a \cdot z_b - z_a \cdot y_b,}_{x_{ab}} \;\; \underbrace{z_a \cdot x_b - x_a \cdot z_b,}_{y_{ab}} \;\; \underbrace{x_a \cdot y_b - y_a \cdot x_b}_{z_{ab}})\,.$$

This product of two vectors is a *vector*, not a scalar (number), (on this account: vector product).

Example 1:

$$(-2, 2, 1) \times (2, 4, 3) = (2, 8, -12)\,.$$

Example 2 (two components only in the x–y plane):

$$(-2, 2, 0) \times (2, 4, 0) = (0, 0, -12)\,.$$

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Short Overview of Special Relativity and Invariant Formulation of Electrodynamics

W. Herr
CERN, Geneva, Switzerland

Abstract
The basic concepts of special relativity are presented in this paper. Consequences for the design and operation of particle accelerators are discussed, along with applications. Although all branches of physics must fulfil the principles of special relativity, the focus of this paper is the application to electromagnetism. The formulation of physics laws in the form of four-vectors allows a fully invariant formulation of electromagnetic theory and a reformulation of Maxwell's equations. This significantly simplifies the treatment of moving charges in electromagnetic fields and can explain some open questions.

Keywords
Special relativity; electrodynamics; four-vectors.

1 Introduction and motivation

As a principle in physics, the laws of physics should take the same form in all frames of reference, i.e., they describe a symmetry, a very basic concept in modern physics. This concept of relativity was introduced by Galileo and Newton in the framework of classical mechanics. Classical electromagnetic theory as formulated by Maxwell's equations leads to asymmetries when applied to moving charges [1,2]. In this context, classical mechanics and classical electromagnetism do not fulfil the same principles of relativity. The theory of special relativity is a generalization of the Galilean and Newtonian concepts of relativity. It also paved the way to a consistent theory of quantum mechanics. It considerably simplifies the form of physics because the unity of space and time as formulated by Minkowski also applies to force and power, time and energy, and last, but not least, to electric current and charge densities. The formulation of electromagnetic theory in this framework leads to a consistent picture and explains such concepts as the Lorentz force in a natural way. Starting from basic considerations and the postulates for special relativity, we develop the necessary mathematical formalism and discuss consequences, such as length contraction, time dilation, and the relativistic Doppler effect, to mention some of the most relevant. The introduction of four-vectors automatically leads to a relativistically invariant formulation of Maxwell's equations, together with the laws of classical mechanics.

Unlike other papers on relativity, this paper concentrates on aspects of electromagnetism; other popular phenomena, such as paradoxes, are left out.

2 Concepts of relativity

The concept of relativity was introduced by Galileo and Newton and applied to classical mechanics. It was proposed by Einstein that a similar concept should be applicable when electromagnetic fields are involved. We shall move from the classical principles to electrodynamics and assess the consequences.

2.1 Relativity in classical mechanics

In the following, the terminology and definitions used are:

– co-ordinates for the formulation of physics laws:

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.27

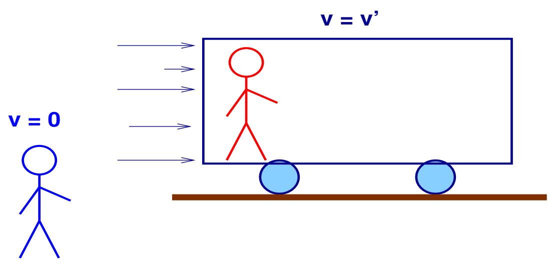


Fig. 1: Two different frames: a resting and a moving observer

  - space co-ordinates: $\vec{x} = (x, y, z)$ (not necessarily Cartesian);
  - time: t
    (side note: it might be better practice to use $\vec{r} = (x, y, z)$ instead of $\vec{x}$ as the position vector to avoid confusion with the x-component but we maintain this convention to be compatible with other textbooks and the conventions used later);
- definition of a *frame*:
  - where we observe physical phenomena and properties as functions of their position $\vec{x}$ and time t;
  - an *inertial frame* is a frame moving at a constant velocity;
  - in different frames, $\vec{x}$ and t are usually different;
- definition of an *event*:
  - something happening at $\vec{x}$ at time t is an 'event', given by four numbers: $(x, y, z), t$.

An example for two frames is shown in Fig. 1: one observer is moving at a constant relative velocity v' and another is observing from a resting frame.

2.2 Galileo transformation

How do we relate observations, e.g., the falling object in the two frames shown in Fig. 2?

- We have observed and described an event in rest frame F using co-ordinates (x, y, z) and time t, i.e., have formulated the physics laws using these co-ordinates and time.
- To describe the event in another frame F' moving at a constant velocity in the x-direction v_x, we describe it using co-ordinates (x', y', z') and t'.
- We need a transformation for:
  (x, y, z) and $t \Rightarrow (x', y', z')$ and t'.

The laws of classical mechanics are invariant, i.e., have the same form with the transformation:

$$\begin{aligned} x' &= x - v_x t\,, \\ y' &= y\,, \\ z' &= z\,, \\ t' &= t\,. \end{aligned} \tag{1}$$

The transformation (Eq. (1)) is known as the Galileo transformation. Only the position in the direction of the moving frame is transformed; time remains an absolute quantity.

Fig. 2: Observing a falling object from a moving and from a resting frame

2.3 Example of an accelerated object

An object falling with an acceleration g in the moving frame (Fig. 2, left) falls in a straight line observed within this frame.

Equation of motion in a moving frame $x'(t')$ and $y'(t')$:

$$\begin{aligned} x'(t') &= 0\,, \\ v'_y(t') &= -g \cdot t'\,, \\ y'(t') &= \int v'_y(t')\mathrm{d}t' = -\frac{1}{2}gt'^2\,. \end{aligned} \tag{2}$$

To get the equation of motion in the rest frame $x(t)$ and $y(t)$, the Galileo transform is applied:

$$\begin{aligned} y(t) &= y'(t')\,, \\ t &= t'\,, \\ x(t) &= x' + v_x \cdot t = v_x \cdot t\,, \end{aligned} \tag{3}$$

and one obtains for the trajectories $y(t)$ and $y(x)$ in the rest frame:

$$y(t) = -\frac{1}{2}gt^2\,, \qquad y(x) = -\frac{1}{2}g\frac{x^2}{v_x^2}\,. \tag{4}$$

From the resting frame, $y(x)$ describes a parabola (Fig. 2, right-hand side).

2.4 Addition of velocities

An immediate consequence of the Galileo transformation (Eq. (1)) is that the velocities of the moving object and the moving frame must be added to get the observed velocity in the rest frame:

$$v = v' + v''\,, \tag{5}$$

because (e.g., moving with the speed v_x in the x-direction):

$$\frac{\mathrm{d}x'}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t} - v_x\,. \tag{6}$$

As a very simple example (Fig. 3), the total speed of the object is 191 m/s.

2.5 Problems with Galileo transformation applied to electromagnetism

Applied to electromagnetic phenomena, the Galileo transformation exhibits some asymmetries. Assume a magnetic field and a conducting coil moving relative to the magnetic field. An induced current will be measured in the coil (Fig. 4). Depending on the frame of the observer, the interpretation of the observation is different.

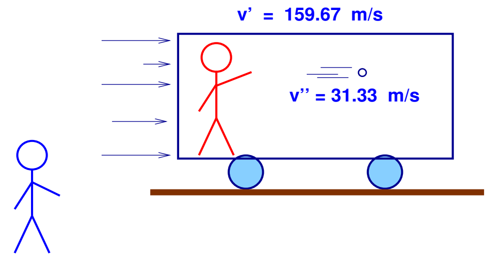


Fig. 3: Measured velocities of an object as observed from the co-moving and rest frames

Fig. 4: Effect of relative motion of a magnetic field and a conducting coil, observed from a co-moving and the rest frame.

– If you sit on the coil, you observe a changing magnetic field, leading to a circulating electric field inducing a current in the coil:

$$\frac{\mathrm{d}\vec{B}}{\mathrm{d}t} \Rightarrow \vec{\nabla} \times \vec{E} \Rightarrow \vec{F} = q \cdot \vec{E} \Rightarrow \text{current in coil}\,. \tag{7}$$

– If you sit on the magnet, you observe a moving charge in a magnetic field, leading to a force on the charges in the coil:

$$\vec{B} = \text{const., moving charge} \Rightarrow \vec{F} = q \cdot \vec{v} \times \vec{B} \Rightarrow \text{current in coil}\,. \tag{8}$$

The observed results are identical but seemingly caused by very different mechanisms! One may ask whether the physics laws are different, depending on the frame of observation.

A quantitative form can be obtained by applying the Galileo transformation to the description of an electromagnetic wave. Maxwell describes light as waves; the wave equation reads:

$$\left(\frac{\partial^2}{\partial x'^2} + \frac{\partial^2}{\partial y'^2} + \frac{\partial^2}{\partial z'^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t'^2}\right)\Psi = 0\,. \tag{9}$$

Applying the Galileo transformation ($x = x' - vt, y' = y, z' = z, t' = t$), we get the wave equation in the moving frame:

$$\left(\left[1 - \frac{v^2}{c^2}\right]\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \frac{2v}{c^2}\frac{\partial^2}{\partial x \partial t} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\Psi = 0\,. \tag{10}$$

The form of the transformed equation is rather different in the two frames.

The Maxwell equations are not compatible with the Galileo transformation.

3 Special relativity

To solve this riddle, one can consider three possible options.

1. Maxwell's equations are wrong and should be modified to be invariant with Galileo's relativity (unlikely).

2. Galilean relativity applies to classical mechanics, but not to electromagnetic effects and light has a reference frame (ether). Was defended by many people, sometimes with obscure concepts…
3. A relativity principle different from Galileo for *both* classical mechanics and electrodynamics (requires modification of the laws of classical mechanics).

Against all odds and with the disbelief of his colleagues, Einstein chose the last option.

3.1 Postulate for special relativity

To arrive at the new formulation of relativity, Einstein introduced three postulates.

- All physical laws in inertial frames must have equivalent forms.
- The speed of light in a vacuum c must be the same in all frames.
- It requires a transformations (not Galilean) that makes *all* physics laws look the same.

3.2 Lorentz transformation

The transformation requires that the co-ordinates must be transformed differently, satisfying the three postulates.

Writing the equations for the front of a moving light wave in F and F':

$$F : x^2 + y^2 + z^2 - c^2t^2 = 0\,, \tag{11}$$

$$F' : x'^2 + y'^2 + z'^2 - c'^2t'^2 = 0\,. \tag{12}$$

The constant speed of light requires $c = c'$ in both equations. This leads to a set of equations known as the Lorentz transformation (Eq. (13)).

$$\begin{aligned}
x' &= \frac{x - vt}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} = \gamma \cdot (x - vt)\,, \\
y' &= y\,, \\
z' &= z\,, \\
t' &= \frac{t - \frac{v \cdot x}{c^2}}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} = \gamma \cdot \left(t - \frac{v \cdot x}{c^2}\right)\,.
\end{aligned} \tag{13}$$

The main difference from the Galileo transformation is that it requires a transformation of the time t. It is a direct consequence of the required constancy of the speed of light. This tightly couples the position and time and they have to be treated on equal footing.

It is common practice to introduce the relativistic variables γ and β_r:

$$\gamma = \frac{1}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} = \frac{1}{\sqrt{(1 - \beta_\mathrm{r}^2)}}\,, \tag{14}$$

where β_r is:

$$\beta_\mathrm{r} = \frac{v}{c}\,. \tag{15}$$

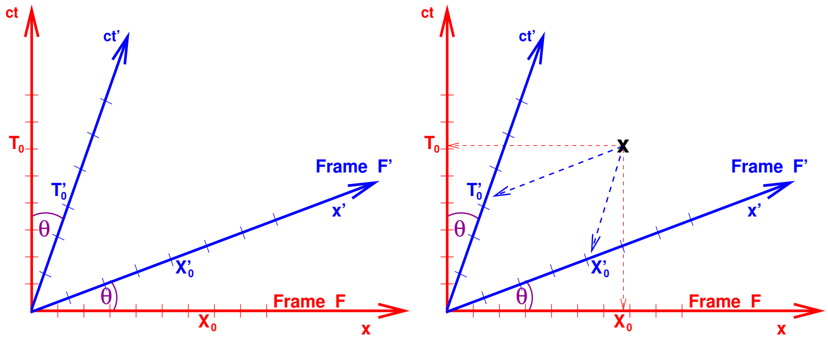


Fig. 5: The Lorentz transformation between frame F and F'. This representation is known as a Minkowski diagram.

3.3 Minkowski diagram—pictorial representation of the Lorentz transformation

An illustration of the Lorentz transformation is shown in Fig. 5. Starting from the orthogonal reference frame and using the transformation of position *and* time, both axes of the new reference system appear tilted, where the tilt angle depends on the velocity of the moving frame:

$$\tan\theta = \frac{v}{c} = \beta\,. \tag{16}$$

The position and time in the two reference frames can easily be obtained by the projection of an event onto the axes of the two frames (Fig. 5, right-hand side).

Contrary to normal (i.e., circular) rotation, where the axes remain perpendicular to each other, this type of rotation is also known as hyperbolic rotation. To quantify such a rotation, another angle ψ is introduced as:

$$\tanh\psi = \frac{v}{c} = \beta\,. \tag{17}$$

This angle ψ is also known as the rapidity. As a consequence we have:

$$\cosh\psi = \gamma \tag{18}$$

and

$$\sinh\psi = \gamma\beta\,. \tag{19}$$

Some applications become easier using this formulation.

3.4 Transformation of velocities

We assume a frame F' moving with constant speed of $\vec{v} = (v, 0, 0)$ relative to frame F. An object inside the moving frame is assumed to move with $\vec{v}' = (v'_x, v'_y, v'_z)$.

The velocity $\vec{v} = (v_x, v_y, v_z)$ of the object in the frame F is computed using the Lorentz transformation (Eq. (13))

$$v_x = \frac{v'_x + v}{1 + \frac{v'_x v}{c^2}}\,, \qquad v_y = \frac{v'_y}{\gamma\left(1 + \frac{v'_x v}{c^2}\right)}\,, \qquad v_z = \frac{v'_z}{\gamma\left(1 + \frac{v'_x v}{c^2}\right)}\,. \tag{20}$$

Adding two speeds v_1 and v_2:

$$v = v_1 + v_2 \qquad \Rightarrow \qquad v = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}\,. \tag{21}$$

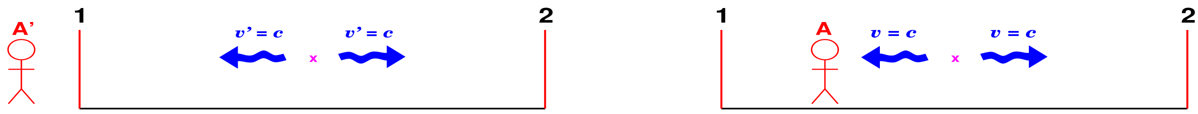


Fig. 6: Flash of light emitted in a resting frame, observed by two observers within and outside the frame

From (Eq. (21)) it can easily be seen that the speed of light can never be exceeded, in agreement with the second postulate.

An interesting result can be observed using the rapidity for this calculation. Since we have $\tanh\psi = \beta = v/c$, we can reformulate Eq. (21) as:

$$\tanh\psi = \frac{\tanh\psi_1 + \tanh\psi_2}{1 + \tanh\psi_1 \tanh\psi_2} = \tanh(\psi_1 + \psi_2)\,, \tag{22}$$

i.e., the rapidities can be added.

4 Consequences of special relativity

The use of the Lorentz transformation between two inertial frames and the required transformation of position *and* time has very significant consequences, which are rather counterintuitive.

- Space and time are *not* independent quantities.
- There is no absolute time and space, no absolute motion.
- Relativistic phenomena (with relevance for accelerators):
  - no speed of moving objects can exceed the speed of light;
  - (non-)simultaneity of events in independent frames;
  - Lorentz contraction;
  - time dilation;
  - relativistic Doppler effect;
  - Lorentz force.
- A formalism with four-vectors is introduced.
- Electrodynamics becomes very simple and consistent.

4.1 Simultaneity in special relativity

Assume that two events in frame F at (different) positions x_1 and x_2 happen simultaneously at times $t_1 = t_2$.

The times t'_1 and t'_2 in F' are obtained from the Lorentz transformation and

$$t'_1 = \gamma \cdot \left(t_1 - \frac{v \cdot x_1}{c^2}\right) \qquad \text{and} \qquad t'_2 = \gamma \cdot \left(t_2 - \frac{v \cdot x_2}{c^2}\right)\,. \tag{23}$$

One finds the surprising result that two events that are simultaneous at *different positions* x_1 and x_2 in F are not simultaneous in F': $x_1 \neq x_2$ in F implies that $t'_1 \neq t'_2$ in frame F'!

Assume the sequence of events depicted in Figs. 6–8. In a resting frame, a flash of light is emitted in the centre of the frame towards two detectors. An observer within the frame and another outside the frame observe the flash of light arriving simultaneously at detectors 1 and 2.

If the frame is moving, the detectors are reached at different times for the observer outside the frame.

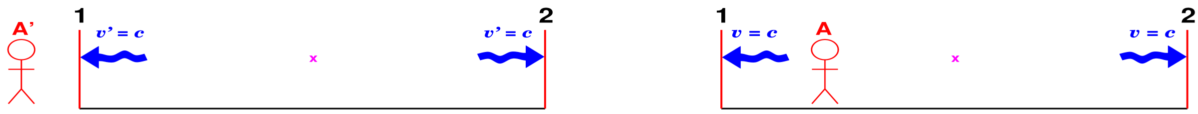


Fig. 7: For both observers, the flash of light reaches detectors 1 and 2 simultaneously

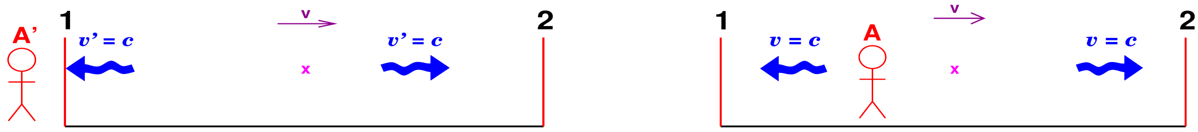


Fig. 8: Emitted in a moving frame, the flash reaches the detectors at different times for the outside observer

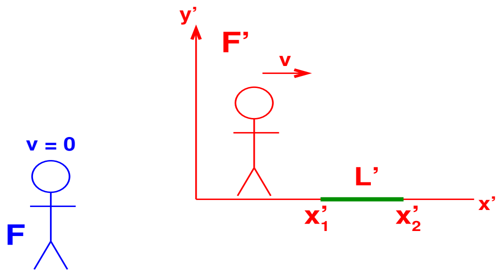


Fig. 9: Measuring the length of an object in a moving and a rest frame

Why should we bother about simultaneity?

- Simultaneity is not frame independent.
- It plays a pivotal role in special relativity.
- Almost all paradoxes are explained by it!
- Different observers see a different reality; in particular, the sequence of events can change!

For $t_1 < t_2$, we may find (not always!) a frame where $t_1 > t_2$ (the concept of before and after depends on the observer).

4.2 Length contraction

To measure the length of an object (Fig. 9), the procedure is to measure the position at *both* ends *simultaneously!*

The measured length of a rod in F' is the difference between the two positions $L' = x'_2 - x'_1$, measured simultaneously at a fixed time t' in frame F'.

From the frame F, we follow the same procedure, i.e., we have to measure simultaneously (!) the ends of the rod at a fixed time t in frame F, i.e.,:

$$L = x_2 - x_1 \,.$$

We therefore make a Lorentz transformation of the 'rod co-ordinates' into the rest frame:

$$x'_1 = \gamma \cdot (x_1 - vt) \qquad \text{and} \qquad x'_2 = \gamma \cdot (x_2 - vt)\,, \tag{24}$$

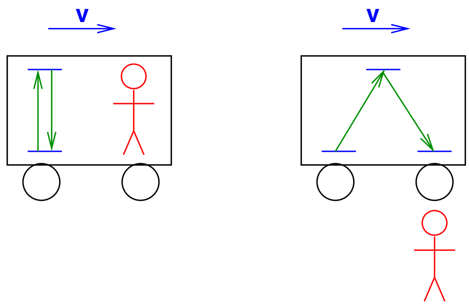


Fig. 10: Path of light between two mirrors in (left) a moving frame and (right) a rest frame

$$L' = x'_2 - x'_1 = \gamma \cdot (x_2 - x_1) = \gamma \cdot L\,. \tag{25}$$

We obtain

$$L = L'/\gamma\,. \tag{26}$$

The length appears contracted in the rest frame (Eq. (26)), depending on the relative velocity, i.e., the relativistic γ.

In accelerators, this has consequences, as objects in the frame of the particle and the frame of the accelerator appear to have different lengths. This is of particular importance for bunch length, electromagnetic fields, magnets, and the distances between magnets and other objects in the accelerator.

4.3 Time dilation

Applying the Lorentz transformation, we have to transform the time t as well as the position.

We assume a moving frame where a flash of light is moving upwards and reflected downwards by a mirror (Fig. 10).

We assume that the frame moves with velocity v. Seen from outside, the flash arrives at the mirror, but at a different position. This means that the apparent total path is longer, but c must be the same. The geometry of this process is shown in Fig. 11.

In frame F': light travels distance L in time $\Delta t'$.
In frame F: light travels distance D in time $\Delta t'$;
the entire system moves distance d in time Δt.

We look at the trajectories in the two frames and simple calculation leads to the result that the time needed by the flash is longer by the factor γ when it is observed from the outside. After a round trip, the full distance observed in the moving frame is $2 \cdot L$; measured from the outside, it is $2 \cdot D$.

$$L = c \cdot \Delta t'\,, \qquad D = c \cdot \Delta t\,, \qquad d = v \cdot \Delta t\,,$$
$$(c \cdot \Delta t)^2 = (c \cdot \Delta t')^2 + (v \cdot \Delta t)^2\,.$$

We obtain:

$$\rightarrow \Delta t = \gamma \cdot \Delta t'\,. \tag{27}$$

4.4 Proper time and proper length

This derivation can lead to some confusion.

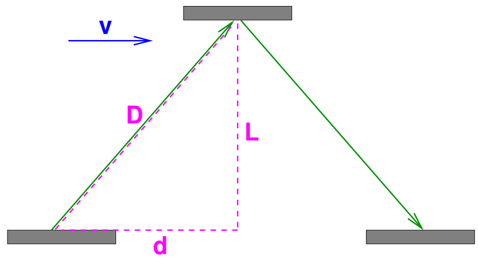


Fig. 11: Path of light between two mirrors, as observed from a rest frame

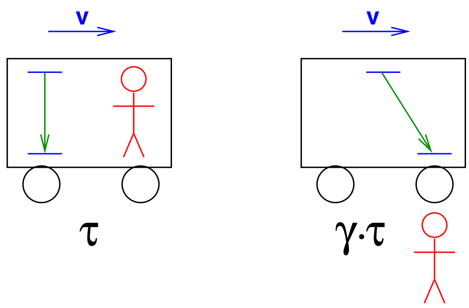


Fig. 12: Time measured within the moving frame (left-hand side) and from the rest frame (right-hand side). In the moving frame, the measured time is always the proper time τ, independent of the velocity of the moving frame.

- The car is moving: $\Delta t = \gamma \cdot \Delta t'$.
- The observer is moving: $\Delta t' = \gamma \cdot \Delta t$.

This seems like a contradiction. This paradox is solved by introducing the concept of proper time τ. The proper time τ is the time measured by the observer at rest relative to the process.

Or: the proper time for a given observer is measured by the clock that travels with the observer:

$$c^2 \Delta\tau^2 = c^2 \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \,. \tag{28}$$

Equation (28) defines the proper time τ.

A similar argument holds for the Lorentz contraction and we define the 'proper length' in the same way.

We illustrate this in Fig. 12, where the time is measured within the frame as the proper time τ. From outside, the measured time is $\gamma \cdot \tau$.

To summarize, it is found that time and distances are relative.

- τ is a fundamental time: the proper time τ.
- The proper time is the time measured by an observer in its own frame.
- From frames moving *relative* to it, the time appears longer.
- $\mathcal{L}$ is a fundamental length: the proper length $\mathcal{L}$.
- The proper length is the length measured by an observer in its own frame.
- From frames moving *relative* to it, the length appears shorter.

4.5 Proper acceleration

Although not very much used in accelerator physics, one has a proper acceleration. It can be computed as the derivative of the rapidity with respect to the proper time:

$$\alpha = \frac{\mathrm{d}\psi}{\mathrm{d}\tau}\,.$$

4.6 Relativistic space travel

The formula for time dilation also holds for an accelerated system! Assuming a spaceship starting from Earth and moving with a constant acceleration, g = 9.81 m/s^2. We denote the time on Earth t and the proper time in the spaceship t_p. For β, we then have:

$$\beta = \tanh\left(\frac{v}{c}\right) = \tanh\left(\frac{g \cdot t_\mathrm{p}}{c}\right)\,, \tag{29}$$

and therefore for γ:

$$\gamma = \cosh\left(\frac{g \cdot t_\mathrm{p}}{c}\right) = \sqrt{1 + \left(\frac{g \cdot t_\mathrm{p}}{c}\right)^2}\,. \tag{30}$$

Finally, for the distance to the start (e.g., Earth), we obtain:

$$d = \left(\frac{c^2}{g}\right) \cdot \left[\cosh\left(\frac{g \cdot t_\mathrm{p}}{c}\right) - 1\right]\,. \tag{31}$$

When 12 years have passed for the passenger of the spaceship, the distance to Earth is $d \approx 120\,000$ light years! This is approximately the diameter of the Milky Way!

4.7 Muon lifetime

A popular example is the lifetime of a μ particle moving at high speed. We can compute the (observed) lifetime of the muon.

In the frame of the muon, the lifetime is $\tau \approx 2 \cdot 10^{-6}$ s. Measured from the laboratory frame, the time $\gamma \cdot \tau$ is observed. The muon appears to have a longer lifetime, in contradiction with the postulates.

A clock in the muon frame shows the *proper time* and the muon decays in $\approx 2 \cdot 10^{-6}$ s, independent of the muon's speed.

Seen from the lab frame, the muon lives γ times longer. In the muon storage ring at CERN, the lifetime of muons circulating with γ = 29.327 was found to be dilated to 64.378 µs, confirming time dilation.

4.8 Moving electron

Important effects are experienced by a fast-moving electron in an accelerator. We assume: $v \approx c$.

Bunch length:
In lab frame: σ_z, In frame of electron: $\gamma \cdot \sigma_z$.

Length of an object (e.g., magnet, *distance* between magnets!):
In lab frame: L, In frame of electron: L/γ.

4.9 Relativistic Doppler effect

A rather important relativistic effect is the Doppler shift of the frequency of a fast-moving particle. In light sources, this effect is most significant. Unlike sound, light does not have a medium of propagation and the nature of this effect is purely relativistic.

The frequency observed ν depends on the velocity γ and the observation angle θ. The frequency of the light to an observer looking against the direction of motion is increased by a factor γ; at high speeds, this is a very significant effect:

$$\nu = \nu_0 \cdot \gamma \cdot (1 - \beta_r \cdot \cos(\theta)) \,. \qquad (32)$$

4.10 Everyday effects

Although the speeds we experience in everyday life are small compared with the speed of light, time dilation has to be taken into account for some applications.

4.10.1 Intercontinental flights

As the time for moving objects is passing more slowly with respect to a reference on Earth, one expects that during an eastward intercontinental flight the passengers age more slowly. This was tested and confirmed experimentally with atomic clocks in 1971 and 1977 [3].

For a 6 hour flight eastwards with a regular aeroplane cruising at $\approx$900 km/h, one computes a difference of 25–30 ns. This can easily be measured [3] but has little effect in everyday life. Taking into account the effects of general relativity, the effect is very different (in the opposite direction, but still small). General relativity is not the topic of this lecture, but the relevant equations are given in Appendix A for the gifted reader.

4.10.2 GPS

Contrary to flights with an aeroplane, other everyday effects experience very strong relativistic effects. A prominent example is the global positioning system (GPS). The flight parameters of the satellites of the system and the effects are:

– Orbital speed, i.e., relative to a reference on Earth is 14000 km/h $\approx$ 3.9 km/s

$$\Rightarrow \beta \approx 1.3 \cdot 10^{-5} \,, \qquad \gamma \approx 1.000000000084 \,.$$

– This is very small, but it accumulates 7 μs during one day compared with the reference time on Earth!
– After one day, the position is wrong by $\approx$2 km!
– Including general relativity, the error is as large as 10 km per day; the computation is given in Appendix A.

Without corrections for the effects of special and general relativity, the satellite navigation cannot work. Countermeasures:

– A minimum of four satellites is used (avoid reference time on Earth);
– The data transmission frequency is reduced from 1.023 MHz to 1.022999999543 MHz prior to launch.

5 Relativistic mass and momentum

After one has established the relativistic corrections for position and time, it is necessary to evaluate the consequences for momentum, energy, and mass of a moving particle.

5.1 Consequences of momentum conservation

To simplify the computation, we assume that an object moves in frame F' with $\vec{u}' = (0, u'_y, 0)$. To have an invariant form, one requires that the expression

$$\vec{F} = m \cdot \vec{a} = m \cdot \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} \tag{33}$$

has the same form in all frames.

The transverse momentum must be conserved; this demands:

$$\begin{aligned} p_y &= p'_y\,, \\ m u_y &= m' u'_y\,, \\ m u'_y/\gamma &= m' u'_y\,. \end{aligned} \tag{34}$$

This implies that

$$m = \gamma m'\,. \tag{35}$$

As a consequence of momentum conservation, *mass must also be transformed*! Using the expression for the mass m (using $m' = m_0$):

$$m = m_0\sqrt{1 - \left(\frac{v}{c}\right)^2} = \gamma \cdot m_0\,, \tag{36}$$

and expand it for small speeds:

$$m \cong m_0 + \frac{1}{2} m_0 v^2 \left(\frac{1}{c^2}\right)\;; \tag{37}$$

multiplied by c^2:

$$mc^2 \cong m_0 c^2 + \frac{1}{2} m_0 v^2 = m_0 c^2 + T\,. \tag{38}$$

The second term is the kinetic energy T :

$$E = mc^2 = m_0 c^2 + T\,. \tag{39}$$

5.2 Interpretation of relativistic mass and energy

- The total energy E is $E = mc^2$.
- It is the sum of the kinetic energy and the rest energy.
- The energy of a particle in its rest frame is $E_0 = m_0 c^2$.

Using the definition of the relativistic mass $m = \gamma m_0$, we can write:

$$E = m \cdot c^2 = \gamma m_0 \cdot c^2\,. \tag{40}$$

For any object, $m \cdot c^2$ is the total energy, and this follows directly from momentum conservation. We can say that m is the mass (energy) of the object 'in motion', while m_0 is the mass (energy) of the object 'at rest'. The mass m is not the same in all inertial systems; the *rest mass* m_0 is. Following previous arguments, m_0 is the 'proper mass'.

5.3 Relativistic relations

Starting from

$$p = mv\,, \tag{41}$$

with $m = \gamma m_0$:

$$p = \gamma \cdot m_0 v = \gamma \cdot \beta \cdot c \cdot m_0\,. \tag{42}$$

We rewrite:

$$E^2 = m^2c^4 = \gamma^2 m_0^2 c^4 = (1 + \gamma^2\beta^2) m_0^2 c^4\,, \tag{43}$$

and finally get:

$$E^2 = (m_0c^2)^2 + (pc)^2 \qquad \Rightarrow \qquad \frac{E}{c} = \sqrt{(m_0c)^2 + p^2}\,. \tag{44}$$

This is a rather important formula in practice, e.g., for kinematics in accelerators.

5.4 Units used in special relativity and particle physics

Standard SI units are not very convenient; other units are much easier to use and have become a standard in accelerator and particle physics:

$$[E] = \text{ eV}\,, \qquad [p] = \text{ eV}/c\,, \qquad [m] = \text{ eV}/c^2\,.$$

Then one can use a very convenient form for Eq. (44):

$$E^2 = m_0^2 + p^2 m\,. \tag{45}$$

Examples for elementary particles (proton):
Mass of a proton: $m_\mathrm{p} = 1.672 \cdot 10^{-27}$ kg,
Energy(at rest): $m_\mathrm{p}c^2 = 938$ MeV $= 0.15$ nJ.

We can take an everyday object like a ping-pong ball:
Ping-pong ball: $m_\mathrm{pp} = 2.7 \cdot 10^{-3}$ kg ($\approx 1.6 \cdot 10^{24}$ protons),
Energy (at rest): $m_\mathrm{pp}c^2 = 1.5 \cdot 10^{27}$ MeV $= 2.4 \cdot 10^{14}$ J.
This corresponds to an energy of:

- $\approx$750 000 times the full LHC beam;
- $\approx$60 kilotons of TNT.

The typical kinetic energy of a ping-pong ball is completely negligible.

5.5 Mass of a moving particle in a particle accelerator

The mass of a fast-moving particle increases as:

$$m = \gamma m_0 = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}\,. \tag{46}$$

When we accelerate:

- for $v \ll c$:
  - E, m, p, v increase...
- for $v \approx c$:
  - E, m, p increase, but v (almost) does not!

Table 1: Relations between relativistic parameters

	cp	T	E	γ
β	$\frac{1}{\sqrt{\left(\frac{E_0}{cp}\right)^2+1}}$	$\sqrt{1-\frac{1}{\left(1+\frac{T}{E_0}\right)^2}}$	$\sqrt{1-\left(\frac{E_0}{E}\right)^2}$	$\sqrt{1-\gamma^{-2}}$
cp	cp	$\sqrt{T(2E_0+T)}$	$\sqrt{E^2-E_0^2}$	$E_0\sqrt{\gamma^2-1}$
E_0	$\frac{cp}{\sqrt{\gamma^2-1}}$	$T/(\gamma-1)$	$\sqrt{E^2-c^2p^2}$	E/γ
T	$cp\sqrt{\frac{\gamma-1}{\gamma+1}}$	T	$E-E_0$	$E_0(\gamma-1)$
γ	$cp/E_0\beta$	$1+T/E_0$	E/E_0	γ

Table 2: Logarithmic relations between relativistic parameters

	$\frac{\mathrm{d}\beta}{\beta}$	$\frac{\mathrm{d}p}{p}$	$\frac{\mathrm{d}T}{T}$	$\frac{\mathrm{d}E}{E}=\frac{\mathrm{d}\gamma}{\gamma}$
$\frac{\mathrm{d}\beta}{\beta}$	$\frac{\mathrm{d}\beta}{\beta}$	$\frac{1}{\gamma^2}\frac{\mathrm{d}p}{p}$	$\frac{1}{\gamma(\gamma+1)}\frac{\mathrm{d}T}{T}$	$\frac{1}{(\beta\gamma)^2}\frac{\mathrm{d}\gamma}{\gamma}$
$\frac{\mathrm{d}p}{p}$	$\gamma^2\frac{\mathrm{d}\beta}{\beta}$	$\frac{\mathrm{d}p}{p}$	$[\gamma/(\gamma+1)]\frac{\mathrm{d}T}{T}$	$\frac{1}{\beta^2}\frac{\mathrm{d}\gamma}{\gamma}$
$\frac{\mathrm{d}T}{T}$	$\gamma(\gamma+1)\frac{\mathrm{d}\beta}{\beta}$	$\left(1+\frac{1}{\gamma}\right)\frac{\mathrm{d}p}{p}$	$\frac{\mathrm{d}T}{T}$	$\frac{\gamma}{(\gamma-1)}\frac{\mathrm{d}\gamma}{\gamma}$
$\frac{\mathrm{d}E}{E}$	$(\beta\gamma)^2\frac{\mathrm{d}\beta}{\beta}$	$\beta^2\frac{\mathrm{d}p}{p}$	$\left(1-\frac{1}{\gamma}\right)\frac{\mathrm{d}T}{T}$	$\frac{\mathrm{d}\gamma}{\gamma}$
$\frac{\mathrm{d}\gamma}{\gamma}$	$(\gamma^2-1)\frac{\mathrm{d}\beta}{\beta}$	$\frac{\mathrm{d}p}{p}-\frac{\mathrm{d}\beta}{\beta}$	$\left(1-\frac{1}{\gamma}\right)\frac{\mathrm{d}T}{T}$	$\frac{\mathrm{d}\gamma}{\gamma}$

$$\beta = \frac{v}{c} \approx \sqrt{1-\frac{m_0^2c^4}{T^2}}\,. \tag{47}$$

A consequence of this effect is that, in a synchrotron, a particle with higher speed takes longer to complete a turn than a particle with lower speed. This leads to the effect of transition in synchrotrons.

For small rest masses (e.g., electrons), this effect is very small; the effect is only relevant for hadrons.

5.6 Kinematic relations

A summary of kinematic relations between the relativistic variables is shown in Tables 1 and 2.

As an example, one can compute the relative spread of particle velocities from the momentum spread:

$$\text{LHC (7 TeV):}\qquad \frac{\Delta p}{p} = 10^{-4}\,,\qquad \frac{\Delta\beta}{\beta} = 10^{-12}\,,$$
$$\text{LEP (100 GeV):}\qquad \frac{\Delta p}{p} = 10^{-4}\,,\qquad \frac{\Delta\beta}{\beta} = 10^{-15}\,.$$

The reason for the much smaller velocity spread in the LEP is, of course, the larger γ factor.

6 First summary

The theory of special relativity is derived from two postulates.

- Physics laws are the same in all inertial frames.
- The speed of light in vacuum c is the same in all frames and requires Lorentz transformation.

Consequences of special relativity are:

- simultaneity is not independent of the frame of observation;
- moving objects *appear* shorter;
- moving clocks *appear* to go slower;
- the mass is not independent of motion ($m = \gamma \cdot m_0$) and the total energy is $E = m \cdot c^2$;
- absolute space or time do not exist: where it happens and when it happens are not independent.

In the following, it is demonstrated how to simplify calculations using the concept of four-vectors.

7 Space–time and four-vectors

Since space and time are not independent, we must reformulate the physics laws, taking both into account on an equal footing:

$$t, \qquad \vec{x} = (x, y, z) \qquad \Rightarrow \qquad \text{Replace by one vector including the time.} \tag{48}$$

All four-vectors are constructed from a temporal and a spatial part, where the temporal part (e.g., time t) is multiplied by c to get the same units.

We need two types of four-vector (here *position four-vector*):

$$X^{\mu} = (ct, x, y, z) \qquad \text{and} \qquad X_{\mu} = (ct, -x, -y, -z)\,. \tag{49}$$

This is due to the 'skewed' reference system; for details see the bibliography.

Four-vectors of the type X^{μ} are called *contravariant* vectors; X_{μ} are called *covariant* vectors. It is common practice to use capital letters for four-vectors to distinguish from the 3D vectors. Using four-vectors, the Lorentz transformation can easily be written in matrix form:

$$X'^{\mu} = \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = X^{\mu}\,, \tag{50}$$

$$X'^{\mu} = \Lambda \circ X^{\mu} \qquad (\Lambda \text{ for 'Lorentz'})\,, \tag{51}$$

but note:

$$X'_{\mu} = \begin{pmatrix} ct' \\ -x' \\ -y' \\ -z' \end{pmatrix} = \begin{pmatrix} \gamma & +\gamma\beta & 0 & 0 \\ +\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ -x \\ -y \\ -z \end{pmatrix} = X_{\mu}\,. \tag{52}$$

It can be verified that the transformation for covariant vectors is the inverse of Λ and:

$$X'_{\mu} = \Lambda^{-1} \circ X_{\mu}\,. \tag{53}$$

7.1 Hyperbolic transformation

Using the definitions from Eq. (17), the transformation becomes:

$$X'^{\mu} = \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(\psi) & -\sinh(\psi) & 0 & 0 \\ -\sinh(\psi) & \cosh(\psi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = X^{\mu} . \tag{54}$$

This takes the form of a rotation.

7.2 Vector operations

Having introduced four-vectors, we have to look at operations on and with four-vectors, in particular scalar products, because they play a key role in special relativity.

7.3 Scalar product

The well-known scalar product using Cartesian co-ordinates and the Euclidean metric is:

$$\vec{x} \cdot \vec{y} = (x_a, y_a, z_a) \cdot (x_b, y_b, z_b) = (x_a \cdot x_b + y_a \cdot y_b + z_a \cdot z_b) . \tag{55}$$

Space–time four-vectors are of the form:

$$A^{\mu} = (ct_a, x_a, y_a, z_a) , \qquad B_{\mu} = (ct_b, -x_b, -y_b, -z_b) .$$

The four-vector scalar product is:

$$A^{\mu} B_{\mu} = \underbrace{\sum_{\mu=0}^{3} A^{\mu} B_{\mu}}_{\text{Einstein convention}} = (ct_a \cdot ct_b - x_a \cdot x_b - y_a \cdot y_b - z_a \cdot z_b) . \tag{56}$$

In Eq. (56), we have used the so-called *Einstein convention* to write the equations in a more compact form: when an index appears more than once in an equation (here: μ) it implies summation over this index.

For many applications, a simplified rule can be used:

$$AB = (ct_a \cdot ct_b - x_a \cdot x_b - y_a \cdot y_b - z_a \cdot z_b) . \tag{57}$$

7.4 Four-vectors

We have important four-vectors:

Co-ordinates: $X^{\mu} = (ct, x, y, z) = (ct, \vec{x}) ,$

Velocities: $U^{\mu} = \dfrac{\mathrm{d}X^{\mu}}{\mathrm{d}\tau} = \gamma(c, \dot{\vec{x}}) = \gamma(c, \vec{u}) ,$

Momenta: $P^{\mu} = mU^{\mu} = m\gamma(c, \vec{u}) = \gamma(mc, \vec{p}) ,$

Force: $F^{\mu} = \dfrac{\mathrm{d}P^{\mu}}{\mathrm{d}\tau} = \gamma \dfrac{\mathrm{d}}{\mathrm{d}\tau}(mc, \vec{p}) ,$

Wave propagation vector: $K^{\mu} = \left(\dfrac{\omega}{c}, \vec{k}\right) ,$

We define the gradient operator as a four-vector: $\partial^{\mu} = \left(\dfrac{1}{c}\dfrac{\partial}{\partial t}, -\vec{\nabla}\right) = \left(\dfrac{1}{c}\dfrac{\partial}{\partial t}, -\dfrac{\partial}{\partial x}, -\dfrac{\partial}{\partial y}, -\dfrac{\partial}{\partial z}\right) .$

It is a key element in this analysis that *all* four-vectors A^{μ} transform as:

$$A'^{\mu} = \Lambda \circ A^{\mu} . \tag{58}$$

7.5 Invariant forms

The main objective of the principle of special relativity is to arrive at invariant laws of physics in different frames.

- The solution: write the laws of physics in terms of four-vectors and use Lorentz transformations between the frames.
- Without proof (it is rather straightforward using Eqs. (51) and (52)): *any* four-vector (scalar) product $Z^\mu Z_\mu$ has the same value in all inertial frames:

$$Z^\mu Z_\mu = Z'^\mu Z'_\mu \,. \tag{59}$$

All scalar products of any four-vectors are invariant!

but: $Z^\mu Z^\mu$ and $Z'_\mu Z'_\mu$ are not!

- Note: the four-vectors in the scalar product do not have to be of the same type.
- $P^\mu X_\mu$ is also an invariant four-vector product.

One should look at two particularly important invariants.

7.5.1 A special invariant—invariant velocity

From the velocity four-vector U^μ:

$$U^\mu = \gamma(c, \vec{u}) \,, \tag{60}$$

we get the scalar product:

$$U^\mu U_\mu = \gamma^2(c^2 - \vec{u}^2) = c^2 \,. \tag{61}$$

We find that the invariant of the velocity has the same value in all inertial frames and is the speed of light. The speed of light c is the same in all frames.

7.5.2 Invariant momentum

Starting from the four-momentum P^μ:

$$P^\mu = m_0 U^\mu = m_0\gamma(c, \vec{u}) = (mc, \vec{p}) = \left(\frac{E}{c}, \vec{p}\right) \,, \tag{62}$$

$$P'^\mu = m_0 U'^\mu = m_0\gamma(c, \vec{u'}) = (mc, \vec{p'}) = \left(\frac{E'}{c}, \vec{p'}\right) \,. \tag{63}$$

We can get another important invariant:

$$P^\mu P_\mu = P'^\mu P'_\mu = m_0^2 c^2 \,. \tag{64}$$

The invariant of the four-momentum vector is the mass m_0. It follows that the rest mass is the same in all frames (it has to be, otherwise we could not tell whether we are moving or not!)

7.6 Four-vectors and kinematics

The use of four-vectors enables a very straightforward and simple procedure to compute the kinematics of moving particles.

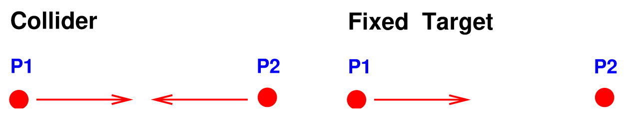


Fig. 13: Schematic illustration of particle collisions. Left: for the case of a colliding beam facility; right: when one particle is at rest (fixed-target collisions).

Table 3: Centre of mass energies for fixed-target collisions and colliding beams for different energies and particle types.

Collision	E **beam energy**	$E_{\rm cm}$ **(collider)**	$E_{\rm cm}$ **(fixed target)**
pp	315 (GeV)	630 (GeV)	24.3 (GeV)
pp	6500 (GeV)	13000 (GeV)	110.4 (GeV)
pp	90 (PeV)	180 (PeV)	13000 (GeV)
e^+e^-	100 (GeV)	200 (GeV)	0.320 (GeV)

7.6.1 Particle collisions

For collisions of particles in high energy physics experiments, the available centre of mass energy $E_{\rm cm}$ is the relevant parameter.

We distinguish two types of collision: when a particle hits a particle at rest and when two high energy particles collide, either head on or at an angle. This is shown in Fig. 13.

The computation of $E_{\rm cm}$ is rather straightforward when four-vectors are used. The centre of mass energy must be an invariant of the collision process and is therefore the scalar product of two four-vectors.

It was shown before that the relevant four-vector is the momentum four-vector. Taking the scalar product of the sum of the four-momentum of the colliding particles gives the centre of mass energy.

$$\begin{aligned}
P_1^\mu &= (E,\vec{p})\,, \qquad P_2^\mu = (E,-\vec{p})\,, & P_1^\mu &= (E,\vec{p})\,, \qquad P_2^\mu = (m_0,0)\,,\\
P^\mu &= P_1^\mu + P_2^\mu = (2E,0)\,, & P^\mu &= P_1^\mu + P_2^\mu = (E+m_0,\vec{p})\,,\\
E_{\rm cm} &= \sqrt{P^\mu P_\mu} = 2\cdot E\,, & E_{\rm cm} &= \sqrt{P^\mu P_\mu} = \sqrt{2m_0 E}\,.
\end{aligned} \tag{65}$$

This procedure also works for more than two colliding particles: $P^\mu = P_1^\mu + P_2^\mu + P_3^\mu + \ldots$ It works for any configuration; also, for particle decay.

In Table 3, a comparison is made of the centre of mass energies for fixed-target collisions and colliding beams for different energies and particle types. The differences are very significant; centre of mass energies as provided by the LHC cannot be reached by fixed-target collisions.

7.6.2 Particle decay

We consider a particle P_0 decaying into two (or more) particles (Fig. 14): $P_0 \Rightarrow P_1 + P_2$.

We can measure the properties of the decay products, (i.e., $\vec{p_2}$, $\vec{p_1}$, m_1, m_2, E_1, E_2). The parameters (in particular the mass m_0) of the original particle are unknown.

For every decay product, we construct the corresponding four-momentum from the measured values:

$$\begin{aligned}
P_1^\mu &= (E_1\,,\vec{p_1}), & E_1 &= \sqrt{m_1^2 + \vec{p_1^2}}\,,\\
P_2^\mu &= (E_2\,,\vec{p_2}), & E_2 &= \sqrt{m_2^2 + \vec{p_2^2}}\,.
\end{aligned} \tag{66}$$

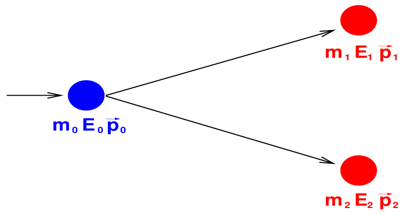


Fig. 14: A particle decaying into two particles

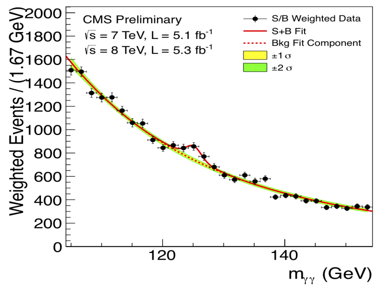


Fig. 15: Invariant mass of a $\gamma\gamma$ decay, showing an enhancement at the Higgs mass

The centre of mass of the decaying particle is an invariant: the rest mass m_0 of the particle.

We obtain the *sum* of the four-momenta, which is the four-momentum of the original particle:

$$P_0^\mu = (P_1^\mu + P_2^\mu) = (E_1 + E_2, \vec{p_1} + \vec{p_2})\,. \tag{67}$$

The particle mass we get from the scalar product and, since we know $P_0^2 = m_0^2 c^4$:

$$m_0 c^2 = \sqrt{(E_1 + E_2)^2 - (\vec{p_1}c + \vec{p_2}c)^2}\,. \tag{68}$$

When we draw m_0 for every observed decay, we obtain a histogram with the mass of the original particle (Fig. 15).

As an example, Fig. 15 shows the invariant mass of two photons, an expected channel for the decay of the Higgs particle. At the mass of the Higgs particle, we see an enhancement above the background. This procedure works for any number of decay products.

7.6.3 Particle scattering and four-vectors

In theoretical physics, the Mandelstam variables are numerical quantities that encode the energy, momentum, and angles of particles in a scattering process in a Lorentz-invariant fashion. They are used for scattering processes of two incoming particles and two outgoing particles.

Assume that there are two particles (P_1 and P_2) in the initial state of an interaction and two particles (P_3 and P_4) after the interaction. The two most important relations are described by the variables s and t, defined as:

$$s = (P_1 + P_2)^2 = (P_3 + P_4)^2\,, \tag{69}$$
$$t = (P_1 - P_3)^2 = (P_2 - P_4)^2\,. \tag{70}$$

The physical interpretation of these variables is rather straightforward. Compared with earlier results, the variable $s = (P_1 + P_2)^2$ is the square of the centre of mass energy in the collision process (Fig. 16).

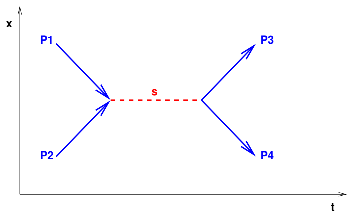


Fig. 16: Invariant variable s for colliding particles

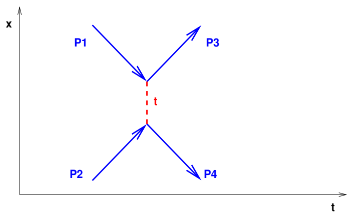


Fig. 17: Invariant variable t for colliding particles

The variable $t = (P_1 - P_3)^2$ describes an interaction where an incoming particle is scattered from another particle (Fig. 17), exchanging momentum t, typically through the exchange of an intermediate particle (vector boson).

7.6.4 Cross-sections and luminosity

The probability for a collision process between particles is characterized by the corresponding cross-section for this process.

According to the postulates, the cross-section must be an invariant in all frames. This requires the correction term from Eq. (71):

$$K = \sqrt{(\vec{v_1} - \vec{v_2})^2 - (\vec{v_1} \times \vec{v_2})^2/c^2}\,. \tag{71}$$

It is easily verified that in the case of two colliding beams with the same energy, this factor becomes $K = 2$.

Since the luminosity [4] must also be an invariant, the same correction factor must be used in the calculation of the luminosity.

8 Special relativity and electrodynamics

We come back to the original starting point and derive a relativistic formulation of Maxwell's equations.

8.1 Four-vector formulation of electromagnetic quantities

According to the rules for four-vectors, one can write the potentials and currents as four-vectors:

$$\phi, \vec{A} \Rightarrow A^{\mu} = \left(\frac{\phi}{c}, \vec{A}\right)\,, \tag{72}$$

$$\rho, \vec{j} \Rightarrow J^{\mu} = (\rho \cdot c, \vec{j})\,. \tag{73}$$

What about the transformation of current and potentials? Since we have formulated the potentials and currents as four-vectors, we transform the four-current as:

$$\begin{pmatrix} \rho' c \\ j'_x \\ j'_y \\ j'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho c \\ j_x \\ j_y \\ j_z \end{pmatrix} .$$

It transforms via: $J'^{\mu} = \Lambda J^{\mu}$ and the potential is transformed correspondingly: $A'^{\mu} = \Lambda A^{\mu}$. Since scalar products of four-vectors are invariant, one writes:

$$\partial_{\mu} J^{\mu} = \frac{\partial \rho}{\partial t} + \vec{\nabla}\vec{j} = 0 . \tag{74}$$

Equation (74) implies the conservation (invariance) of charge.

8.2 Electromagnetic quantities and field tensor

Having written the currents and potentials as four-vectors [1], we derive a formulation for the fields. The *Magnetic field is derived from the potential*:

$$\vec{B} = \nabla \times \vec{A} ,$$

e.g., written explicitly for the x-component:

$$B_x = \frac{\partial A_3}{\partial y} - \frac{\partial A_2}{\partial z} = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} .$$

The scalar potential provides the *electric field:*

$$\vec{E} = -\nabla \phi - \frac{\partial \vec{A}}{\partial t} ,$$

e.g. written explicitly for the x-component:

$$E_x = -\frac{\partial A_0}{\partial x} - \frac{\partial A_1}{\partial t} = -\frac{\partial A_t}{\partial x} - \frac{\partial A_x}{\partial t} .$$

These operations can be performed for all components of the fields. Then the electromagnetic fields can be condensed to form a field tensor $F^{\mu\nu}$:

$$F^{\mu\nu} := \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} = \begin{pmatrix} 0 & \frac{-E_x}{c} & \frac{-E_y}{c} & \frac{-E_z}{c} \\ \frac{E_x}{c} & 0 & -B_z & B_y \\ \frac{E_y}{c} & B_z & 0 & -B_x \\ \frac{E_z}{c} & -B_y & B_x & 0 \end{pmatrix} . \tag{75}$$

It transforms via: $F'^{\mu\nu} = \Lambda F^{\mu\nu} \Lambda^T$ (using the same transformation Λ as before). This corresponds to the transformation of the components of the fields into a moving frame.

Using the Lorentz transformation for $F^{\mu\nu}$, and writing the components explicitly, one gets easily:

$$\begin{aligned} E'_x &= E_x , & B'_x &= B_x , \\ E'_y &= \gamma(E_y - v \cdot B_z) , & B'_y &= \gamma \left(B_y + \frac{v}{c^2} \cdot E_z \right) , \\ E'_z &= \gamma(E_z + v \cdot B_y) z , & B'_z &= \gamma \left(B_z - \frac{v}{c^2} \cdot E_y \right) z . \end{aligned} \tag{76}$$

For the example of a Coulomb field (a charge moving with constant speed, Fig. 18):

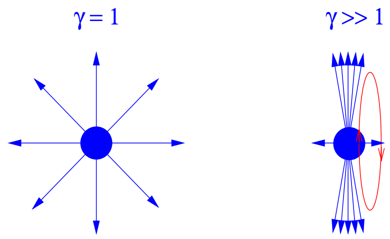


Fig. 18: Transformation of a Coulomb field to a moving frame

- in the rest frame, one has only electrostatic forces;
- in the moving frame, $\vec{E}$ is transformed and a magnetic field $\vec{B}$ appears.

When we rewrite the second component of Eq. (76):

$$B'_y = \gamma \cdot \frac{v}{c^2} \cdot E_z \,. \tag{77}$$

For small velocities v, this corresponds to the component of the Biot–Savart law. The Lorentz transformation automatically yields this law, including relativistic corrections for larger v.

8.3 An important consequence for moving charges in an accelerator

$$\begin{aligned} E'_x &= E_x \,, & B'_x &= B_x \,, \\ E'_y &= \gamma(E_y - v \cdot B_z) \,, & B'_y &= \gamma \left(B_y + \frac{v}{c^2} \cdot E_z \right) \,, \\ E'_z &= \gamma(E_z + v \cdot B_y) \,, & B'_z &= \gamma \left(B_z - \frac{v}{c^2} \cdot E_y \right) \,. \end{aligned} \tag{78}$$

Assuming that $\vec{B}' = 0$, we get for the *transverse* forces:

$$\vec{F}_{\mathrm{mag}} = -\beta^2 \cdot \vec{F}_{\mathrm{el}} \,. \tag{79}$$

For particles close to the speed of light, β = 1, and electric and magnetic *forces* cancel.

This has many important consequences, (e.g., space charge effects) because transverse fields generated by the charges vanish as the beam is accelerated.

This is most important for the stability of beams and (for $\beta \ll 1$) it cannot be ignored.

8.4 Retarded potentials

To compute these fields of moving charges, one can start with the four-potential of the charges at rest and apply the transformation.

For the static charge, we have the Coulomb potential [1] and $\vec{A} = 0$. The transformation into the new frame (moving in the x-direction) gives, for the new potentials:

$$\frac{\phi'}{c} = \gamma \left(\frac{\phi}{c} - A_x \right) = \gamma \frac{\phi}{c} \,, \tag{80}$$

$$A'_x = \gamma \left(A_x - \frac{v\phi}{c^2} \right) = -\gamma \frac{v}{c^2} \phi = -\frac{v}{c^2} \phi' \,. \tag{81}$$

i.e., all that is needed to compute the fields is the new scalar potential ϕ':

$$\phi'(\vec{x}) = \gamma \phi(\vec{x}) = \gamma \cdot \frac{1}{4\pi\epsilon_0} \cdot \frac{q}{|\vec{x} - \vec{x}_q|} \,. \tag{82}$$

After transformation of the co-ordinates, e.g., $x = \gamma(x' - vt')$, the resulting potentials can be used to compute the fields, as observed in the system at rest. However, one has to take care of causality.

The field observed at a position $\vec{x}$ at time t was caused at an *earlier* time $t_x < t$ at the location $\vec{x}_0(t_{\mathrm{r}})$ and the potentials have to be written as:

$$\phi(\vec{x}, t) = \frac{qc}{|\vec{X}|c - \vec{X}\vec{v}}, \qquad \vec{A}(\vec{x}, t) = \frac{q\vec{v}}{|\vec{X}|c - \vec{X}\vec{v}}. \tag{83}$$

The potentials $\phi(\vec{x}, t)$ and $\vec{A}(\vec{x}, t)$ depend on the state at the *retarded* time t_{r}, not t. Here ,$\vec{v}$ is the velocity at time t_{r} and $\vec{X} = \vec{x} - \vec{x}_0(t_{\mathrm{r}})$ relates the retarded position to the observation point.

8.5 Invariant formulation of Maxwell's equations

One can now rewrite Maxwell's equations using four-vectors and $F^{\mu\nu}$:

$$\nabla\vec{E} = \frac{\rho}{\epsilon_0} \qquad \text{and} \qquad \nabla \times \vec{B} - \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t} = \mu_0 \vec{J},$$

$$\stackrel{1+3}{\Rightarrow} \partial_\mu F^{\mu\nu} = \mu_0 J^\nu \qquad \text{(inhomogeneous Maxwell equation)}; \tag{84}$$

$$\nabla\vec{B} = 0 \qquad \text{and} \qquad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0,$$

$$\stackrel{1+3}{\Rightarrow} \partial_\gamma F^{\mu\nu} + \partial_\mu F^{\nu\lambda} + \partial_\nu F^{\lambda\mu} = 0 \qquad \text{(homogeneous Maxwell equation)}. \tag{85}$$

We have Maxwell's equations in a very compact form; transformation between moving systems is now very easy. The equivalent formulation in matter (macroscopic Maxwell's equation) is shown in Appendix B.

8.5.1 *Derivation of Gauss' law*

Starting from Eq. (84):

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu. \tag{86}$$

Written explicitly (Einstein convention, sum over μ):

$$\partial_\mu F^{\mu\nu} = \sum_{\mu=0}^{3} \partial_\mu F^{\mu\nu} = \partial_0 F^{0\nu} + \partial_1 F^{1\nu} + \partial_2 F^{2\nu} + \partial_3 F^{3\nu} = \mu_0 J^\nu. \tag{87}$$

As an example, one can choose ν = 0 and replace $F^{\mu\nu}$ with the corresponding elements:

$$\partial_0 F^{00} + \partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30} = \mu_0 J^0,$$

$$0 + \partial_x \frac{E_x}{c} + \partial_y \frac{E_y}{c} + \partial_z \frac{E_z}{c} = \mu_0 J^0 = \mu_0 c\rho. \tag{88}$$

This corresponds exactly to:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \qquad (c^2 = \epsilon_0\mu_0). \tag{89}$$

8.5.2 *Derivation of Ampère's law*

For $\nu = 1, 2, 3$, one obtains Ampère's law.

For example in the x-plane (ν = 1) and the F frame:

$$\partial_y B_z - \partial_z B_y - \partial_t \frac{E_x}{c} = \mu_0 J^x. \tag{90}$$

After transforming ∂^γ and $F^{\mu\nu}$ to the F' frame:

$$\partial'_y B'_z - \partial'_z B'_y - \partial'_t \frac{E'_x}{c} = \mu_0 J'^x \,. \tag{91}$$

It should be mentioned that now Maxwell's equations have the identical form in F and F'.

8.5.3 Combining the equations

Finally: since we have $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$,

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \,, \tag{92}$$

$$\partial_\gamma F^{\mu\nu} + \partial_\mu F^{\nu\lambda} + \partial_\nu F^{\lambda\mu} = 0 \,. \tag{93}$$

We can rewrite them two-in-one in a new form:

$$\frac{\partial^2 A^\mu}{\partial x_\nu \partial x^\nu} = \mu_0 J^\mu \,. \tag{94}$$

This expression contains all four Maxwell's equations, and it is the only one that stays the same in all frames!

One can conclude that there are *no* separate electric and magnetic fields; they are just a frame-dependent manifestation of a single electromagnetic field.

8.6 Electromagnetic forces in the framework of relativity

Starting with the four-force as the time derivative of the four-momentum:

$$\mathcal{F}^\mu_\mathcal{L} = \frac{\partial \mathcal{P}^\mu}{\partial \tau} \,. \tag{95}$$

We get the four-vector for the Lorentz force, where the spatial part is the well-known expression:

$$\mathcal{F}^\mu_L = \gamma q \left(\frac{\vec{E} \cdot \vec{u}}{c}, \vec{E} + \vec{u} \times \vec{B} \right) = q \cdot F^{\mu\nu} U_\nu \,. \tag{96}$$

8.7 Interpretation

To quote Einstein [2]:

> For a charge moving in an electromagnetic field, the force experienced by the charge is equal to the electric force, transformed into the rest frame of the charge.

Therefore, the Lorentz force is not an add-on to Maxwell's equations but just a consequence of two reference frames.

9 Summary

9.1 Summary—relativity basics

- Special relativity is very simple; there are a few basic principles.
  - Physics laws are the same in all inertial systems.
  - The speed of light in vacuum is the same in all inertial systems.
- Everyday phenomena lose their meaning (do not ask what is 'real').
  - Only the union of space and time preserve an independent reality: *space–time*.

- Electric and magnetic fields do not exist!
- They are simply different aspects of a *single electromagnetic field*.
- The manifestation of the electromagnetic field, i.e., division into electric $\vec{E}$ and magnetic $\vec{B}$ components, depends on the chosen reference frame.

9.2 Summary—consequences for particle accelerators

Write everything as four-vectors, blindly follow the rules and you get it all easily, in particular, transformation of fields, etc.

- Relativistic effects in accelerators (used in later lectures):
  - Lorentz contraction and time dilation (e.g., free-electron laser, ...);
  - relativistic Doppler effect (e.g., free-electron laser, ...);
  - invariants!
  - relativistic mass effects and dynamics;
  - new interpretation of electric and magnetic fields, in particular 'Lorentz force'.
- If you do not take relativity into account, you are sunk...

References

[1] W. Herr, Classical electrodynamics and applications to particle accelerators, these proceedings.

[2] A. Einstein, *Ann. Phys.* **322** (1905) 891. http://dx.doi.org/10.1002/andp.19053221004

[3] J. Hafele, *Science* **177**(4044) (1972) 166. http://dx.doi.org/10.1126/science.177.4044.166

[4] W. Herr, in Proc. of the CAS-CERN Accelerator School: Intermediate Course on Accelerator Physics, Zeuthen, Germany, 16–26 September 2003, edited by D. Brandt, CERN-2006-002 (CERN, Geneva, 2006), pp. 361-378. http://dx.doi.org/10.5170/CERN-2006-002.361

Bibliography

A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific Publishing, New Jersey, 1998).

R.P. Feynman, *Feynman Lectures on Physics* (Basic Books, New York, 2001), Vols. 1 + 2.

J. Freund, *Special Relativity for Beginners: A Textbook for Undergraduates* (World Scientific, New Jersey, 2008), http://dx.doi.org/10.1142/6601

J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998).

L. Landau and E. Lifschitz, *The Classical Theory of Fields* (Butterworth-Heinemann, Oxford, 1975), Vol. 2.

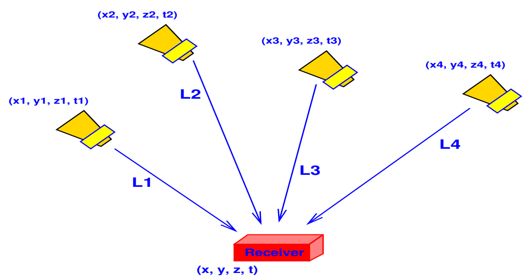


Fig. A.1: Schematic view of the GPS

Appendices

A Time dilation and general relativity

The time dilation due to the difference of height between two systems can be calculated as:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - \frac{2GM}{Rc^2}}\,, \tag{A.1}$$

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} \approx 1 - \frac{GM}{Rc^2}\,. \tag{A.2}$$

where G is the gravitational constant and R the distance from the centre of the Earth, and

$$\Delta\tau = \frac{GM}{c^2} \cdot \left(\frac{1}{R_{\mathrm{Earth}}} - \frac{1}{R_{\mathrm{GPS}}} \right)\,. \tag{A.3}$$

With the following parameters for the GPS (Fig. A.1):

$$R_{\mathrm{Earth}} = 6357000\ \mathrm{m}\,, \qquad R_{\mathrm{GPS}} = 26541000\ \mathrm{m}\,,$$

$$G = 6.674 \cdot 10^{-11}\ \mathrm{N\,m^2/\,kg^2}\,, \qquad M = 5.974 \cdot 10^{24}\ \mathrm{kg}\,,$$

we have:

$$\Delta\tau \approx 5.3 \cdot 10^{-10}\ \mathrm{s}\,. \tag{A.4}$$

B Tensors and macroscopic Maxwell's equations

It was mentioned in a previous lecture [1] that electric displacement $\vec{D}$ and magnetic field $\vec{H}$ are linked to the electric field $\vec{E}$ and induction $\vec{B}$ as:

$$\begin{aligned} \vec{D} &= \epsilon_0 \vec{E} + \vec{P}\,, \\ \vec{H} &= \frac{1}{\mu_0}\vec{B}\vec{M}\,, \end{aligned} \tag{B.1}$$

where $\vec{P}$ and $\vec{M}$ are the polarization and magnetization, respectively.

Each can be represented as a covariant tensor:

$$M^{\mu\nu} := \begin{pmatrix} 0 & P_x c & P_y c & P_z c - P_x c & 0 & -M_z & M_y \\ -P_y c & M_z & 0 & -M_x \\ -P_z c & -M_y & M_x & 0 \end{pmatrix}\,, \tag{B.2}$$

$$D^{\mu\nu} := \begin{pmatrix} 0 & -D_x c & -D_y c & -D_z c \\ D_x c & 0 & -H_z & H_y \\ D_y c & H_z & 0 & -H_x \\ D_z c & -H_y & H_x & 0 \end{pmatrix}\,. \tag{B.3}$$

The tensors are linked as:

$$D_{\mu\nu} = \frac{1}{\mu_0} F_{\mu\nu} - M^{\mu\nu}\,. \tag{B.4}$$

It can easily be verified that Eq. (B.4) is equivalent to Eq. (B.1).

The Gauss–Ampère law (Eq. (84)) becomes:

$$\partial_\mu D_{\mu\nu} = J^\nu\,. \tag{B.5}$$

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Undulator Technology

J. Pflueger
European XFEL GmbH, Schenefeld, Germany

Abstract
Since the 1970s undulators are used in storage rings and free-electron lasers as sources of intense radiation. This tutorial gives an elementary introduction and describes the principles as well as electromagnetic, superconducting and permanent magnet technologies, which are used in practice. Special emphasis is put on permanent magnet technology, which is most developed and used in most practical applications. An overview illustrated by many examples of the state of the art is given.

Keywords
Undulators; free-electron lasers; permanent magnets.

1 Introduction

The word ‘undulator’ originates from the Latin word for wave, ‘unda’. Its meaning is thus ‘wave maker’. By using a series of magnet poles with the same lengths and strengths but alternating field directions an ultra-relativistic electron beam is forced on a wave like, wiggling but overall straight orbit as shown schematically in Fig. 1.

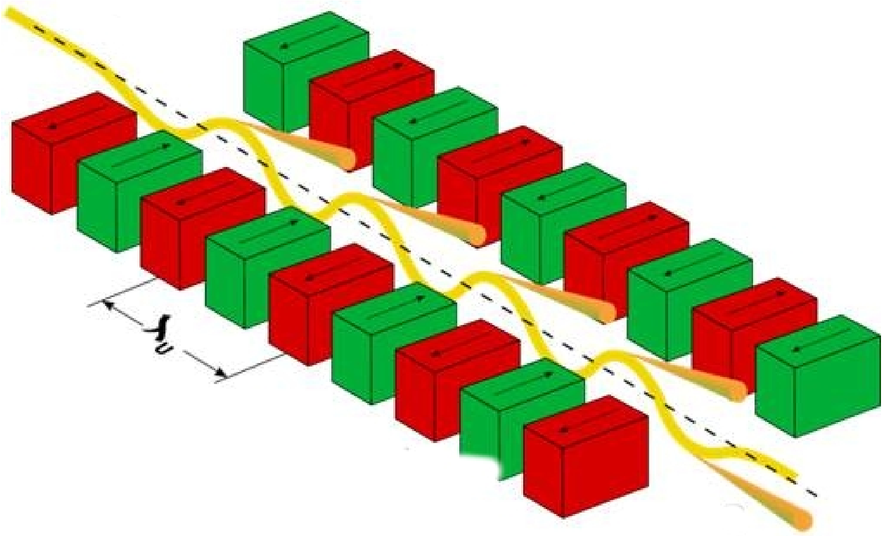


Fig. 1: Schematic of an undulator

Common and often used synonyms are ‘insertion device’ or ‘wiggler’. There are three technologies to create the periodic magnetic field: permanent magnet (PM), electromagnetic (EM) and superconducting (SC) technologies. They will be treated in this report. Typical device lengths are in the range of about 0.5 to 5 m. The development and use of undulators as intense sources of synchrotron radiation began in the late 1970s when first devices were developed for use in storage rings at the Budker Institute for Nuclear Physics (BINP) in Novosibirsk and the Lawrence Berkeley National Laboratory (LBNL) in co-operation with Stanford Synchrotron Radiation Laboratory (SSRL).

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.55

In a storage ring an undulator requires a straight section of an appropriate length. Such straight sections were very rare at that time. Therefore since the 1990s, dedicated third-generation storage rings were developed, which were optimized to accommodate a large number of straight sections and provide space for many insertion devices. Nowadays large facilities like the European Synchrotron Radiation Facility, (ESRF), the Advanced Photon Source (APS), the Super Photon Ring-8 GeV, (Spring8) or the reconstructed Positron Electron Tandem Ring Anlage, (PETRA III) accommodate dozens of such devices with lengths up to 5 m. An early historic insertion device built at Deutsches Elektronen Synchrotron (DESY) in 1983 is shown in Fig. 2. It is the 2.3 m long W1 wiggler. It occupied the only straight section available at that time in the storage ring "Doppelringspeicher", (DORIS).

Fig. 2: The W1 wiggler in use at DESY/HASYLAB in the DORIS storage ring from 1983 to 2012

In the last 15 years X-ray free-electron lasers (XFELs) using the principle of self-amplified spontaneous emission (SASE) were developed. They require very long systems of undulators and generate soft and hard X-ray beams with extreme properties allowing for revolutionary new experimental techniques. Examples are the **F**ree Electron **Las**er in **H**amburg, (FLASH) at DESY, Germany, the Linac Coherent Light Source, (LCLS) in Stanford, USA or the **S**pring8 **A**ngstroem **C**ompact Free Electron **La**ser (SACLA) at Spring8 in Harima, Japan. All are already in operation since several years. New projects with even more improved properties are in construction at the European XFEL/DESY (EXFEL) in Schenefeld/Hamburg, Germany, SwissFEL at Paul Scherrer Institute (PSI) in Villingen, Switzerland and XFEL at the Pohang Accelerator Laboratory (PAL-XFEL) in Pohang, Korea.

Depending on beam parameters and radiation properties the lengths of undulator systems vary from 30 m for FLASH to about 220 m for EXFEL. Although the requirements and specifications are different, the undulator technology has a lot in common.

This contribution will give a basic understanding of undulator properties and evaluation criteria. For a deeper insight the books by Onuki/Elleaume [1], Clarke [2] and the *Handbook of Synchrotron Radiation* [3] treat many theoretical as well as practical aspects, which were omitted on purpose in this contribution on technology: calculation of emission properties, magnetic measurement techniques of insertion devices, magnet design and special insertion devices. The state of the art of undulator technology is illustrated by a number of examples.

2 Basics

2.1 Equations of motion

In this section some fundamental relations for commonly used key parameters are derived. The motion of an electron in a periodic field of an undulator is sketched in Fig. 3.

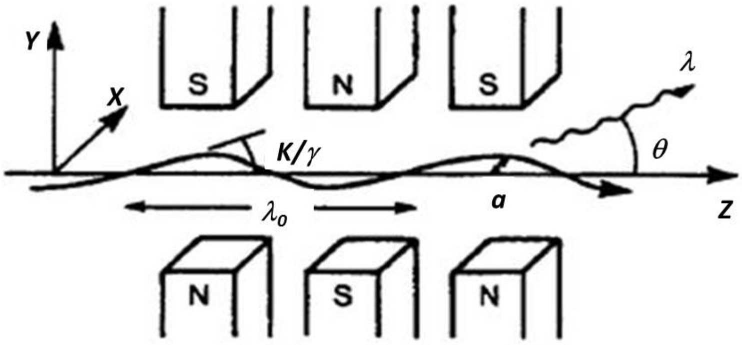


Fig. 3: Electron motion in an undulator

This figure also defines the coordinate system used in this contribution. The motion of a single electron in a magnetic field is controlled by the Lorentz equations:

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = m_0\gamma\frac{\mathrm{d}}{\mathrm{d}t}\vec{v} = m_0\gamma\begin{pmatrix}\ddot{x}\\ \ddot{y}\\ \ddot{z}\end{pmatrix} = e\left[\vec{v}\times\vec{B}\right] = e\cdot\begin{pmatrix}v_yB_z - v_zB_y\\ v_zB_x - v_xB_z\\ v_xB_y - v_yB_x\end{pmatrix}. \tag{1}$$

Here $\vec{F}$ is the force acting on an electron, $\vec{p}$ its momentum and $\vec{v}$ its velocity vector. m_0 is the electron rest mass, γ is the kinetic energy in units of the rest mass and e its charge. $\vec{B}$ is the vector of the magnetic field.

2.2 Transverse motion

For the motion in a planar undulator Eq. (1) is solved by making the assumption for a planar field:

$$\vec{B} = \begin{pmatrix}0\\ B_0\sin\left(\frac{2\pi}{\lambda_0}z\right)\\ 0\end{pmatrix}, \tag{2}$$

where λ_0 is the period length of the field and B_0 its amplitude. $B_x,\ B_z = 0$. $\vec{B}$ is a purely transverse field which varies along *z*, the direction of propagation. The initial conditions are

$$v_x,\ v_y = 0;\ v_z = \beta c\,, \tag{3}$$

where *c* is the speed of light and the following relations are used:

$$\beta = \sqrt{1-\frac{1}{\gamma^2}} \cong 1-\frac{1}{2\gamma^2};\quad \gamma = \frac{E_{\mathrm{Kin}}}{m_0c} \gg 1\,. \tag{4}$$

For multi-GeV beams this condition is very well satisfied and γ is in the order of many times 10^3. In practice for most undulators the period length, λ_0, is in the range of 10 to 400 mm and B_0 of the order of 1–2 T. As will be shown quantitatively below, transverse velocities are small enough so that the small-angle approximation can be made:

$$\frac{v_x, v_y}{\beta c} = x', y' << 1\,. \tag{5}$$

Then Eq. (1) can be simplified:

$$\begin{aligned}\ddot{x} &= -\frac{e}{\gamma m_0} v_z B_y \,, \\ \ddot{y} &= 0 \,, \\ \ddot{z} &= \frac{e}{\gamma m_0} v_x B_y \cong 0 \,.\end{aligned} \tag{6}$$

For the deflection angle x' it can be solved by integration:

$$x'(z) = \frac{v_z(z)}{\beta c} = -\frac{e}{\gamma m_0 c} \int_{-\infty}^{z} B_y(z')\mathrm{d}z' \,. \tag{7}$$

The integral

$$\int_{-\infty}^{z} B_y(z')\mathrm{d}z' = \boldsymbol{I}1(z) \tag{8}$$

is called the first field integral and can be calculated using measured field data, if available.

For the sinusoidal field as assumed in Eq. (2),

$$x'(z) = \frac{eB_0\lambda_0}{\gamma\ m_0 c 2\pi} \cos\left(\frac{2\pi}{\lambda_0} z\right) = \frac{K}{\gamma} \cos\left(\frac{2\pi}{\lambda_0} z\right) \tag{9}$$

is obtained, where

$$K = \frac{eB_0\lambda_0}{m_0 c 2\pi} = 0.0934\, B_0[\mathrm{T}] \cdot \lambda_0[\mathrm{mm}] \tag{10}$$

defines the undulator K-parameter for a purely sinusoidal field as assumed in Eq. (2). The maximum excursion angle of the electron beam is given by $K/_{\gamma}$. For non-sinusoidal but periodic B_y fields it is given by

$$K_{\mathrm{Def}} = \mathrm{Max}\left(\frac{e}{m_0 c} \boldsymbol{I}1(z) \right), \tag{11}$$

where the suffix ‘Def’ marks the definition using the deflection criterion. Real fields of undulators are periodic, but in general contain higher harmonics and therefore the K-parameter differs from the definition given in Eq. (10).

The electron trajectory in the X–Z plane is obtained by a second integration of Eq. (9):

$$x(z) = -\frac{e}{\gamma m_0 c} \int_{-\infty}^{z} \left(\int_{-\infty}^{z'} B_y(z'')\mathrm{d}z'' \right) \mathrm{d}z' \,. \tag{12}$$

Here the second field integral is defined as

$$\boldsymbol{I}2(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{z'} B_y(z'')\mathrm{d}z'' \right) \mathrm{d}z' \,. \tag{13}$$

For the sinusoidal field of Eq. (2), the result is

$$x(z) = -\frac{eB_0\lambda_0^2}{\gamma m_0 c\, 4\pi^2} \cdot \sin\left(\frac{2\pi}{\lambda_0} z\right) = -\frac{K}{\gamma}\ \frac{\lambda_0}{2\pi} \cdot \sin\left(\frac{2\pi}{\lambda_0} z\right), \tag{14}$$

so that the amplitude of the trajectory oscillation, A, is given by

$$A = \frac{K}{\gamma}\ \frac{\lambda_0}{2\pi} \,. \tag{15}$$

A short comment on non-monochromatic field contributions: Eqs. (10) and (15) are commonly used and give estimates better than 10–15%. Details, however, depend on the presence of higher field harmonics, which in turn depend on technological details such as the minimum gap and the period length λ_0 as well as on dimensions of the magnetic active parts. It therefore needs a careful analysis if higher precision is required.

2.3 Longitudinal motion

In a magnetic field the total velocity of an electron is constant. This connects the longitudinal and transverse motions:

$$v_y^2 + v_z^2 = (\beta c)^2 \,. \tag{16}$$

Using Eqs. (7) and (9), the evolution of the longitudinal speed can be written:

$$v_z = \sqrt{(\beta c)^2 - v_x^2} \cong \beta c\left(1 - \frac{v_x^2}{2\cdot(\beta c)^2}\right) = c\beta\left[1 - \frac{K^2}{4\gamma^2} - \frac{K^2}{4\gamma^2}\cos\left(\frac{4\pi}{\lambda_0}z\right)\right] . \tag{17}$$

This result shows two consequences:

1. The average longitudinal speed is reduced since the oscillations in the undulator increase the path length. This can be accounted for using $\bar{\beta}$ defined as

$$\bar{\beta}=\beta\left[1 - \frac{K^2}{4\gamma^2}\right] . \tag{18}$$

2. In addition, the longitudinal speed is modulated by the factor $\frac{K^2}{4\gamma^2}\cos\left(\frac{4\pi}{\lambda_0}z\right)$ with two oscillations per period λ_0. As compared to Eq. (9), the longitudinal amplitude is reduced by $\frac{K}{4\gamma}$.

Equation (17) can be rewritten:

$$v_z = c\left[\bar{\beta} - \frac{K^2}{4\gamma^2}\cos\left(2\cdot\frac{2\pi}{\lambda_0}z\right)\right] . \tag{19}$$

2.4 Slippage, optical phase and phase errors

Light travels at light speed c. An electron in an undulator travels at lower average speed given by $\bar{\beta}c$. This gives rise to an effect called 'slippage', which is explained in Fig. 4. The black full line sketches the oscillating trajectory of an electron in an undulator. The time for light to travel the distance from A to B, one period length, λ_0, is given by $t_c = \frac{\lambda_0}{c}$. In that time the electron travels only the distance $\frac{\lambda_0}{c}\bar{\beta}c$. The difference is called 'slippage'. Using Eqs. (4) and (18) and neglecting the term proportional to $\frac{1}{\gamma^4}$ the slippage of one period is given by

$$\Delta = \lambda_0\left(1 - \bar{\beta}\right) = \frac{\lambda_0}{2\gamma^2}\left(1 + \frac{K^2}{2}\right) . \tag{20}$$

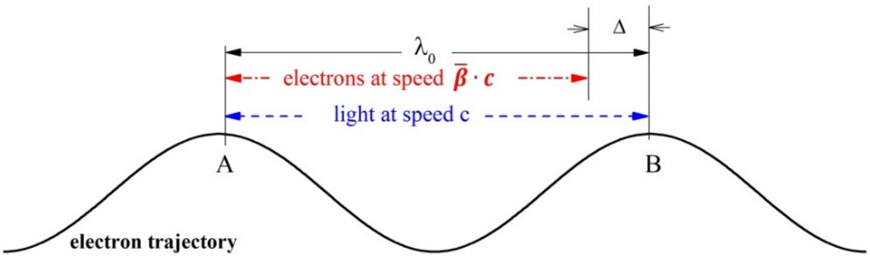


Fig. 4: Slippage in an undulator

Since the slippage in all periods of an ideal undulator is the same, the light emitted by different periods constructively interferes at the wavelength given by Eq. (20). It is therefore more common to rewrite Eq. (20) using the radiation wavelength, λ_{Rad}:

$$\lambda_{\mathrm{Rad}} = \frac{\lambda_0}{2\gamma^2}\left(1 + \frac{K^2}{2}\right) , \tag{21}$$

which is the well-known resonance condition for the first harmonic of an undulator.

2.5 Optical phase

The accumulated slippage in a magnetic field extending from z_0 to z is given by

$$\Delta(z) = \int_{z_0}^{z} \left(c - v_z(z')\right) \mathrm{d}z' . \tag{22}$$

The optical phase is the total slippage normalized to $\frac{\lambda_{\mathrm{Rad}}}{2\pi}$ and defined as

$$\varphi(z) = 2\pi \frac{\Delta(z)}{\lambda_{\mathrm{Rad}}} . \tag{23}$$

Combining Eqs. (7), (17), (21) and (22), one obtains the optical phase for $B_y(z)$

$$\varphi(z) = 2\pi \frac{\Delta(z)}{\lambda_{\mathrm{Rad}}} = \frac{2\pi}{\lambda_0 (1 + \frac{K^2}{2})} \left(z + \left(\frac{e}{m_0}\right)^2 \int_{-\infty}^{z} \left(\int_{-\infty}^{z''} B_y(z') \mathrm{d}z' \right)^2 \mathrm{d}z'' \right) . \tag{24}$$

This result can again be applied to measured magnetic field data $B_y(z)$. The normalization to λ_{Rad}, the first harmonic, eliminates γ and leads to an energy-independent form.

The double integral

$$\mathrm{PI}(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{z''} B_y(z') \mathrm{d}z' \right)^2 \mathrm{d}z'' \tag{25}$$

is commonly called 'Phase Integral'.

2.6 Phase errors

The optical phase is very important for evaluating the quality of an undulator in terms of its emission properties. This is understood by a qualitative argument: field errors lead to changes in the transverse velocity and thus result in changes of the longitudinal velocity as seen by Eq. (17). This will perturb the phase advance per period of 2π and lead to a phase mismatch of the radiation emitted by different periods and the quality of the produced radiation will be degraded. The criterion which controls this degradation is the phase jitter, also called phase error. It is defined as the RMS difference between the ideal and the actual phases on the poles of an undulator as determined by Eq. (24).

This can be written as

$$\mathrm{PJ} = \frac{1}{N} \sum_{i=1}^{N} (i \cdot \pi - \varphi(z_i))^2 . \tag{26}$$

Here i labels the poles and z_i are the corresponding positions. The nominal phase advance per pole is π. Phase errors are a reliable quality criterion for the evaluation of emission properties of an undulator. For spontaneous emission of undulators in synchroton radiation (SR) sources this was established by Walker; see Ref. [4]. For undulators in SASE-FELs, Li *et al.* [5] have made a thorough investigation. Today magnetic measurement and tuning techniques allow for RMS phase jitters of 1° or 0.0175 rad or even less. However, this should not be overstressed. Such a small PJ is only needed if an undulator will be operated at a high harmonic, which is often the case in SR sources, see Ref. [4]. In contrast, SASE-FELs are operated on the first harmonic only. Here a PJ of 11° or 0.192 rad is sufficient. It should be mentioned that the phase-error criterion avoids the unnecessary over specification of very small peak field errors, which is sometimes found. As shown in Ref. [5] some errors at the proper location are tolerable.

2.7 Demonstration and example

The relations derived in the previous sections are illustrated by two examples in Fig. 5. A short model of 20 periods of a XFEL U40 undulator with λ_0 = 40 mm and B_0 = 1 T is shown. The *K*-parameter is 3.72. For simplicity and explanation this short model field rather than that of a real 5 m long undulator with 120 full periods was selected.

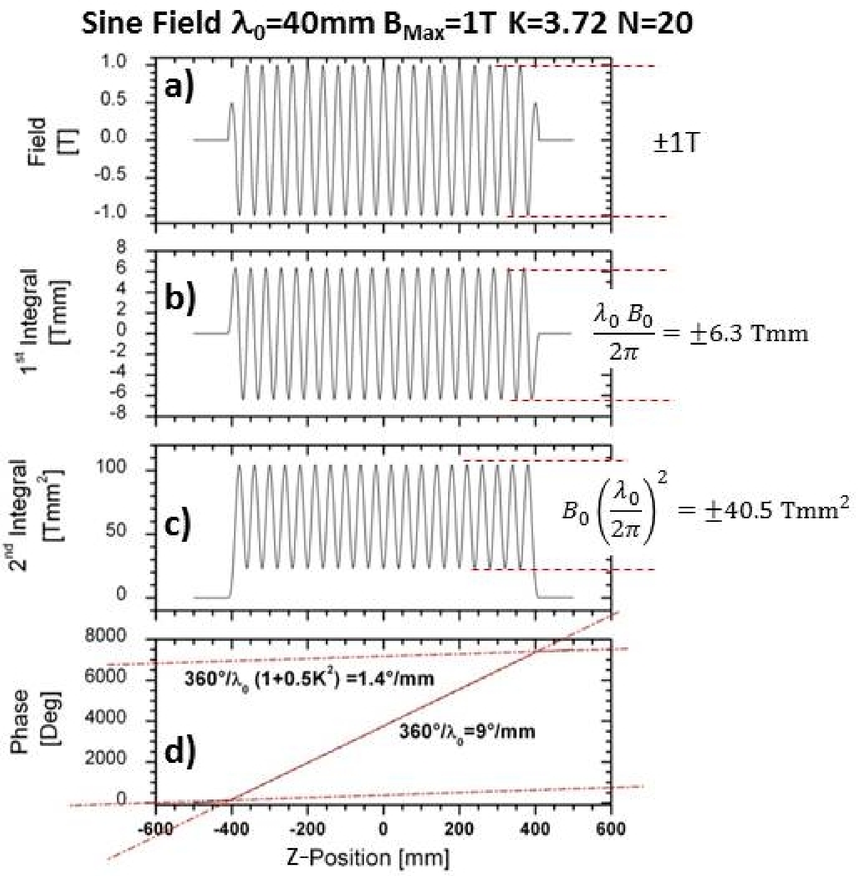


Fig. 5: Field, first, second field integrals and phase integral for a short model with 20 periods

Figure 5(a) shows the field with an amplitude of ±1 T. The first field integral with an amplitude of ±6.3 T mm calculated using Eqs. (7) and (9) is shown in Fig. 5(b). The amplitude of the second field integral, Eqs. (12) and (13), is ±40.5 T mm2 and shown by Fig. 5(c). Finally, Fig. 5(d) shows the phase advance and illustrates Eq. (24). On both ends, outside the undulator, there is zero field and the phase evolution is that in free space represented by a straight line with a slope given by $\frac{360°}{\lambda_0(1+0.5K^2)} = 1.4°/\mathrm{mm}$. Inside the undulator, the phase advance is 360° per period or a slope of $\frac{360°}{\lambda_0} = 9°/\mathrm{mm}$. It is seen that over the 20 periods the phase advance is about 7200°.

The measured phase errors of a typical XFEL U40 are shown in Fig. 6. The phase error on each pole i is calculated using Eqs. (24) and (26). There are about 240 poles. The phase errors on all these poles were measured at six different gaps and their RMS values are shown in Fig. 6.

It is seen that there is some systematic variation along the undulator. Its amplitude changes with gap. The smallest value is at 14 mm gap leading to an RMS error of 1.77°. The results show that the phase error of this undulator is well within the XFEL specifications, which require an RMS phase error of less than 8°.

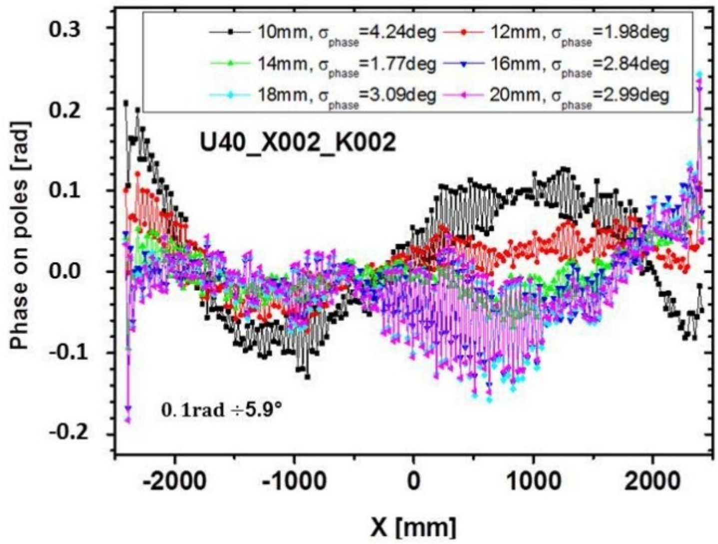


Fig. 6: Phase errors of a XFEL U40 measured at different gaps

3 Hardware technology

3.1 Technological limitations

EM, SC and PM technologies are used for the technical realization of periodic fields in undulators. In this section their Pros and Cons as well as technological limits will be discussed.

The first two, EM and SC technologies, are magnet systems excited by currents in conductors. Apart from the much higher currents and current densities in SC systems there is no fundamental difference between an EM and a SC system. In an EM system it is the current in a copper conductor, in a SC system a typical wire material is NbTi. Both need a sufficiently large cross-section to carry the total excitation current. This marks a difference to PM systems, where PM material based on SmCo or NdFeB with a remanent field of 0.9–1.25 T is used. Homogeneous PM material can be described by a surface current given by

$$j_{\mathrm{Surface}} = \frac{M}{\mu_0} , \tag{27}$$

where M is the magnetization and $\mu_0 = 1.256 \times 10^{-6}$ V s/A m the vacuum permeability. For a remanent field of 1.25 T a surface current of about 10^6 A/m results. This results in a fundamental difference in the scaling properties of EM and PM systems, which is explained in Fig. 7. Here the scaling properties of a simple EM and a PM dipole system using an iron yoke are investigated if all dimensions are scaled down, i.e. divided by a factor a, and the scaled down systems are required to have the same field in the gap.

For the original EM system the field in the gap is obtained by integrating the magnetic field strength, H, around a closed path containing the enclosed current as indicated in Fig. 7. The iron contribution is negligible due to its very large permeability and only the field in the gap contributes.

$$\oint \; H\mathrm{d}s = \frac{B_{\mathrm{Gap}} \cdot \mathrm{Gap}}{\mu_0} = j_{\mathrm{Area}} A = I \;\; \rightarrow \;\; B_{\mathrm{Gap}} = \frac{j_{\mathrm{Area}} A}{\mathrm{Gap}} \mu_0 \, . \tag{28}$$

A is the cross-section of the conductor, j_{Area} is the current density and I is the total enclosed current. For the scaled down system the linear dimensions are divided by the scaling factor a, $\mathrm{Gap}_{\mathrm{Scaled}} = Gap/a$ and $A_{\mathrm{Scaled}} = A/a^2$. This leads to the requirement for the current density:

$$j_{\text{Area,Scaled}} = j_{\text{Area}} \cdot a \tag{29}$$

if the requirement $B_{\text{Gap}} = B_{\text{Gap,Scaled}}$ needs to be fulfilled. So, for the system scaled down by $a > 1$, the current density needs to increase proportional to a. There are, however, technical limits for current densities, as will be shown below.

A PM system behaves differently. Integration of H along a closed path results in

$$\oint \ H\text{d}s = \frac{B_{\text{Gap}}\text{Gap} + l_{\text{m}}B_{\text{m}}}{\mu_0} = j_{\text{Surface}}l_{\text{m}} = \frac{M}{\mu_0}l_{\text{m}} = I\,. \tag{30}$$

The magnet is treated like an infinitely thin air coil of length l_{m} and the surface current density is given by Eq. (27). Since the flux in the gap and magnet, i.e. the number of field lines, is preserved, $B_{\text{m}} = B_{\text{Gap}}$,

$$B_{\text{Gap}} = \frac{Ml_{\text{m}}}{\text{Gap} + l_{\text{m}}} = \frac{M}{\frac{\text{Gap}}{l_{\text{m}}} + 1}\,. \tag{31}$$

In this result only the geometric ratio Gap/l_{m} determines the field, not the absolute coordinates. So, the fields of PM structures are invariant under a change of scale. The consequence of this scaling property is seen when the geometries are miniaturized: EM structures are limited by the current density for copper of $\lesssim 10\,\frac{\text{A}}{\text{mm}^2}$; for superconductors it may exceed $1500\ \frac{\text{A}}{\text{mm}^2}$ but the basic principle of limitation is the same. For a PM system such a limitation does not exist.

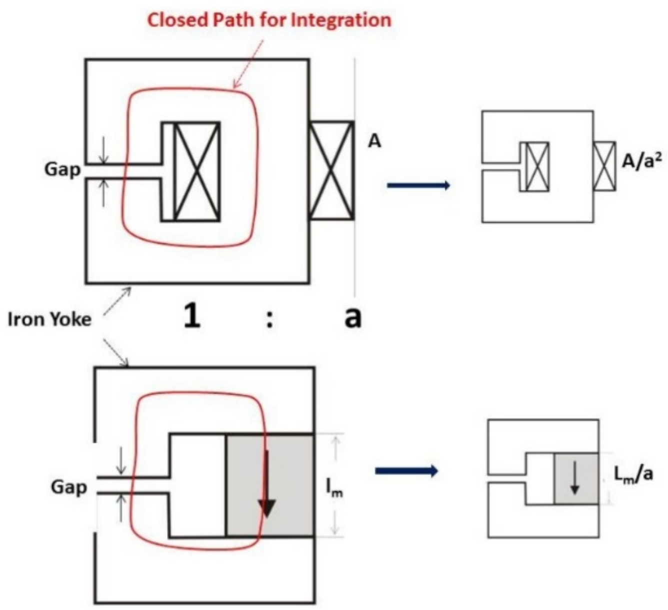


Fig. 7: Scaling properties of EM and PM systems

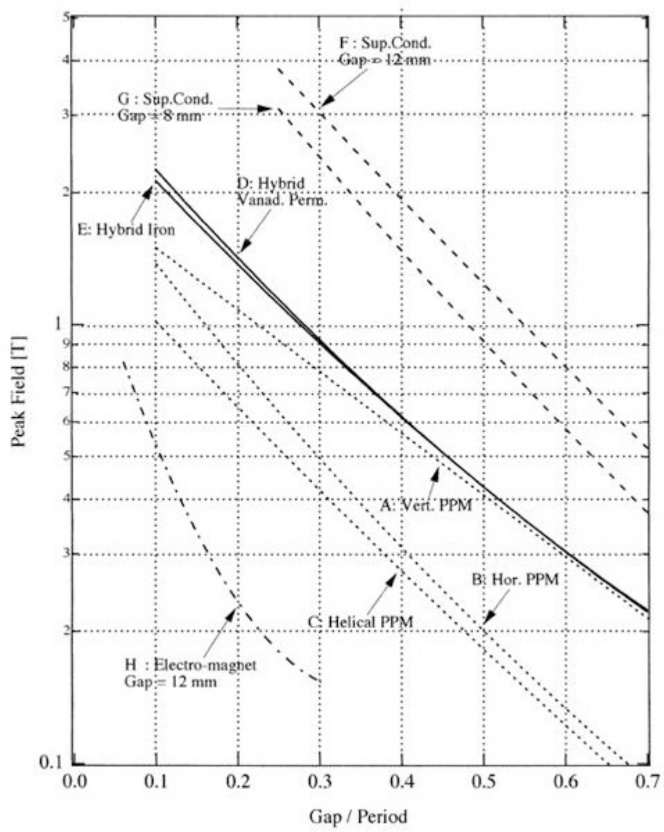


Fig. 8: Comparison of EM, SC and PM technologies

For illustration and without going into any design details, Fig. 8 shows a comparison of the peak field of undulators using EM, SC and PM technologies. The figure is taken from Ref. [6]. The scaling is gap/period, which is appropriate for PM. For EM and SC the comparison is made at gap = 12 mm. It is seen that EM shows the lowest values. A peak field of 0.8 T is only possible if the period length is about 200 mm. This is of practical use only for special cases.

SC offers the highest fields. At a gap of 12 mm a period length of about 50 mm would result in almost 4 T and at a period length of 17 mm still about 0.5 T is possible. For PM technology several curves are shown for different magnet designs, which will be explained below. They are somewhere in between EM and PM.

In summary, EM technology may be useful for undulators with long or even very long period lengths, $\lambda_0 > 200$ mm, and for very special applications. Its technology is well established and has much in common with classical EM magnets for accelerator applications.

SC has two regions: at periods larger than about 40 mm, fields of several Teslas may be reached. Wavelength shifters are such extreme applications, as will be shown below. There are numerous such long-period, high-field devices in operation in various laboratories. Typically, they have very few periods, sometimes only one. This technology is well established.

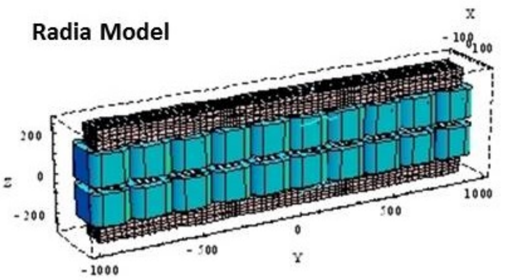
In contrast, the technology for short-period SC undulators is much more demanding, still under development and far away from being ‘state of the art’. At present there are only very few laboratories world-wide working on short-period SC undulators. There are, however, promising technological developments, which need time to get established. At period lengths down to 10–15 mm SC outperforms the other technologies but with decreasing difference the smaller the periods are.

PM technology is well established and is the workhorse technology for insertion devices. An estimated 90–95% of all devices are built in this way. Very large systems have been built in co-operation with industrial suppliers. For example, the European XFEL requires 91 5 m long undulators with a total magnetic length of 455 m. Typical period lengths for PM undulators are 10–250 mm. In contrast to EM and SC undulators, PM technology has one important and obvious advantage: since a PM system is permanently excited there are no operation costs for electric power or cryogenics as for EM and SC systems and maintenance costs are low.

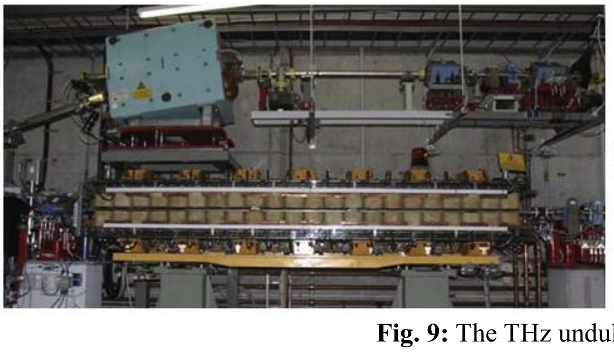
3.2 EM examples

Two examples for EM undulators will be explained. The first is shown in Fig. 9. It is the infrared/THz undulator for FLASH. It is a planar device with a very large period length of 400 mm. Water-cooled coils are wound around the poles, as can be seen schematically in the RADIA model in the upper left. The device has a gap of 40 mm, which is needed to guide the infrared radiation without losses. The maximum field at full excitation is 1.2 T. For FLASH operated at 1.2 GeV very intense coherent infrared radiation with wavelengths as long as 4.2 μm can be generated with this device. Operation cost should not be neglected: its power consumption at full excitation is 80 kW. At current electricity prices and 100% operation this amounts to around 400 €/day or 140 k€/year.

The second example is the EM helical undulator built for the APS and is shown in Fig. 10. It is a more sophisticated and rather exotic device. The cross-section is shown in the upper left part. There is a vertical structure for the B_y field similar to that in Fig. 9 but there is a second structure for the horizontal B_x field, which is displaced by a quarter period. The overall geometry is such that it allows for the insertion of a vacuum chamber with a very large horizontal aperture as required in storage rings and as shown in the upper left part of Fig. 10. Therefore, the B_x coils are split into an upper and a lower part. The period length is 125 mm, which is short for an EM undulator but the resulting field at 10.5 mm gap is only 0.3 T resulting in a K-parameter of 3.5. At the APS with 7 GeV this structure will create soft X-ray radiation around 500 eV. There is another speciality: the iron core of this device is laminated. This allows for operation of the coils with AC at 10 Hz rather than DC. So, the helicity of the field can be reversed quickly, which is important for studying magnetic phenomena.



Gap	40 mm
Period length	400 mm
Pole length/width	100/140 mm
Number of full periods	9
Number of poles	44
End termination pattern	+1/8,-1/2,+1,...,-1,+1/2,-1/8
Iron yoke length	4.3 m
Maximum field/K-value	12 kG/49
Number of turns of central main coils	64
Conductor cross-section	$8.5 \times 8.5\ \mathrm{mm}^2$, ⌀5.3 mm bore
Maximum current density	$8.7\ \mathrm{A/mm}^2$
First/second field integral	< 200 G cm, $< 20\ \mathrm{kG\,cm}^2$
Maximum magnetic force	237 kN
Cooling water flow	100 l/min
Water temperature rise at 435 A	20 °C
Maximum temperature gradient (water cut-off)	0.4 °C/s
Maximum current	435 A
Voltage at 435 A	208 V
Maximum total power	87 kW
Total weight	4490 kg



O. Grimm, N.Morozov, A.Chesnov, Y.Holler, E.Matushevsky, D.Petrov, J. Rossbach, E.Syresin , M.Yurkov, NIMA 615 (2010) 105

Fig. 9: The THz undulator for FLASH

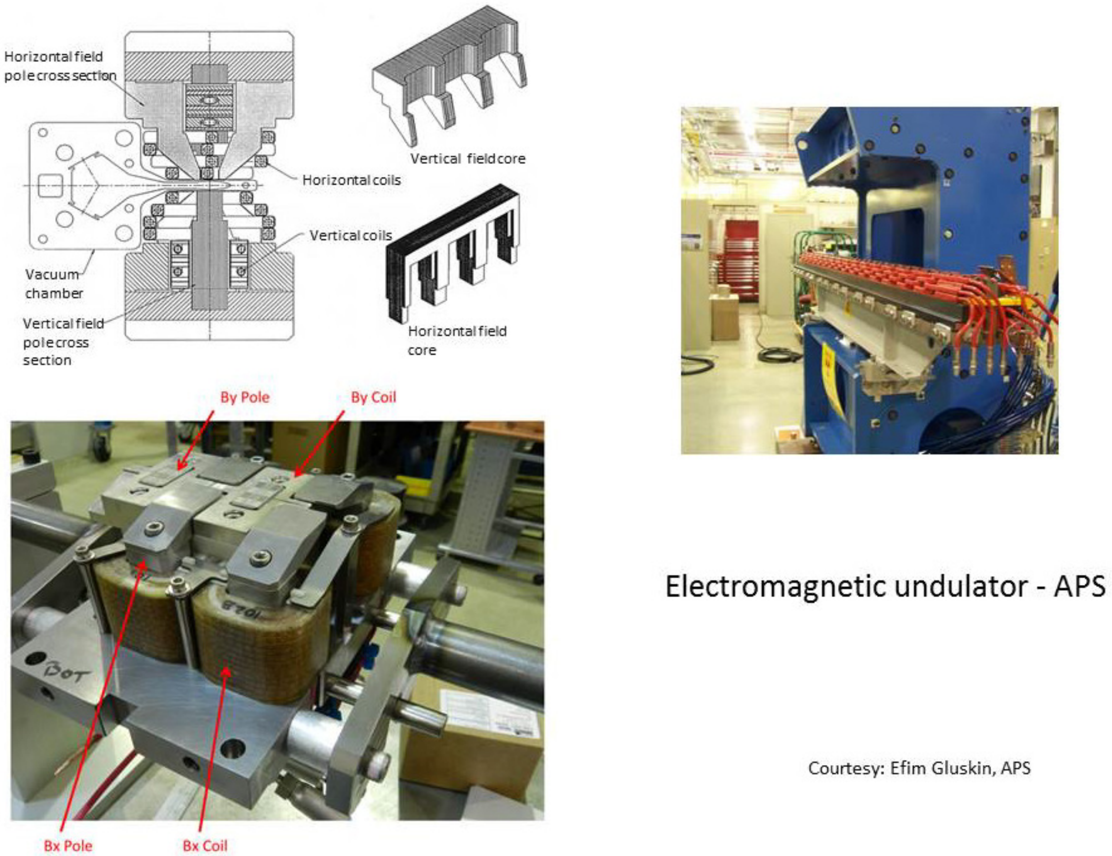


Fig. 10: EM helical undulator for the APS

3.3 SC devices

The technological advantages of SC devices were already described in Section 3.1. As an example of long-period/high-field devices Fig. 11 shows the wavelength shifter built for Spring8. It has a magnetic length of about 1 m only and accommodates just three poles. The central pole is designed for highest field and reaches up to 10.2 T. It is surrounded by two side poles with lower field and longer length, which balance and control the first and second field integrals. The field distribution is seen in the lower left of Fig. 11. At 8 GeV this device is used to create ultra-hard X-rays with a critical energy of 440 keV, which should be compared with about 64 keV from a conventional bending magnet at 1.5 T. In general wavelength shifters are good choices if the spectrum of X-rays emitted by bending magnets is too soft. An example: in a typical soft X-ray storage ring with 2 GeV the critical photon energy of a conventional 1.5 T bending magnet is about 4 keV only. This is much too low for special X-ray techniques such as protein crystallography, which requires photon energies of about 24 keV (0.5 Å). With a 10.2 T wavelength shifter as described above the critical energy is shifted to 27 keV.

The technology for these devices is well established and used in many storage rings throughout the world.

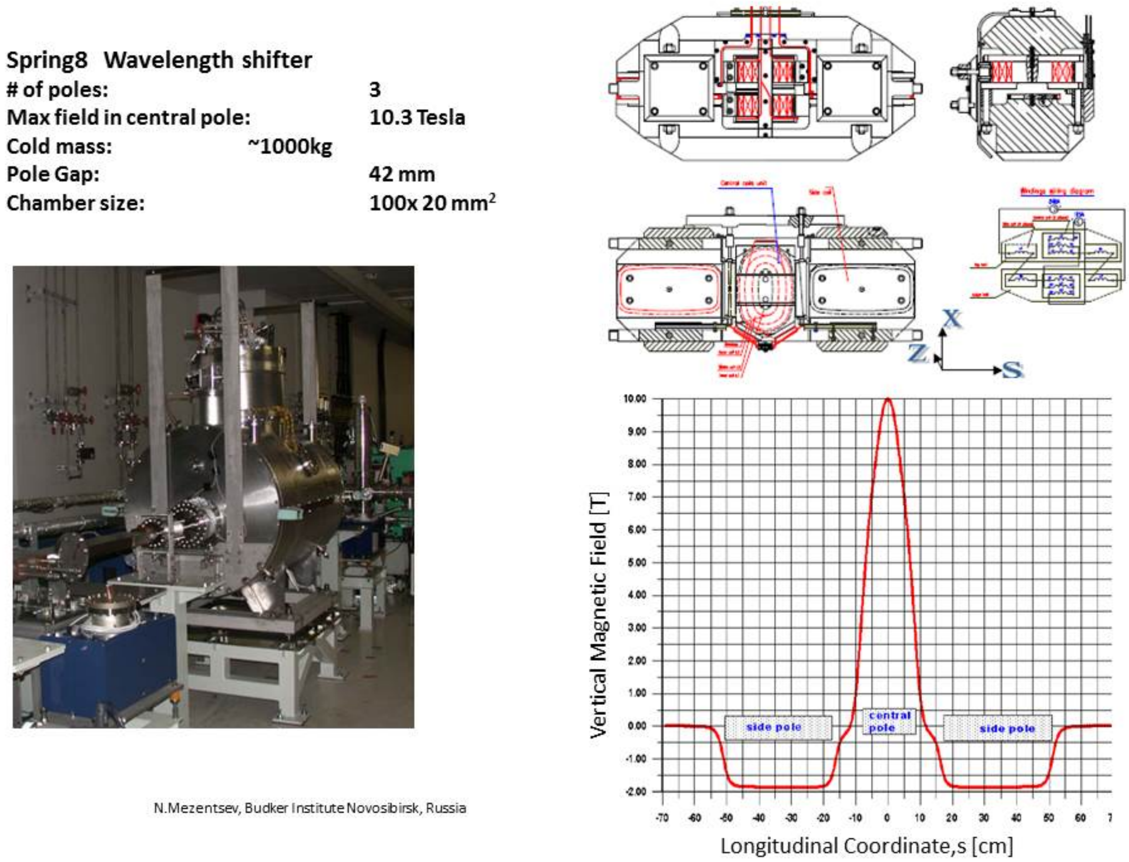


Fig. 11: The SC wavelength shifter built by BINP for Spring8

For short-period SC undulators the situation is quite different. Two examples are shown in Fig. 12. The upper one was built for the Anstroemquelle Karlsruhe (ANKA) at the Karlsruhe Institute of Technology (KIT) and is in operation since 2005. The lower one is a short prototype under development for the APS upgrade. It has only 20 periods and a magnetic length of about 300 mm but is inserted in a full-size 2 m cryostat, the planned final length of the device.

This device is part of a systematic development at the APS. The plan is to use a large number of SC devices for the planned upgrade of the APS storage ring expected to start in 2020. It will be the first large-scale use of SC undulators.

Some of the technological challenges which need to be solved are briefly mentioned.

i) Proper radiation shielding is required to prevent heat load from SR originating from upstream magnets especially in long small-gap devices.

ii) A vacuum chamber is needed to separate the accelerator vacuum from the cryogenic part, which reduces the usable gap.

iii) Magnetic measurements in such a device require substantial effort.

iv) There is no compensation scheme for field errors. Field quality can only be guaranteed by perfect manufacturing.

v) The production technology needs to be further developed. At present (2016) it is not yet mature enough to be used for a large number of devices.

The target parameters for the final APS device are: λ_0 = 15 mm; $B_{\text{Peak}} \leq 1.5$ T; $K \leq 2.1$; magnetic gap = 7.3 mm; vacuum gap = 5.0 mm. This should be compared to an in-vacuum hybrid PM device where magnetic and vacuum gaps are the same, that is, 5.0 mm: λ_0 = 15 mm; gap = 5.0 mm: B_{Peak} = 0.82 T; K = 1.15. So peak field and K-parameter of a SC device are significantly larger in spite of the larger magnetic gap of the SC undulator. This demonstrates the advantage of SC technology.

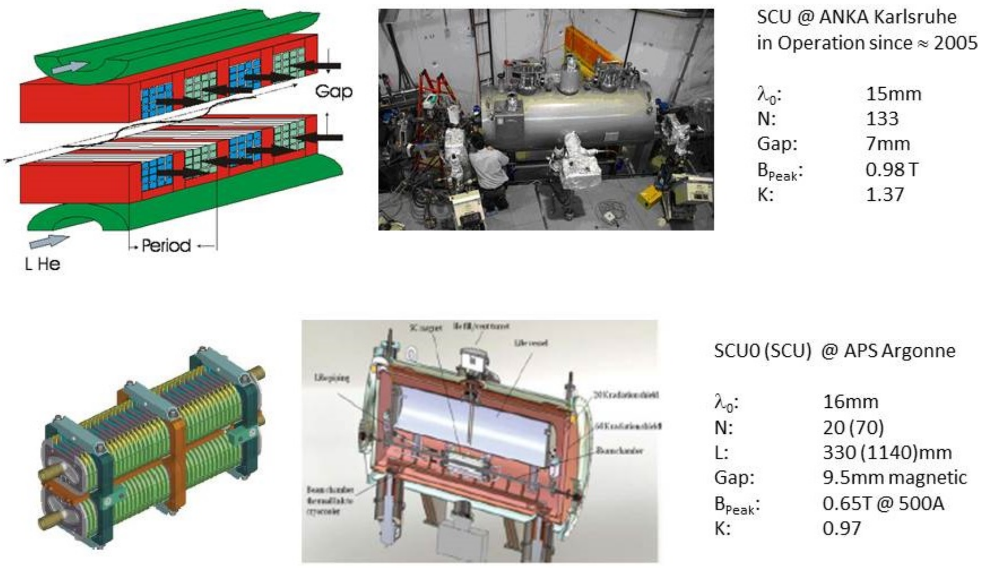


Fig. 12: SC undulator development at ANKA and APS

3.4 PM undulators

3.4.1 Magnet design

New PM materials based on sintered SmCo and NdFeB compounds were developed in the 1970s and 1980s. They offer much higher energy products resulting in higher remanent and coercive fields and therefore much increased magnetic performance as compared to traditional ferrite or AlNiCo materials, which were used before. This development revolutionized many applications using PM technology.

The use of these materials for insertion devices was pioneered by Klaus Halbach, who proposed the two different magnet configurations, which are still used without modification for planar undulators worldwide [7, 8]. The pure permanent magnet (PPM) structure is shown in Fig. 13, left. The structure is assembled from parallelepipeds arranged in two rows and magnetized as shown in the figure. The space between the upper and lower rows is called the ‘gap’. By changing the gap mechanically, the field strength in the gap can be accurately controlled. The field of a homogenously magnetized parallelepiped with $\mu_{\text{r}} = 1.0$ can be calculated analytically using the current-sheet method. The field of a complete structure is obtained by superposition of the fields. This is good for many calculations. For precise data, however, the finite permeability μ_{r} needs to be taken into account; $\mu_{\text{r}} \approx$ 1.05–1.07 for NdFeB and 1.02–1.03 for SmCo.

Obviously the field of a PPM undulator is fully determined by the PM material, its quality, homogeneity, magnetic orientation, mechanical dimensions and manufacturing accuracy. These are important quality criteria and depend on many details of the production process. There might be large variations for commercially available materials.

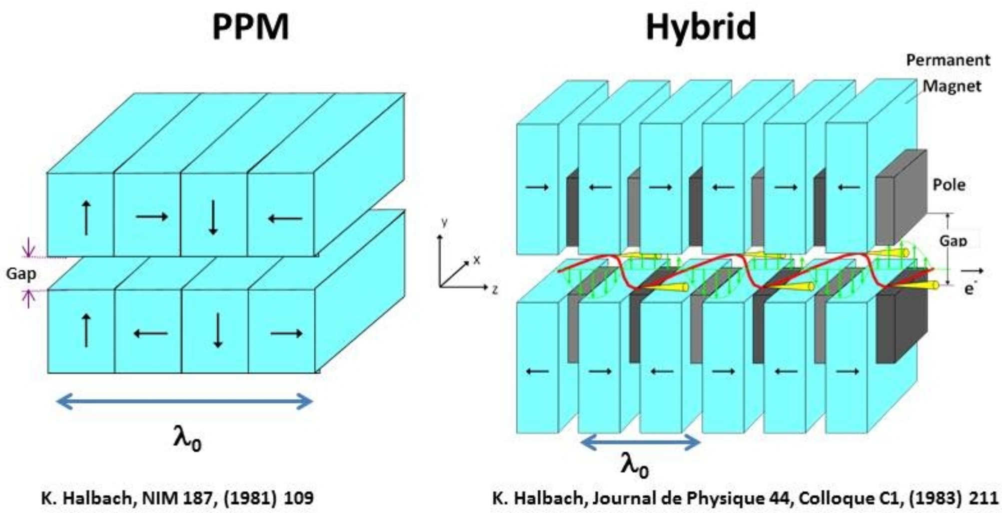


Fig. 13: Left-hand side: PPM array. Right-hand side: hybrid array

An alternative is the hybrid configuration, Fig. 13, right, which avoids some of the drawbacks of the PPM design but requires more effort. It uses a combination of soft iron poles and PM material arranged again in a bottom and a top array with a gap in between. The iron poles concentrate the flux of the magnets and conduct it to the gap. The fields in general are higher than for PPM. In contrast to PPM undulators the field is dominated by the geometry of the poles and only to a lesser extent by material quality. This allows a better and more direct control of field errors as well as eventually some compromises on material quality.

For field calculations of hybrid undulators there are no analytic methods and numeric codes need to be used. Since these codes generally allow for $\mu_{\mathrm{r}} > 1$ they are often used for PPM structures as well. A very popular code in the insertion device community is RADIA developed at the ESRF [9].
Today for most undulator applications NdFeB magnet material is used because it offers the highest energy product. Only for special applications requiring high temperatures or extremely high coercive fields SmCo material is an alternative.

For design work and parameter determination it is very useful to have an analytic formula which describes the peak field as a function of the gap. A convenient form was already given in Ref. [8], see also Ref. [6]:

$$B\left(\frac{g}{\lambda_0}\right)[\mathrm{T}] = a\,\mathrm{e}^{b\frac{g}{\lambda_0}+c\left(\frac{g}{\lambda_0}\right)^2} \tag{32}$$

The normalization to $\lambda_0$visualizes the scaling properties described in Section 3.1. The constants a, b and c are determined by fitting using either calculated or measured data of a specific design and geometry. Using scaling the same magnet design can in principle be scaled to different λ_0.

Figure 14 shows a selection of normalized gap dependencies fitted using Eq. (32). They are taken from the literature as well as from results obtained at DESY and the European XFEL. The coefficients a, b and c are shown in Table 1 for some of the curves shown in Fig. 14. Here are some remarks

1. As expected, the PPM curve is well below the hybrid curve.
2. The majority of the curves are close together. They all use hybrid magnet designs optimized for relatively small period lengths, λ_0 < 48 mm.
3. In contrast, the curve for BW5 was optimized for a large-period device with λ_0 = 230 mm. The objective was a high field of about 1.98 T and a gap of 20 mm, gap/$\lambda_0 = 0.087$.

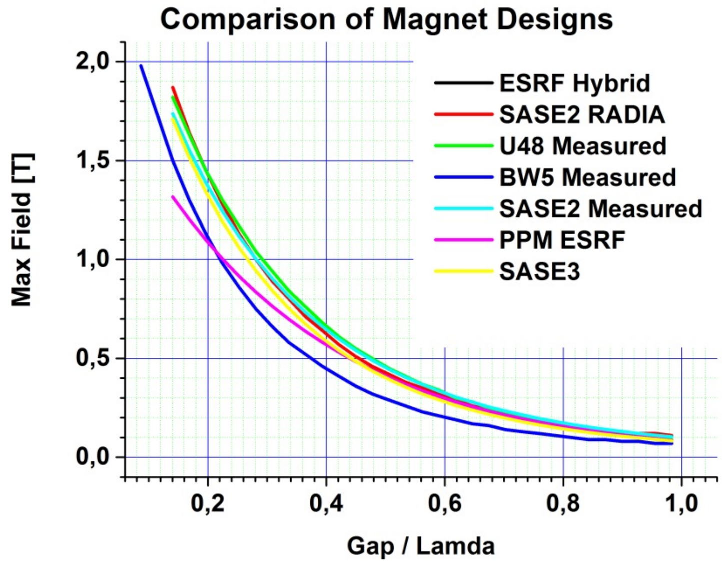


Fig. 14: Examples of gap dependences normalized to $\boldsymbol{\lambda_0}$

Limits to scaling are only set by physical dimensions. Applying scaling from a small to a large period length requires increasing the volume of magnets and poles proportional to the third power. At material cost of 200 to 500 €/kg for magnet material this sets an economic limit for large-period devices, which therefore always are a compromise between magnet volume/cost and achievable field. An example for scaling is given for the DORIS BW5 wiggler: with λ_0 = 230 mm the weight of one magnet block was ≈15 kg. Using the SASE2 design, which would allow for slightly higher fields and scaling, its weight would have to be about 57 kg.

Table 1: Examples for fitting the parameters *a*, *b*, *c* in Eq. (32)

Magnet structure	***a***	***b***	***c***
Hybrid FeCo poles[a]	3.694	–5.068	1.52
XFEL SASE2 measured[b]	3.10487	–4.24914	0.80266
XFEL SASE3 measured[c]	3.2143	–4.62305	0.92541
DORIS III BW5[d]	3.1852	–5.6036	1.6891
PPM[a]	2.076	–3.24	0

[a] See Ref. [6].
[b] Measured data of EXFEL U40; λ_0 = 40 mm.
[c] Measured data of EXFEL U68; λ_0 = 68 mm.
[d] Measured data of DORIS III 2 T wiggler BW5; λ_0 = 230 mm.

For the design of the mechanics of PM undulators, no matter whether they use PPM or hybrid technology, two points are important:

1. The magnets of the top and bottom structures need to be mounted on girders. In the case of the European XFEL the length is 5 m. There are significant attractive magnetic forces between these girders, which need an accordingly massive and stiff support in order to guarantee homogeneous field properties. The sinusoidal magnetic field applies an attractive magnetic 'pressure' between the upper and lower structures, which can be estimated by

$$F[\mathrm{N/m^2}] = \frac{B_0^2[\mathrm{T}]}{4\mu_0} \tag{33}$$

As a rule of thumb, a field of 0.5 T corresponds to about 0.5 bar, which for a 5 m long and 70 mm wide structure amounts to an attractive force of 17500 N. For 1.7 T, which applies to the worst case for the SASE3 undulator for the European XFEL at 10mm gap, it amounts to about 200 kN. Girder stiffness and the supports need to be designed in such a way that under a dynamic load change with these forces the dynamic girder deformation is typically such that the resulting peak to peak homogeneity does not significantly exceed $\frac{\Delta B}{B} \leq 10^{-3}$.

2. The K-parameter of a PM undulator is tuned by the mechanical adjustment of the gap. For X-ray FELs a typical requirement on adjustment accuracy is $\frac{\Delta K}{K} \leq 10^{-4}$. As a result a typical specification on the mechanics, the drive motors, measurement systems and the motion control system is to allow for a gap-adjustment accuracy better than ±1 μm.

3.4.2 *Some examples*

3.4.2.1 *Open C-frame*

Figure 15 shows some examples of the C-frame geometry. It is a very common way to arrange the magnet structures. The principle is shown in Fig. 15(a). There is a stiff frame which is a good support for the guide rails, to which the girders are connected to. The gap is adjusted via spindles. Nowadays usually gears are avoided and axes are coupled and synchronized electronically by the control system rather than by hardware such as shafts and gears. There might be four motors, one for each spindle, as seen in the two examples for LCLS II, Fig. 15(b) and the European XFEL, Fig. 15(d) or alternatively two motors in combination with two right/left spindles, as seen in Fig. 15(c), which shows a standard carriage of the ESRF. This device in addition is equipped with four spring systems, which are used for partial compensation of the magnetic forces. The great advantage of the C-geometry is its good access from the open side. This is seen in Fig. 15(d), which shows a 5 m long undulator segment aligned on the magnetic bench. Magnetic measurements and tuning can be done in an alternating fashion. The C-bracket with the hand-off sign seen in the foreground on the end of the girder is part of the gap-measurement system with a verified accuracy of ±1 μm.

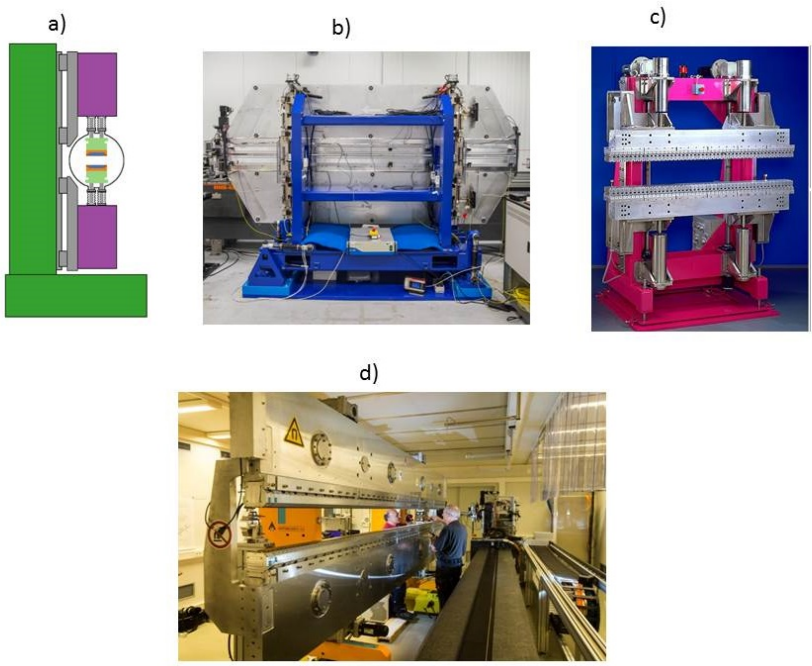


Fig. 15: Undulators with C-frame geometry. (a) Principle; (b) 2.3m LCLS II prototype; (c) 1.6m ESRF standard carriage; (d) A 5m U40 undulator for the European XFEL aligned on the magnetic bench.

3.4.2.2 Revolver undulators

A special case of C-frame devices are revolver undulators. They allow for the use of different magnet structures which are mounted on rotatable drums or girders so they can be changed rapidly. The principle is shown in Fig. 16(a). Different magnet structures are arranged on drums. Different ones can be selected by proper rotation of the upper and lower drums. Figure 16(b) shows the 4 m long BW3 undulator used in DORIS III from 1990 to 2012. It provides four positions. The bearings for the drum rotation need to be at the ends and allow for four positions and continuous rotation.

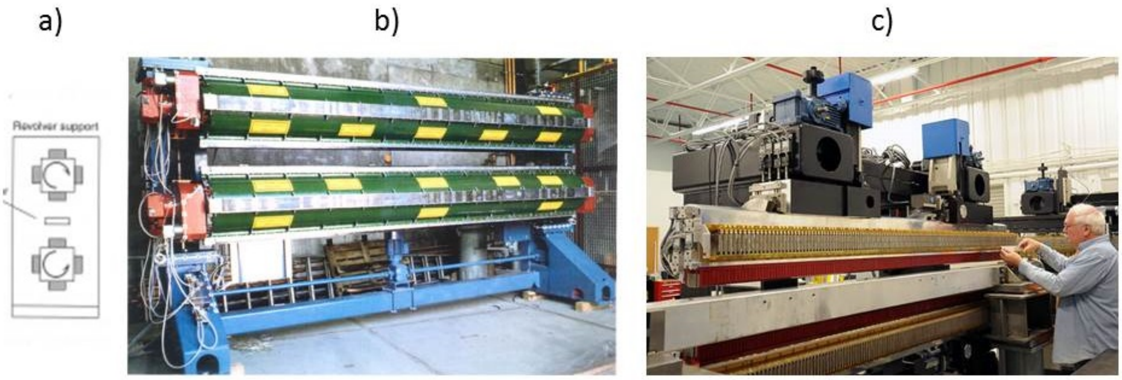


Fig. 16: Revolver undulator. (a) Principle; (b) DORIS III BW3 revolver with four positions; (c) APS revolver with cradle support with two positions.

However, but their locations are non-optimum for mechanical deformation under attractive forces. Enforced drums as well some extra length of about 0.3 m for the bearings on either end are required. Deformation properties can be improved by shifting the support to the Bessel points using cradle-type circular guide rails as seen in Fig. 16(c). But in this case only two structures can be accommodated and continuous rotation is not possible. Revolvers are used in many laboratories to extend the scan range of undulators.

3.4.2.3 H-frame geometry

The H-frame geometry is shown schematically in Fig. 17(a). It uses a symmetric and closed frame and therefore in contrast to the C-frame it is more compact and less prone to deformation. Figure 17(b) shows an early example built in 1987: it is the hard X-ray wiggler (HARWI) used in DORIS III during 1987–2004 for coronary angiography. For highest performance requiring smallest gaps in dedicated SR runs it was equipped with a variable vacuum chamber. Lateral access in H-frame devices is very restricted. Only limited magnetic measurements can be performed. High-precision systems as shown in Fig. 15(d) cannot be used. For these reasons compact H-frame devices as shown in Fig. 17 are rare.

Recently a very slim measurement system was reported, which can be used under such spatial conditions. Although it was built for in situ measurements in in-vacuum undulators, which encounter the same problem, see also next section, it might be a perfect system for the magnetic measurement of H-frame devices as well.

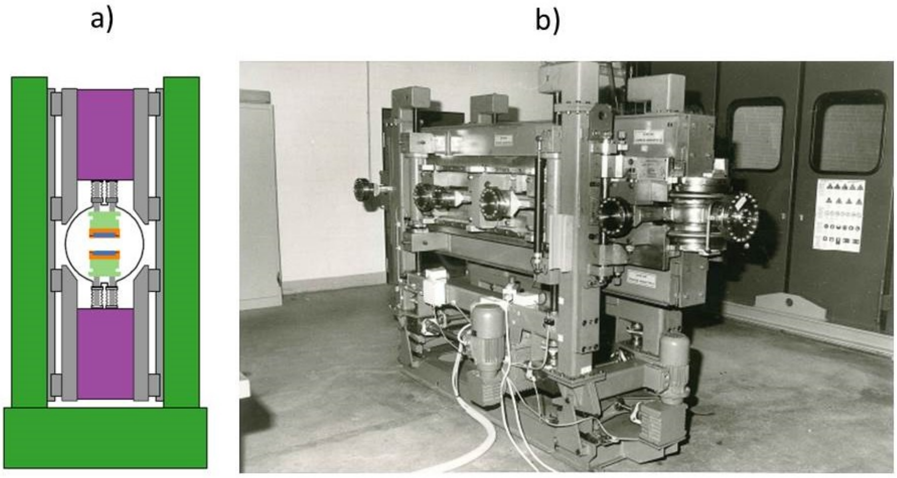


Fig. 17: (a) H-Frame geometry; (b) 2.4 m long Hard X-ray Wiggler (HARWI) used at DORIS III for coronary angiography 1987–2004.

3.4.2.4 In-vacuum undulators (IVUs)

The relativistic electron beam, which is passed through an undulator in order to generate light, needs a vacuum chamber. This aspect has not been treated so far. There exist two alternatives, which are illustrated in Fig. 18. In most undulators a separate vacuum chamber with an aperture as small as possible is used; see Fig. 18(a). Such an out of vacuum chamber requires space for the vacuum chamber wall thickness plus some tolerances. This limits the usable magnetic gap. An example is given in Fig. 18(c). It shows the dimensions of the vacuum chamber for the undulator segments of the European XFEL. It is made of an extruded AlMg alloy, which has been machined to exact final dimensions. The minimum magnetic gap of the undulator is 10.000 mm, the vertical beam stay clear is 8.6 mm and the vertical outside dimension of the vacuum chamber is 9.5 mm. This results in a wall thickness of 0.45 mm only. There is a 0.5 mm tolerance for the chamber to fit in the undulator. However, in this optimized example the usable gap is reduced by 1.4 mm.

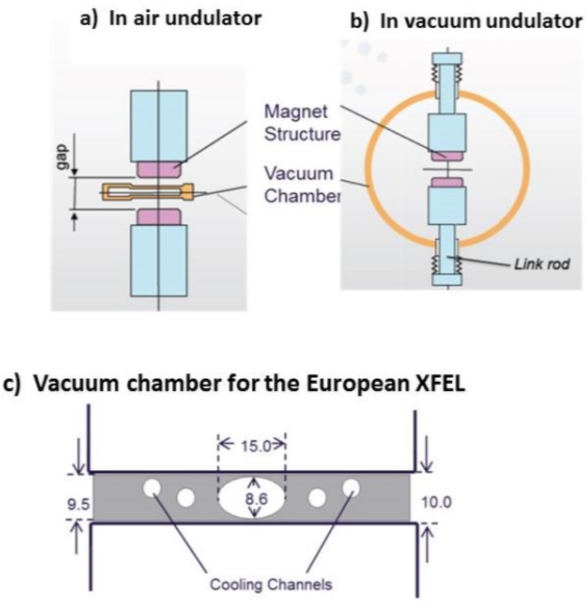


Fig. 18: (a) Conventional out of vacuum chamber; (b) in-vacuum undulator; (c) vacuum chamber of the European XFEL: magnetic gap: 10mm, chamber outside dimension: 9.5mm, alignment tolerance: 0.5mm, beam stay clear: 8.6mm. The resulting wall thickness is 0.45mm only.

An alternative is to place the magnet structure inside the vacuum, Fig. 18(b). In an IVU the magnet structure is completely inside the vacuum chamber and there is no loss in usable gap. There are several problems which should be mentioned.

1. The complete magnet structure needs to be compatible with the ultra-high-vacuum (UHV) conditions of the accelerator.
2. All magnets need special coating in order to prevent outgassing of the sintered magnet material.
3. The whole structure must be designed to avoid virtual leaks.
4. Compatibility with bake out requires selection of adequate magnet material with high H_c.
5. In order to keep the diameter of the vacuum chamber small the magnet structure is supported by a number of link rods. They are connected to a massive external girder using feedthroughs. The assembly of the magnet structures inside the vacuum vessel implies problems with alignment and reproducibility.
6. Magnetic measurements and magnetic tuning are special challenges.

As an example the assembly of an IVU at SACLA is demonstrated in Fig. 19. On the left the pre-assembled IVU is aligned on the magnetic bench without vacuum chamber to provide lateral access for magnetic measurements and tuning. Some of the details already mentioned above are seen: there are the massive external magenta support girders, 12 pairs of link rods with their feedthroughs protected by aluminium foil and there is the very slim magnet structure, which will go inside the vacuum chamber. After measurement and tuning the magnet structure needs to be completely detached from the 2 × 12 link rods, transferred inside the vacuum chamber and re-attached. Obviously there is only finite reproducibility of the re-attachment of the link rods to the girders. This may induce errors in the magnetic field, which would stay undetected if no further steps are made. The final assembled IVU is seen on the right of Fig. 19. The endcap of the vacuum vessel is still opened. Most IVUs in use today use the techniques described above and end at this point.

A big step forward was the development of the self-aligned field analyser with laser instrumentation, (SAFALI) system which allows in situ magnetic measurements; see Ref. [10]. In this way the re-attachment errors can be measured and compensated. Its principle is shown in Fig. 20, left. There is a very slim guide system, which carries the field probe and is supported by three adjustable posts. Their supports go through three flanges in the vacuum vessel. During measurements the probes are kept on axis using two laser positioning systems and feedback loops to adjust the posts in such a way as to keep the probe on axis. This system in operation is seen in Fig. 20, right.

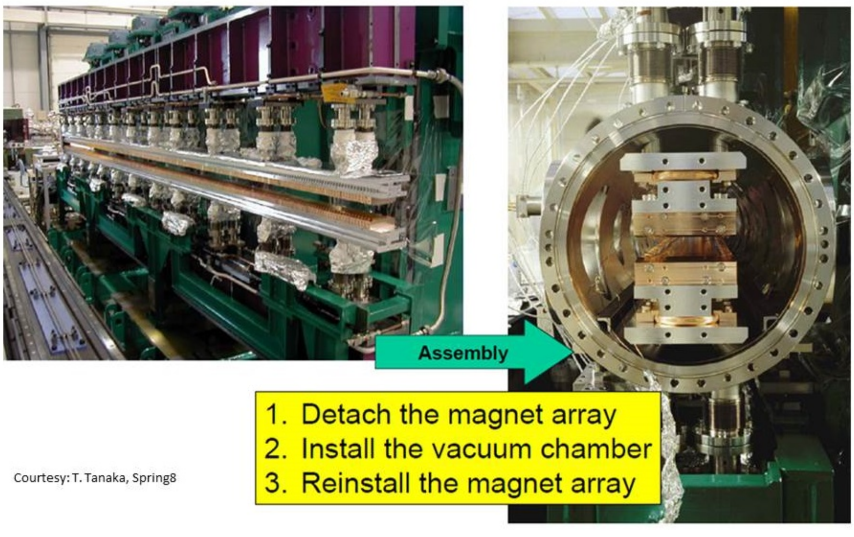


Fig. 19: Magnetic measurement and final assembly of an in-vacuum undulator

Figure 21 shows one of the SACLA IVU systems. It consists of 18 undulator segments of 5 m length. The period length is 18 mm. The minimum operational gap is as small as 3.7 mm and the maximum K-parameter is 2.1. At 8 GeV X-rays with a wavelength as short as 1 Å can be produced.

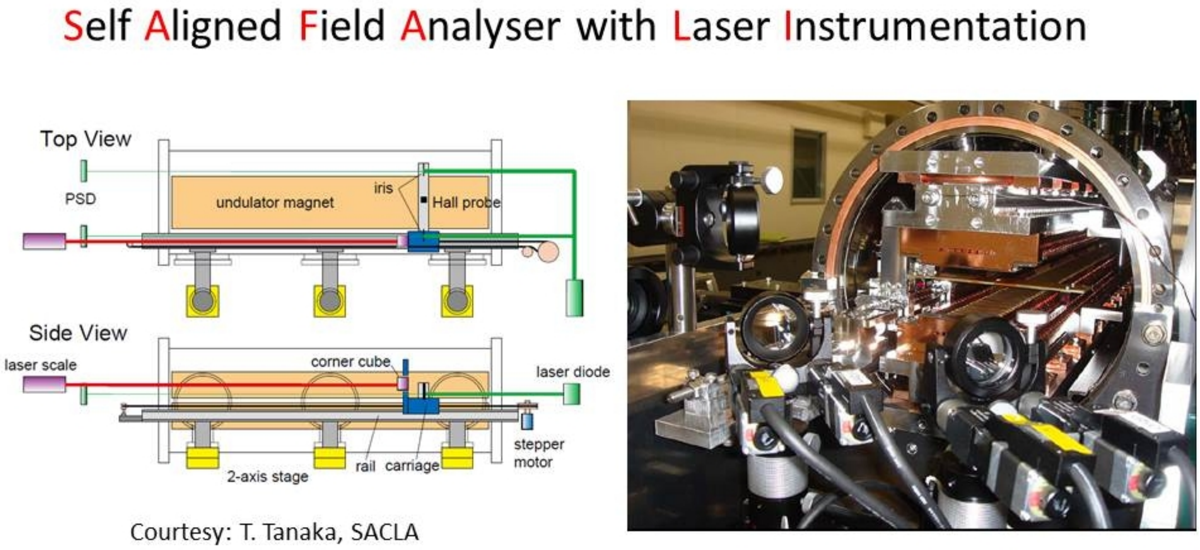


Fig. 20: SAFALI system. Left-hand side: principle; right-hand side: measurement on an IVU

Fig. 21: One of the SACLA IVU systems

3.4.2.5 APPLE undulators

For PPM structures the superposition principle allows the combination of fields produced by different magnet structures. This gave rise to a number of different types of insertion devices, which have been proposed over the years. A full treatment and explanation is beyond the scope of these proceedings. However, one example, the advanced polarized light emitter, APPLE, undulator [11] became popular for spectroscopic applications and is briefly explained. It allows the creation of light with any polarization state: planar, right/left circular, elliptical and planar with arbitrary plane of polarization. The principle is shown in Fig. 22. An APPLE undulator consists of a PPM structure, in which each half is subdivided into two rows. Each row can be moved individually. This is shown schematically in Fig. 22(a). If all rows have the same shift the geometry is equivalent to a PPM undulator and there is only a B_y field component. If the rows are shifted diagonally as indicated in Fig. 22(a) a horizontal field component is generated resulting in elliptical polarization. If the shift is $\pm\lambda_0/4$ helical radiation with right/left helicity is generated. Planar fields can be generated by mutually shifting the diagonal pairs.

An APPLE undulator provides a planar gap and therefore is ideal for storage rings, which require a large horizontal aperture. Its total field is controlled by the gap. The horizontal gap together with the C-frame geometry allows good access for magnetic measurements. However, depending on the shift of the rows there might be large forces, which are not present in planar undulators. Therefore they require substantially enforced massive frames. Figure 22(b) gives an impression. It shows the 5 m long APPLE undulator built for upgraded Positron Electron Tandem Ring Accelerator PETRA III.

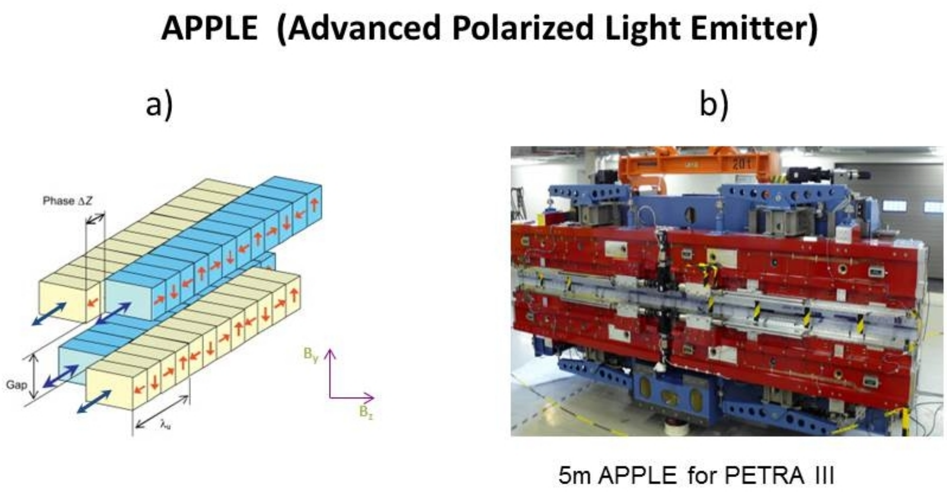


Fig. 22: APPLE undulator

4 Summary and outlook

Undulators are indispensable components in storage rings and in SASE FELs, where they are used in very long systems. In this contribution an overview of the current status of undulator technology is given. The most important parameters such as period length, λ_0, the *K*-parameter, radiation wavelength λ_{Rad} , first and second field integrals, the optical phase and the RMS phase errors were explained.

The three different technologies for building undulators were introduced: EM, SC and PM technologies. Their domains and pros and cons were discussed. Special emphasis was put on PM technology, which is used in the majority of applications.

Stimulated by new accelerator developments, which allow smaller emittances and beam sizes, there is a trend to reduce the undulator gaps to less than 4 mm as in the case of SACLA and SwissFEL. Together with short-period IVUs and beam energies of 5.8 GeV, radiation in the Angström regime and below can be produced. However, a severe limitation for small-gap PM undulators is radiation damage, which is caused by halo electrons and secondary particles colliding with residual gas and vacuum-chamber atoms. Such radiation damage has been observed already in conventional warm accelerators based on copper technology with repetition rates of typically 50–120 Hz but will become much more important for projects using SC accelerators such as the European XFEL or LCLS II, where repetition rates up to the MHz range will be used. Obviously the protection of undulators from radiation damage will be an important challenge for these new projects. It will require elaborate countermeasures such as collimators, doglegs, loss detectors and active protection systems to prepare a well-collimated beam without any contamination from particles outside an allowed, well-defined phase space which may hit the vacuum chamber. Here the minimum gap is a critical parameter.

Once the technology is mature enough SC undulators might be a good choice: at given gap and λ_0 they offer higher fields or alternatively they offer the same field at a significantly larger gap than PM devices. In addition, the superconductor material NbTi is believed to be much less sensitive to radiation damage than PM material. There are, however, no direct comparative measurements yet.

At present, autumn 2016, PM technology is the method of choice for all applications requiring short period length. This may change once SC technology is mature enough to be used routinely on large-scale systems for SASE FELs. This will require time and stimulation by the requirements of new projects such as the APS upgrade or future LCLS II extensions.

References

[1] H. Onuki and P. Elleaume, *Undulators, Wigglers and their Applications* (Taylor & Francis, New York, 2003). https://doi.org/10.4324/9780203218235

[2] J. Clarke, *The Science and Technology of Undulators and Wigglers* (Oxford University Press, Oxford, New York, 2004).

[3] S. Krinsky, M.L. Perlman and R.E. Watson, in *Handbook of Synchrotron Radiation*, Eds. Y. Farges and E.E. Koch (North Holland, Amsterdam, 1983), Vol. 1, Chap. 2.

[4] R.P. Walker, *Nucl. Instrum. Methods A* **335** (1993) 328. https://doi.org/10.1016/0168-9002(93)90288-S

[5] Y. Li, B. Faatz and J. Pflueger, *Phys. Rev. ST Accel. Beams* **11** (2008) 100701. https://doi.org/10.1103/PhysRevSTAB.11.100701

[6] P. Elleaume, J. Chavanne and B. Faatz, *Nucl. Instrum. Methods A* **455** (2000) 503. https://doi.org/10.1016/S0168-9002(00)00544-1

[7] K. Halbach, *Nucl. Instrum. Methods* **187** (1981) 109. https://doi.org/10.1016/0029-554X(81)90477-8

[8] K. Halbach, *J. Phys. Colloques* **44**(C1) (1983) C1-211. https://doi.org/10.1051/jphyscol:1983120

[9] P. Elleaume, O. Chubar and J. Chavanne, Computing 3D magnetic field from insertion devices, Proc. PAC97 Conf., May 1997, pp. 3509–3511. The code is available under http://www.esrf.eu/Accelerators/Groups/InsertionDevices/Software/Radia

[10] T. Tanaka, T. Seike and H. Kitamura, Measurement of Spring-8 XFEL undulator prototype with the SAFALI system, FEL2009 Proc., 24–29 August 2009, Gyeongju, Korea; see also T. Tanaka, Spring8 EFST on Insertion Devices Workshop, 3–4 December 2009, Shanghai, China, http://www.sinap.ac.cn/meeting/EFST2009/142nd-EFST/EFST_TTanaka.pdf

[11] S. Sasaki, K. Miyata and T. Takada, *Jpn J. Appl. Phys.* **31** (1992) 1794. https://doi.org/10.1143/JJAP.31.L1794

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Linear Accelerator Technology

D. Alesini
Laboratori Nazionali di Frascati dell'INFN, Frascati, Italy

Abstract
The expression 'linear accelerator technology' usually addresses all technological topics related to linear accelerators' (linac) realization. These topics comprise, for instance, accelerating structures, Radio Frequency (RF) sources, magnets, diagnostics systems, controllers, power supplies, vacuum systems, particle sources, etc. Obviously, the extent of all these subjects is too wide to be covered in a single one-hour lecture and, therefore, this paper will only present an overview of the RF structures (including their fabrication technology) with a few examples of particle sources, RF sources and waveguide components. Moreover, the focus will be on linac technology for Free-Electron Lasers (FELs) and Energy Recovery Lasers (ERLs), which are electron accelerators. Electrons, in fact, even at very low energies (a few MeV) travel very close to the speed of light, and this fact has a direct consequence in the accelerating structures design, since they can be identical all along the linac.

Keywords:
Radio Frequency; particle accelerators; cavity; electromagnetic field; superconductivity; linac.

1 Accelerating structures

To accelerate charged particles, Electromagnetic (EM) fields are used. This method is effective only if, however, there is a component of the electric field along the direction of propagation of the particles. In common practice, this particular configuration is obtained with two different solutions, schematically represented in Fig. 1.

i) Using Standing Wave (SW) TM_{010}-like modes in a resonant cavity (or multiple resonant cavities), in which the beam's passage in the cavities is synchronous with the resonating field. The electric field oscillates in each cell at a given frequency (from hundreds of megahertz to several gigahertz) and it is synchronous with the bunches' passage in each cell, so that the beam will experience a net accelerating gradient. A single coupler feeds the structure, as will be described below.

ii) Using Travelling Wave (TW) TM_{01}-like modes in a disk-loaded structure, in which the RF wave co-propagates with the beam with a phase velocity equal to the beam velocity ($\sim c$ for relativistic electrons). The power flows into the structure through an input coupler, and is collected and dissipated at the end of the section by an output coupler connected to an RF load, to avoid reflections that can strongly affect the beam dynamics and might damage the RF power source, as will be described below.

The structures can be made of different materials: copper (Cu) for Normal Conducting (NC) cavities and niobium (Nb) for Superconducting (SC) cavities; and are powered by RF generators (e.g. klystrons, solid-state amplifiers). The choice between the NC and the SC technologies depends on the machine's required performance:

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.79

i) Average accelerating field.

ii) RF pulse length, which is directly related to the bunch train length that it is possible to accelerate: from a few μs to hundreds of ms up to a continuous wave (CW) operation.

iii) Repetition rate (i.e. the number of RF pulses that feed the structures in one second). In this regard, it is useful to define the Duty Cycle (DC) as the ratio between the RF pulse length and its period. Low DC accelerators (DC=10^{-3} to 10^{-5}) operate in a pulsed regime with very short RF pulses (~μs) and a relatively low repetition rate (10 Hz to 100 Hz). On the contrary, high DC machines (DC>10^{-3}) operate with long RF pulses (>ms) up to CW operation (i.e. DC=1).

iv) Average beam current (from μA to hundreds of mA).

v) Available space, etc.

Typically, NC structures are employed in pulsed low DC Free-Electron Laser (FEL) linacs while SC structures are used for long RF pulses and high DC FEL linacs or Energy Recovery Lasers (ERLs).

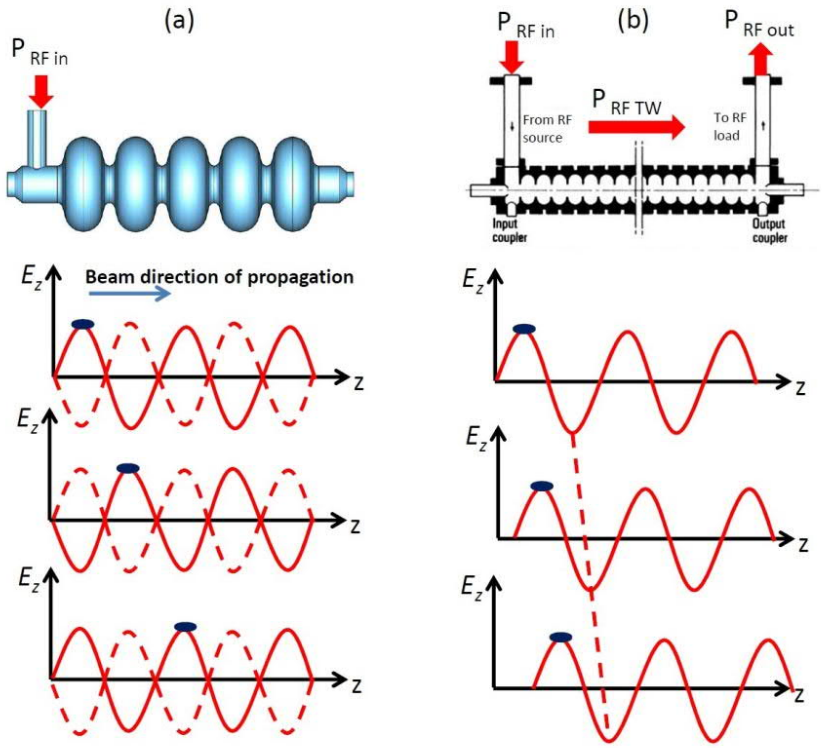


Fig. 1: Acceleration principle using (a) SW or (b) TW structures. In a SW cavity the cell length has to be equal to half the wavelength of the RF field (for π-mode structures). In this way, the bunch in each cell is always synchronous with the electric field positive half-wave. In a TW structure the EM field travels together with the bunch, continuously transferring its energy to the particles. If the bunch velocity matches the EM wave velocity this transfer is obviously maximized.

2 Standing wave cavities

Before comparing different accelerating technologies, it is necessary to define a set of parameters useful for characterizing the various structures. Let us start with SW cavities.

These are metallic, closed volumes where the EM field has very well-defined spatial configurations (resonant modes) whose components, including the accelerating field on axis (E_z) rigidly oscillates at a specific frequency called the resonant frequency (f_{res}). The modes are excited coupling RF generators to the cavity through waveguides or coaxial cables. The resonant modes are called SW modes because they have a spatial configuration that is fixed, and oscillate in time. For a SW cavity the accelerating field on the axis can be written as

$$E_z(z,t) = E_{\mathrm{RF}}(z)\cos\left(\underbrace{2\pi f_{\mathrm{RF}}}_{\omega_{\mathrm{RF}}} t + \varphi\right) = \mathrm{Real}\left[\tilde{E}_z(z)e^{j\omega_{\mathrm{RF}}t}\right] \tag{1}$$

where f_{RF} is the excitation frequency of the generator equal to (or close to) the resonant frequency of the cavity, ω_{RF} is the angular excitation frequency, $E_{\mathrm{RF}}(z)$ (or the phasor $\tilde{E}_z$) is a real (complex) function related to the spatial configuration of the mode.

For a pure cylindrical structure (usually called a 'pillbox cavity') the first accelerating mode (i.e. the mode with non-zero longitudinal electric field on the axis) is the TM_{010} mode. It has a well-known analytical solution from Maxwell's equations, and its spatial configuration is given in Fig. 2. For this mode the electric field has a longitudinal component only, while the magnetic one is purely azimuthal. The corresponding complex phasors are given by [1, 2]

$$\begin{cases} \tilde{E}_z = AJ_0\left(p_{01}\dfrac{r}{a}\right) \\ \tilde{H}_\theta = -jA\dfrac{1}{Z_0}J_0'\left(p_{01}\dfrac{r}{a}\right) \end{cases} \tag{2}$$

where a is the cavity radius, A is the mode amplitude and p_{01} (= 2.405) is the first zero of the Bessel function $J_0(x)$. The resonant frequency of this mode is given by: $f_{\mathrm{res}} = \dfrac{p_{01}c}{2\pi a}$.

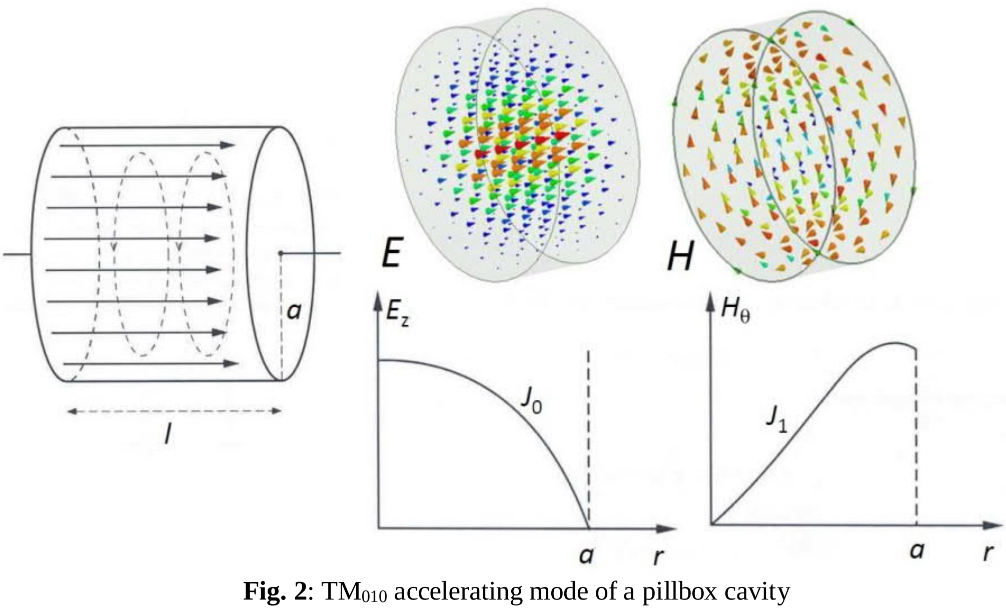


Fig. 2: TM_{010} accelerating mode of a pillbox cavity

The geometry of real cylindrical cavities is somewhat different to that of a pillbox. In fact, one has to also consider the perturbation introduced by the beam pipe, the power couplers to RF generators and any antenna or pick-up used to monitor the accelerating field inside the cavity. For this reason, the actual accelerating mode is called the TM_{010}-like mode.

Nowadays, real cavities and their couplers to the RF generators are designed using numerical codes that solve the Maxwell equations with the proper boundary.

A sketch of a cavity fed through a loop or a coaxial probe to an external generator is given in Fig. 3, where there are also reported the electric and magnetic field lines and the longitudinal electric field profile on the axis. Details of the coupler design can be found in [3].

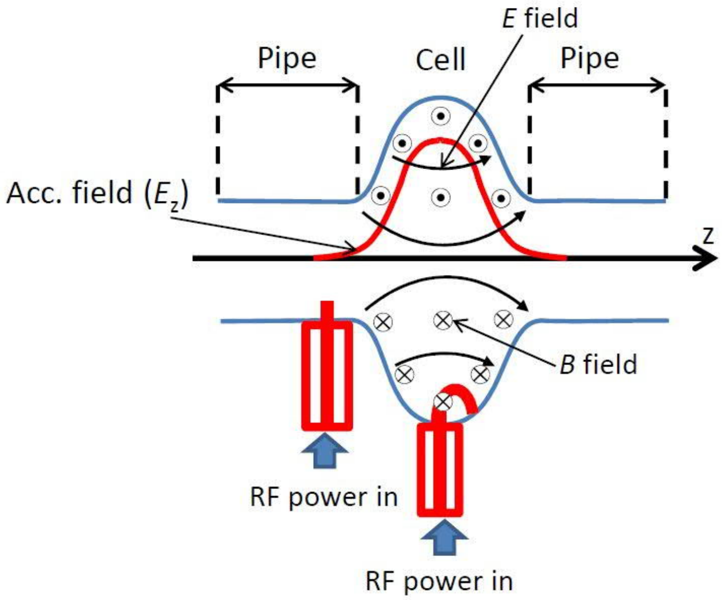


Fig. 3: Sketch of a real cavity operating on the TM_{010}-like mode with two type of coaxial couplers

2.1 Standing wave cavity parameters

To define the cavity parameters, we suppose that it is excited with a pure sinusoidal tone at f_{RF}. The maximum energy gain per unit charge (i.e. the accelerating voltage V_{acc}) of a particle crossing the cavity at a velocity v is obtained by integrating the time-varying accelerating field sampled by the charge along the trajectory

$$V_{acc} = \left| \int_{cavity} \tilde{E}_z(z) e^{j\omega_{RF}\frac{z}{v}} dz \right| \tag{3}$$

Real cavities have losses. Surface currents experience a surface resistance R_s and dissipate energy, so that a certain amount of RF power must be provided from outside to keep the accelerating field at the desired level. The total average dissipated power P_{diss} is given by

$$P_{diss} = \int_{\substack{cavity\\wall}} \overbrace{\frac{1}{2} R_s \left|\tilde{H}_{tan}\right|^2}^{\text{power density}} dS \tag{4}$$

where $\tilde{H}_{tan}$ is the tangential magnetic field components on the surface. The value of the surface resistance depends on several factors, such as frequency and cavity material. As an example, at 1 GHz for a NC cavity in copper $R_s \cong 8$ mΩ while, for a SC cavity made in niobium at 2 K, $R_s \cong 10$ nΩ.

For a SW cavity the first figure of merit is the shunt impedance defined by

$$R = \frac{V_{acc}^2}{P_{diss}} \qquad [\Omega] \tag{5}$$

This parameter qualifies the efficiency of the cavity: the higher its value, the larger is the achievable accelerating voltage for a given dissipated power. Traditionally, it is the quantity to optimize in order to maximize the accelerating field for a given dissipated power. As an example, at 1 GHz for a NC single cell cavity a typical shunt impedance is of the order of 2 MΩ, while a SC single cell cavity at the same frequency is of the order of 1 TΩ, due to the extremely lower dissipated power.

The total energy (W) stored in the cavity is given by

$$W = \int_{\substack{\text{cavity}\\ \text{volume}}} \overbrace{\left(\frac{1}{4}\varepsilon\left|\vec{E}\right|^2 + \frac{1}{4}\mu\left|\vec{H}\right|^2\right)}^{\text{energy density}} dV \tag{6}$$

The quality factor of the accelerating mode is then defined by the ratio of the cavity stored energy and the dissipated power on the cavity walls

$$Q_0 = \omega_{RF}\frac{W}{P_{diss}} \tag{7}$$

For a NC cavity operating at 1 GHz the quality factor is of the order of 10^4 while, for an SC cavity, values of the order 10^9 to 10^{10} can be achieved.

It can be easily demonstrated that the ratio R/Q is a pure geometric factor and it does not depend upon the cavity wall conductivity or operating frequency. This is the reason why it is always taken as a geometric design qualification parameter. The R/Q of a single cell is of the order of 100.

2.2 Standing wave cavity equivalent circuit

The quantities described above play crucial roles in the evaluation of the cavity performances. Let us consider a cavity powered by a source (klystron) at a constant frequency f_{RF} in CW and at a fixed power (P_{in}) as shown in Fig. 4. It can be demonstrated that the equivalent circuit of this system is that of a parallel RLC resonant circuit, as reported in Fig. 4. In the equivalent circuit the resistance R is exactly the shunt impedance of the cavity, and the quality factor is the quality factor of the RLC circuit. The transformer models the coupling between the waveguide and the cavity. With simple calculations, it is easy to demonstrate that the maximum accelerating voltage V_{acc} for a given input power P_{in} is given by

$$V_{\mathrm{acc}} = \frac{\dfrac{2\sqrt{\beta}}{(1+\beta)}}{\sqrt{1+\left(Q_{\mathrm{L}}\delta\right)^2}}\sqrt{RP_{\mathrm{in}}} \tag{8}$$

where β is the generator–cavity coupling coefficient [3], $Q_{\mathrm{L}} = Q_0/(1+\beta)$ is the loaded quality factor and $\delta = \left(f_{\mathrm{RF}}/f_{\mathrm{res}} - f_{\mathrm{res}}/f_{\mathrm{RF}}\right)$.

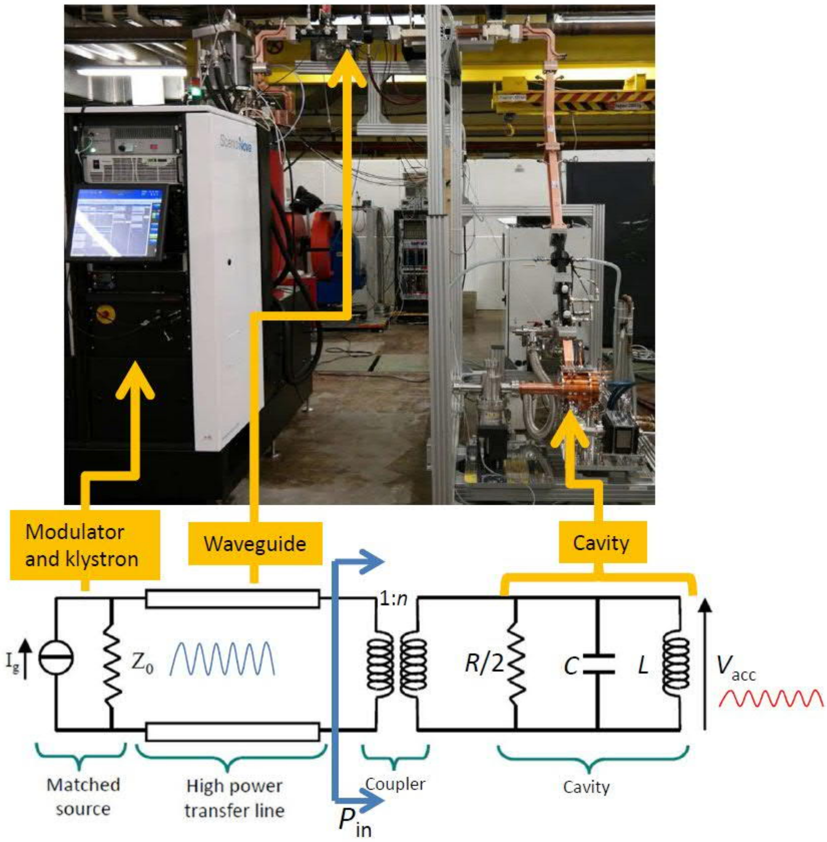


Fig. 4: Equivalent circuit of a cavity fed by a generator

The dissipated power into the cavity is given by:

$$P_{\text{diss}} = \frac{\dfrac{4\beta}{(1+\beta)^2}}{1+(Q_{\text{L}}\delta)^2} P_{\text{in}} \tag{9}$$

The plot of the accelerating voltage as a function of the excitation frequency is given in Fig. 5 for three different values of the cavity quality factor supposing P_{in} = 1 MW, R/Q = 100 and β = 1 (critical coupling), and f_{res} = 1 GHz.

From previous formulas we easily see that, at the resonant frequency and in the case β = 1, we have

$$V_{\text{acc}} = \sqrt{RP_{\text{in}}} = \sqrt{\left(\frac{R}{Q}\right) Q_0 P_{\text{in}}} \tag{10}$$

This means that, for a given cavity, the accelerating voltage is proportional to $\sqrt{Q_0}$.

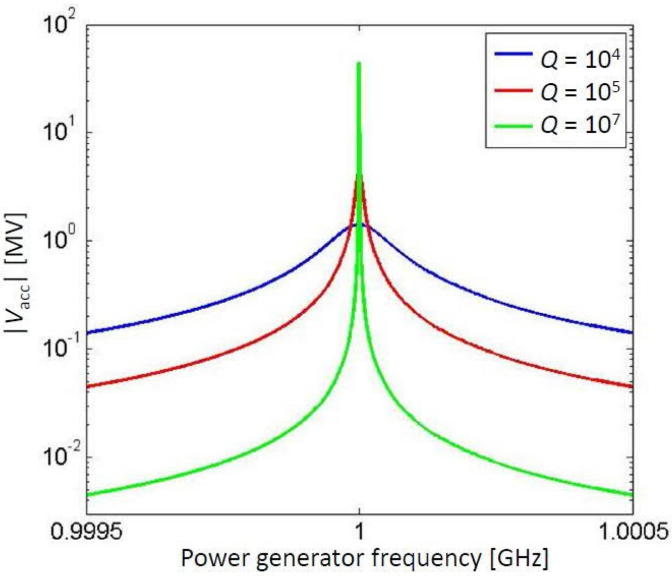


Fig. 5: Accelerating voltage as a function of the excitation frequency for a cavity with $R/Q = 100$, $\beta = 1$, $f_{res} = 1$ GHz, supposing that $P_{in} = 1$ MW.

On the other hand, the bandwidth of the resonance, defined as the δ-frequency interval corresponding to the points with an average dissipated power in the cavity of a factor of two lower than the dissipated power at resonance, is inversely proportional to the cavity quality factor according to the formula

$$\frac{\Delta f_{RF}\big|_{3dB}}{f_{res}} = \frac{1}{Q_L} \qquad \Rightarrow \begin{cases} \Delta f_{RF}\big|_{3dB}\Big|_{NC} \cong 100kHz \\ \Delta f_{RF}\big|_{3dB}\Big|_{SC} < 1Hz \end{cases} \tag{11}$$

The bandwidth of the cavity is, also, labeled 3 dB bandwidth because, usually, we refer to the normalized quantity $10\log_{10}[P_{diss}(f)/P_{diss}(f_{res})]$: the bandwidth correspond to the frequency interval related to the −3 dB points below the peak. In other words, the dissipations and the external coupling cause the cavity to oscillate in a band of frequencies ($\Delta f\big|_{3dB} = f_{RF}/Q_L$) whose width is a function of the loaded quality factor.

Let us now consider a cavity powered by a source (klystron) in pulsed mode at a frequency $f_{RF} = f_{res}$. If we suppose that the generator is switched on at time $t = 0$ with a peak power P_{in} we obtain the following expressions for the accelerating voltage, and dissipated and reflected powers [4]

$$V_{acc}(t) = \frac{2\sqrt{\beta}}{(1+\beta)}\left(1-e^{-\frac{t}{\tau}}\right)\sqrt{RP_{in}} \xrightarrow{\beta=1} V_{acc}(t) = \left(1-e^{-\frac{t}{\tau}}\right)\sqrt{RP_{in}}$$

$$P_{diss}(t) = P_{in}\frac{4\beta}{(1+\beta)^2}\left(1-e^{-\frac{t}{\tau}}\right)^2 \xrightarrow{\beta=1} P_{diss}(t) = P_{in}\left(1-e^{-\frac{t}{\tau}}\right)^2$$

$$P_{refl}(t) = P_{in}\left[\left(1-e^{-\frac{t}{\tau}}\right)\frac{2\beta}{(1+\beta)}-1\right]^2 \xrightarrow{\beta=1} P_{refl}(t) = P_{in}e^{-\frac{2t}{\tau}} \tag{12}$$

where we define the filling time $\tau = 2Q_L/\omega_{res}$.

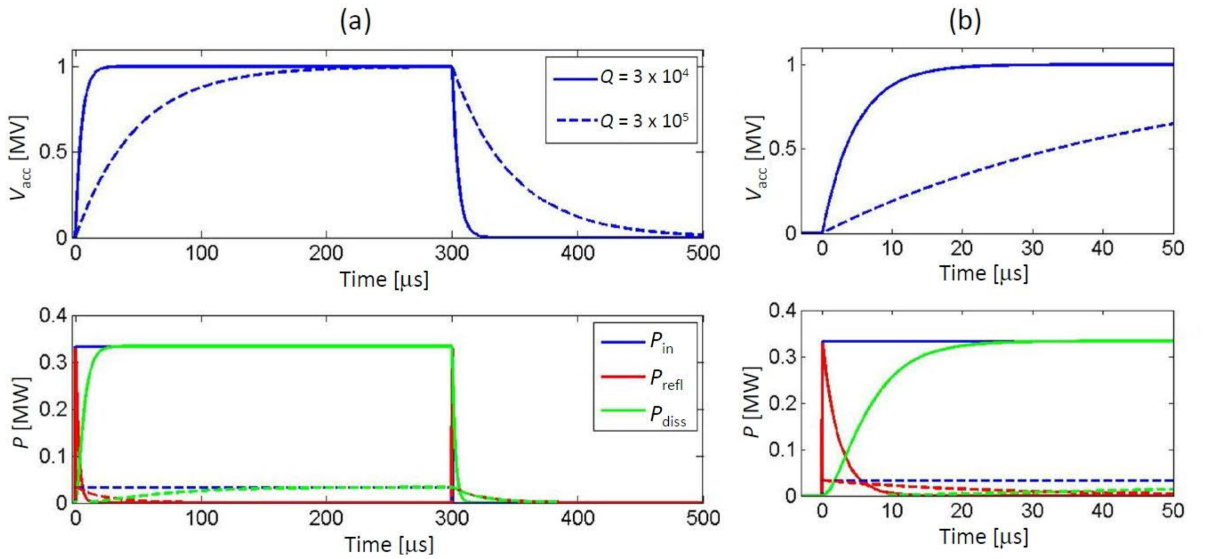


Fig. 6: Accelerating voltage and dissipated and reflected powers as a function of time for two different values of the cavity quality factors, $R/Q = 100$, $\beta = 1$, $f_{res} = 1$ GHz and an accelerating voltage in the regime $V_{acc} = 1$ MV.

The behaviour for a 1 GHz NC cavity with a quality factor $Q_0 = 3 \times 10^4$ is given in Fig. 6, where the accelerating voltage and the dissipated and reflected powers are given as a function of time, assuming $\beta = 1$. In the plot we fixed the accelerating voltage at a regime equal to $V_{acc} = 1$ MV and we have calculated from Eqs. (9) and (12) the needed input power to reach this value. In the same plot we also reported, for reference, a quality factor 3×10^5 that is still, at least, four orders of magnitude lower than the Q-factor of a typical superconducting structure. The plots and the previous formulas clearly show the following important results:

i) The input power we need to reach the desired voltage is inversely proportional to the quality factor of the cavity according to

$$P_{in} = \frac{V_{acc}^2}{\left(\frac{R}{Q}\right)} \frac{1}{Q_0} \propto \frac{1}{Q_0} \tag{13}$$

and the dissipated power into the structure follows the same scaling. This means that, as example, to reach a 1 MV accelerating voltage with a NC cavity at 1 GHz we need an input power of the order of a few hundreds of kilowatts, while for a superconducting cell it scales to a few watts.

ii) There is a peak of reflected power back to the generator at the beginning (and at the end) of the input pulse that requires protection for the generator itself to avoid damage.

iii) The voltage in the cavity grows with a filling time proportional to the quality factor of the cavity $\tau = 2Q_L/\omega_{res}$.

Typical filling times for NC cavities are of the order of a microsecond while for a SC cavity they are hundreds of milliseconds. This is also the reason why it is difficult to represent, in the same plot, a NC cavity and an SC one; and we have chosen a value of $Q = 3 \times 10^5$ instead of 10^9 to 10^{10}.

2.3 Multi-cell standing wave cavities

In a multi-cell structure, there is one RF input coupler that feeds a system of coupled cavities as sketched in Fig. 1(a). The field of adjacent cells is coupled through the cell irises (and/or through properly designed coupling slots). It is quite easy to demonstrate that the shunt impedance is N times the impedance of a single cavity; moreover, with one source, it is possible to feed a set of cavities with a simplification of the power distribution system and layout. On the other hand, the fabrication of multi-cell structures is more complicated than single-cell cavities.

The N-cell structure behaves like a system of N coupled oscillators with N coupled multi-cell resonant modes. As an example, the field configuration of a two-cell resonator is shown in Fig. 7. The mode in which the two cells oscillate with the same phase is called 0 mode, while the one with 180° phase shift is called π-mode. It is possible to demonstrate that the most efficient configuration (generally used for acceleration) is the π-mode, which has been shown in Fig. 1(a) for a system of five cells.

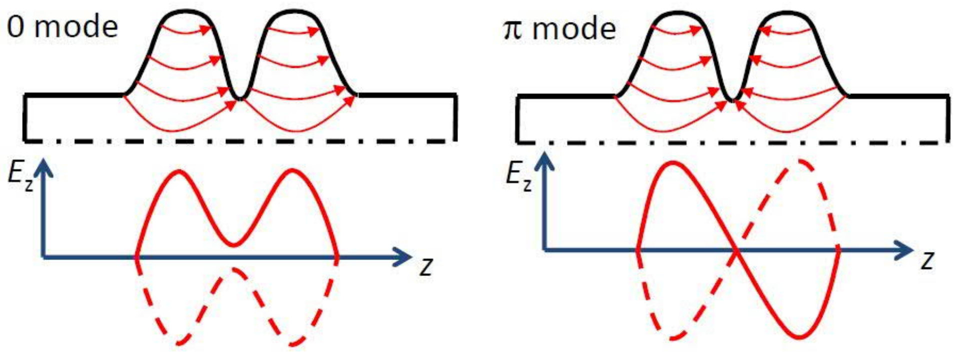


Fig. 7: Resonant mode in a system of two coupled cavities. The mode typically used for acceleration is the π-mode.

In order to have a synchronous acceleration in each cell, the distance (d) between the centre of two adjacent cells has to be $d = v/(2f_{\mathrm{RF}})$ where v is the particle velocity.

Field amplitude variation from cell to cell should be also small, to maximize the acceleration efficiency. This requires a careful realization procedure, which is sometimes not sufficient to reach field flatness below a few percent, and thus requires a tuning process after fabrication.

It is possible to demonstrate that, over a certain number of coupled cavities, the overlap between adjacent modes can introduce problems for field equalization, and this limits the maximum number of multi-cell structures to around 10. It is possible to overcome this limit, for example working in the π/2 mode, but typically these types of structures are not used in FELs or ERLs and their description is beyond the purpose of this paper (further details can be found in the literature, for example in [4]).

For multi-cell structures, it is also useful to introduce another parameter r, the shunt impedance per unit length, simply given by

$$r = \frac{R}{L} = \frac{\left(V_{\mathrm{acc}}/L\right)^2}{P_{\mathrm{diss}}/L} = \frac{E_{\mathrm{acc}}^2}{p_{\mathrm{diss}}} \qquad \left[\frac{\Omega}{m}\right] \tag{14}$$

where L is the total structure length, E_{acc} is the average accelerating field and p_{diss} is the dissipated power per unit length.

As an example, the SC cavities of the European X-FEL operate with modules of nine cells, as sketched in Fig. 8. The parameters of these cavities are given in Table 1. More details can be found in [5, 6].

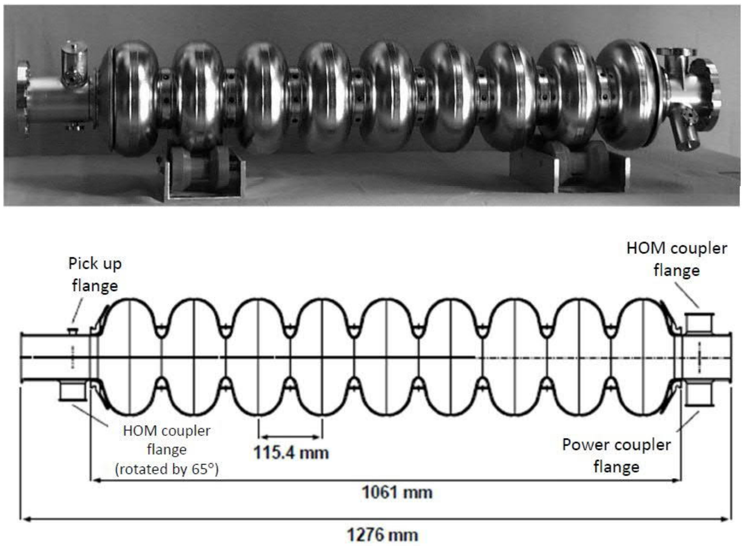


Fig. 8: Multi-cell superconducting cavity of the European X-FEL (courtesy [6])

Table 1: XFel SC cavity parameters

Parameter	**Value**
Frequency of operation	1.3 GHz
Mode of operation	π
R/Q	1036 [Ω]
Q_0	10^{10}
Q_L	4.6×10^6
Number of cells	9
Active length	1.038 [m]
Accelerating gradient	23.6 [MV/m]
RF pulse length	1.4 [ms]
Repetition rate	10 [Hz]
Iris diameter	70 [mm]
E_{peak}/E_{acc}	2
B_{peak}/E_{acc}	4.26 [mT/(MV/m)]

2.4 Beam structure

The discussion so far allows a better understanding of the relation between the beam structure and the type of accelerating sections. RF structures are, in general, fed by pulses, and each pulse includes several RF periods as reported in Fig. 9(a). As already remarked, the duty cycle is defined as the ratio between the pulse width and the period. The corresponding accelerating voltage in the cavity is sketched in Fig. 9(b). To avoid energy modulation along the bunch train, the electrons can be accelerated only in the flat part of the voltage pulse (in the case of high beam currents, injection during the voltage slope, or RF gymnastics, is also possible to compensate for the so-called beam loading effects, but the discussion of these techniques is beyond the scope of the present paper).

From consideration of the paragraphs above, it is straightforward to see that high DC operation (up to CW operation) is feasible only with SC structures, while low DC operation (DC < 10^{-3}) is also feasible with NC structures. The reasons are clearly due to the power available from RF sources, power dissipation and filling times of the structures. In the case of SC, in fact, long RF pulses are required to fill the cavities (the higher the quality factor, the higher the filling time), but relatively low average power is required from the RF source to reach a given gradient, since a small amount of power is dissipated into the structures. In the second case, instead, short RF pulses (~μs) with high peak power are used but, in order to limit the average power required from the RF generator and dissipated into the structure, the overall DC has to be kept small. This implies that the maximum number of pulses per second has to be lower than a few hundred.

From what has been highlighted in the paragraphs above it follows that multi bunch operation (from hundreds of bunches up to CW) is preferred with SC structures, because of the long pulses used.

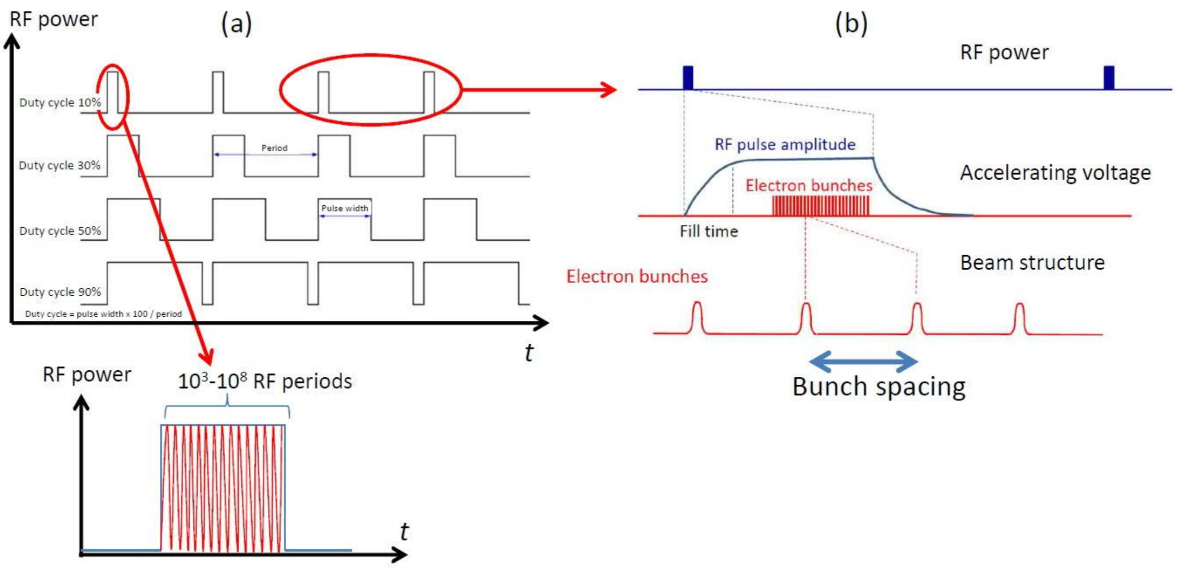


Fig. 9: (a) Sketch of the input power into RF structures and (b) related beam structure

3 Travelling wave structures

As pointed out in the first paragraph, there is another possibility for accelerating particles: using a travelling wave (TW) section. The RF wave is co-propagating with the beam with a phase velocity equal to the beam velocity (c).

In other words, a TW structure is a special waveguide in which the phase velocity of the wave matches the particle velocity. Only in this case, the beam can efficiently absorb energy from the EM wave, and results in continuous acceleration. The solutions of Maxwell's equations for an EM wave propagating into a constant cross-section waveguide, however, give a phase velocity that is always larger than the speed of light. Thus, such an EM wave will never be synchronous with a particle beam. For instance, if we consider a circular waveguide (Fig. 10(a)), it turns out that the first propagating mode with $E_z \neq 0$ is the TM_{01} mode whose longitudinal electric field can be expressed by the well know formula [7]

$$E_z\big|_{TM_{01}} = E_0 J_0\left(\frac{p_{01}}{a} r\right)\cos\left(\omega_{RF} t - k_z^* z\right) \tag{15}$$

where a is the radius of the waveguide and k_z^* is the propagation constant given by

$$k_z^* = \frac{1}{c}\sqrt{\omega_{RF}^2 - \omega_{cut}^2} \tag{16}$$

where ω_{cut} cut-off angular frequency of the waveguide equal to $\omega_{cut} = cp_{01}/a$.

The corresponding phase velocity is given by

$$v_{\mathrm{ph}} = \frac{\omega_{\mathrm{RF}}}{k_z^*} = \frac{c}{\sqrt{1-\omega_{\mathrm{cut}}^2/\omega_{\mathrm{RF}}^2}} \tag{17}$$

which is always larger than c.

The behaviour of the propagation constant as a function of frequency is the well-known dispersion curve, and is sketched in Fig. 10(a). It is important to remark that the phase velocity is not the velocity of the energy propagation into the structure, which, instead, is given by the group velocity (v_g):

$$v_{\mathrm{g}} = \left.\frac{d\omega}{dk_z}\right|_{\omega=\omega_{RF}} = c\sqrt{1-\omega_{\mathrm{cut}}^2/\omega_{\mathrm{RF}}^2} \tag{18}$$

and is always smaller than c.

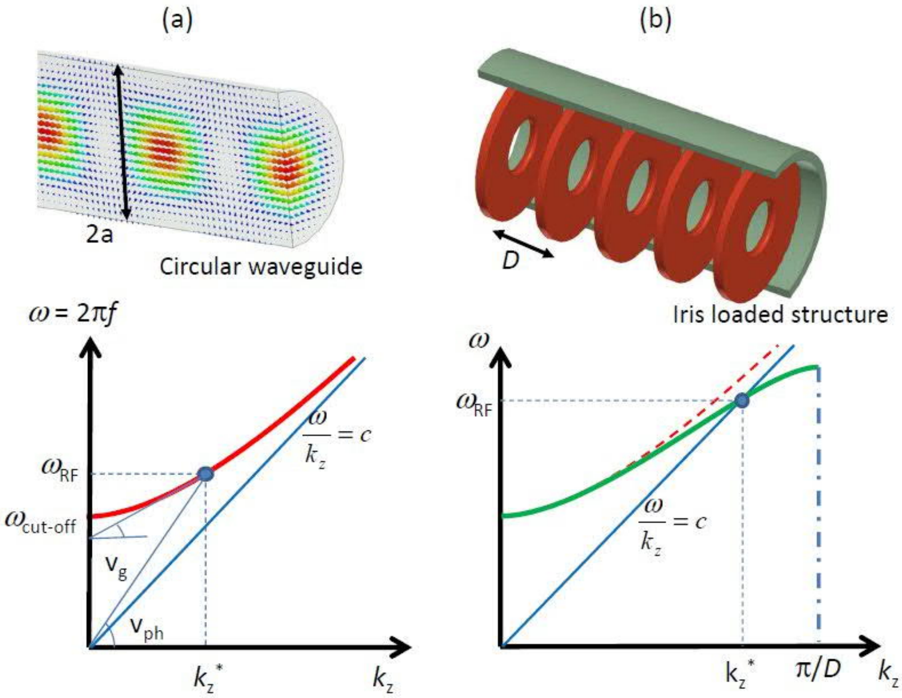


Fig. 10: Typical dispersion curve for (a) a circular waveguide and (b) an iris-loaded structure

In order to slow down the wave phase velocity, the structure through which the wave is travelling is periodically loaded with irises. A sketch of an iris-loaded structure is given in Fig 10(b). The field in this type of structure is that of a special wave travelling within a spatial periodic profile (TM_{01}-like mode). The structure can be designed to have the phase velocity equal to the speed of the particles: this allows a net acceleration over large distances.

In particular, the accelerating field can be expressed by Floquet's theorem [4, 7]

$$E_z\big|_{TM_{01\text{-like}}} = \underbrace{E_P(r,z)}_{\substack{\text{periodic function}\\ \text{with period D}}} \cos\left(\omega_{\mathrm{RF}}t - k_z^* z\right) \tag{19}$$

The dispersion curve for this type of structure is given in Fig. 10(b) and shows that, at a given frequency, the phase velocity can be equal to (or even slower than) c.

In a TW structure, the RF power enters into the cavity through an input coupler and flows (travels) through the cavity in the same beam direction; and an output coupler, at the end of the structure, connected to a matched power load, absorbs the non-dissipated power avoiding reflections, as sketched in Fig. 1(b). If there is no beam, the input power simply dissipates on the cavity walls and the remainder is finally dissipated into the power load. In the presence of a beam current a fraction of this power is, indeed, transferred to the beam.

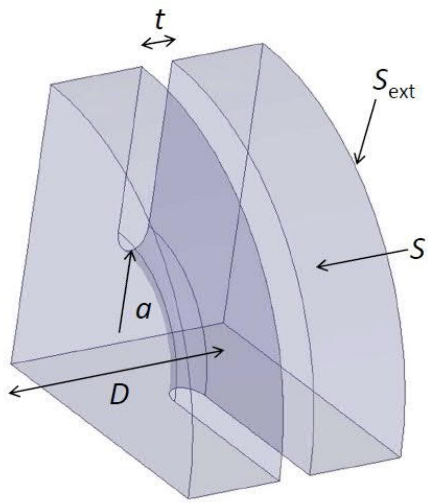


Fig. 11: Sketch of a single cell of a TW structure

3.1 Travelling wave structure parameters

Similarly to what has been done for SW cavities, it is possible to define some figures of merit of TW structures as well. Referring to the single cell sketched in Fig. 11, we can consider the following quantities:

$$V_z = \left| \int_0^D E_z \cdot e^{j\omega_{RF}\frac{z}{c}} dz \right| \qquad \text{single cell accelerating voltage } [V]$$

$$E_{acc} = \frac{V_z}{D} \qquad \text{average accelerating field in the cell } [\frac{V}{m}]$$

$$P_{in} = \int_{\text{Section}} \frac{1}{2} \mathrm{Re}\left(\vec{E} \times \vec{H}^*\right) \cdot \hat{z} dS \qquad \text{average input power (power flux) } [W]$$

$$P_{diss} = \frac{1}{2} R_s \int_{\substack{\text{cavity}\\\text{wall}}} \left| H_{tan} \right|^2 dS \qquad \text{average dissipated power in the cell } [W]$$

$$p_{diss} = \frac{P_{diss}}{D} \qquad \text{average dissipated power per unit length } [\frac{W}{m}]$$

$$W = \int_{\substack{\text{cavity}\\\text{volume}}} \overbrace{\left(\frac{1}{4} \varepsilon \left| \vec{E} \right|^2 + \frac{1}{4} \mu \left| \vec{H} \right|^2 \right)}^{\text{energy density}} dV \qquad \text{stored energy in the cell } [J]$$

$$w = \frac{W}{D} \qquad \text{average stored energy per unit length } [\frac{J}{m}] \tag{20}$$

The shunt impedance per unit length r is defined in Eq. (14) as

$$r = \frac{E_{acc}^2}{p_{diss}} \qquad \left[\frac{\Omega}{m} \right]$$

The higher the value of r, the higher the available accelerating field for a given RF power. Typical values for a 3 GHz structure are ~60 MΩ/m.

In a purely periodic structure, made by a sequence of identical cells (also called a 'constant impedance structure'), the RF power flux and the intensity of the accelerating field decay exponentially along the structure. The process of power filling is represented in Fig. 12. When the RF generator is switched on, the power starts flowing into the structure with a velocity equal to the group velocity, which is typically a small fraction of the velocity of light (a few percent). The time necessary to propagate the RF wavefront from the input coupler to the end of a section of length L, referred to as the filling time, is given by

$$\tau_{\mathrm{F}} = \frac{L}{v_{\mathrm{g}}} \qquad [\mathrm{s}] \tag{21}$$

where the group velocity can be calculated by

$$v_{\mathrm{g}} = \frac{P_{\mathrm{in}}}{w} \qquad \left[\frac{\mathrm{m}}{s}\right] \tag{22}$$

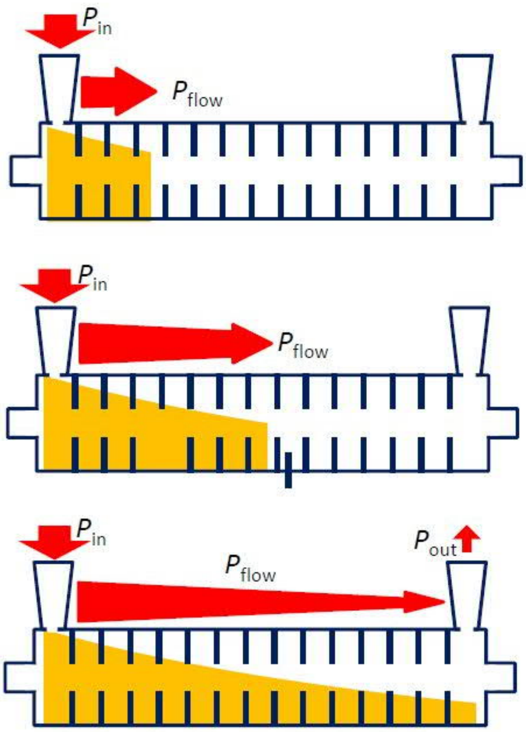


Fig. 12: Sketch of the power filling of a TW structure

After one filling time the structure is completely full of energy.

Increasing the group velocity allows reduction of the duration of the RF pulse powering the structure. Since $w \propto E^2$, however, a low group velocity is preferable to increase the effective accelerating field for a given power flowing in the structure. It is possible to demonstrate that the group velocity scales as a^3 where a is the iris's aperture.

Due to the power attenuation along the structure, after one filling time the accelerating field can be expressed as

$$E_z\big|_{TM_{01\text{-like}}} = E_P(r,z)\cos\left(\omega_{\mathrm{RF}} t - k_z^* z\right) e^{-\alpha z} \tag{23}$$

where the attenuation constant is given by

$$\alpha = \frac{p_{\text{diss}}}{2P_{\text{in}}} \qquad \left[\frac{1}{\text{m}}\right] \tag{24}$$

The field sampled by a particle entering the structure after one filling time is simply given by

$$E_z(z) = E_P(z)e^{-\alpha z} \tag{25}$$

Due to the periodicity of the structure, this field has an amplitude modulation with period D (in the term $E_P(z)$). Also, for TW structures it is possible to define the quality factor

$$Q = \omega_{\text{RF}} \frac{w}{p_{\text{diss}}} \tag{26}$$

which can, from Eqs. (22) and (24), be related to the attenuation constant α

$$Q = \frac{\omega_{\text{RF}}}{2v_g\alpha} \tag{27}$$

Each structure is identified by a field phase advance per cell given by: $\Delta\varphi = k_z^* \text{D}$. For several reasons well-illustrated in [8], one of the most common modes used for acceleration is $2\pi/3$.

As an example, we can consider a 2 m long C-band (f_{RF} = 5.712 GHz) accelerating section made of copper. Using an iris aperture $2a$ = 14 mm, we obtain for the $2\pi/3$ mode typical values of the abovementioned parameters: r = 82 [MΩ/m], α = 0.36 [1/m], v_g/c = 1.7%. This gives a filling time τ_{F} = 150 ns, which is much lower than typical values obtainable with SW structures working at the same frequency. In Fig. 13(a) is shown the power flow along the structure, and the corresponding accelerating field, assuming a pulsed input power of 50 MW. The power dissipated into the structure and the total accelerating voltage integrated by a particle entering into the section at different times are shown in Fig. 13(b). From the figure, it is possible to observe that after one filling time the structure is full of energy, and the integrated voltage sampled by a particle does not change.

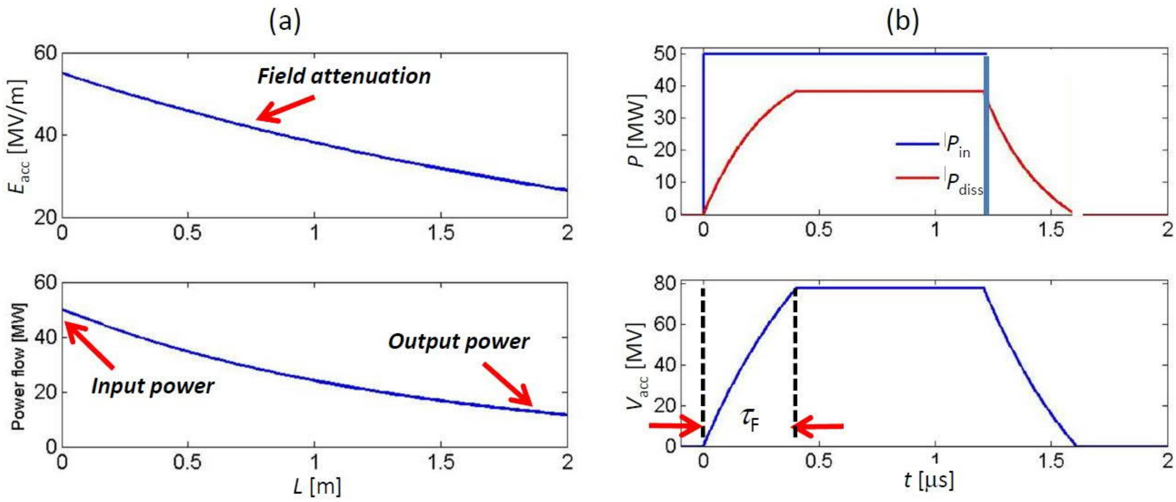


Fig. 13: (a) Accelerating field and power flow; (b) input power, cavity dissipated power and accelerating voltage, for a C-band constant impedance structure working on the $2\pi/3$ mode.

At the end of the structure, the remaining power has to be dissipated into an external load to avoid reflections. These, in fact, can travel back to the RF power unit (causing possible damage), locally increase the peak fields (causing possible discharges) or can produce undesirable perturbations to the beam dynamics.

Typically, TW structures have very short filling time values (<1 μs) and allow operation in pulsed mode at low repetition rate (10 Hz to 100 Hz) with high peak power (tens of MW) and relatively high accelerating field (up to 50 MV/m). Typical DC are very low (10^{-4} to 10^{-5}). For instance, assuming the previously mentioned TW structure being fed by 400 ns RF pulses of 50 MW (P_{in}) at 100 Hz, the resulting DC is 4×10^{-5} with a total average power from the RF generator of 2 kW ($P_{in} \times$ DC). The use of short RF pulses, however, gives the possibility of accelerating just a few bunches per RF bucket.

In conclusion, note the following important remarks.

i) Due to the power dissipation and the consequent reduction of the accelerating field, increasing the TW cavity length too much makes the acceleration process very inefficient. Typically in the S-band (3 GHz) the cavity length is limited to 3 m, in the C-band (6 GHz) to 2 m.

ii) There is no benefit preferring SC materials for TW structure fabrication. This is a direct consequence of the TW working principle, where there is no accelerating field build-up effect limited by ohmic losses, and thus the obtainable gain using SC materials is not relevant. It would be, theoretically, only using a very long TW structure, but this would imply very long RF pulses at high peak voltage (not feasible) together with a dramatic complication of the realization process.

iii) Since the structure is basically a waveguide with irises, at the input port there are no significant power reflections towards the generator, and thus it is possible to connect the power unit directly to the section without circulators/isolators to protect the source.

3.2 Travelling wave constant gradient structures

An example of constant impedance TW structures has been presented in the previous section. Irises of equal dimensions cause the accelerating field to decay exponentially along the section. It is possible to demonstrate that, in order to keep the accelerating field constant in the whole structure, the iris aperture has to be correctly shrunk along the structure [4, 8]. In this way, the field attenuation is compensated for by the increase of the stored energy per unit length, due to the lower group velocity. For instance, in Fig. 14 the iris dimensions and the section parameters for a 100-cell C-band constant gradient and constant impedance structure have been reported as a function of the cell number. In both cases the average accelerating field is the same and equal to 40 MV/m for 50 MW input power. Because of the different iris dimensions, the group velocity and shunt impedance also change along the structure.

In general, constant gradient structures are more efficient than constant impedance structures, because of the more uniform distribution of RF power in the longitudinal direction, but they require a more complicated mechanical realization due to the irises' profile modulation.

As a reference, the parameters of the TW accelerating structures of the SLAC linac [8] (used for the LCLS FEL), the PSI SwissFEL [9–12] and Spring 8 XFEL linac [13, 14] are reported in Table 2. The former operates in the S-band, while the others are in the C-band.

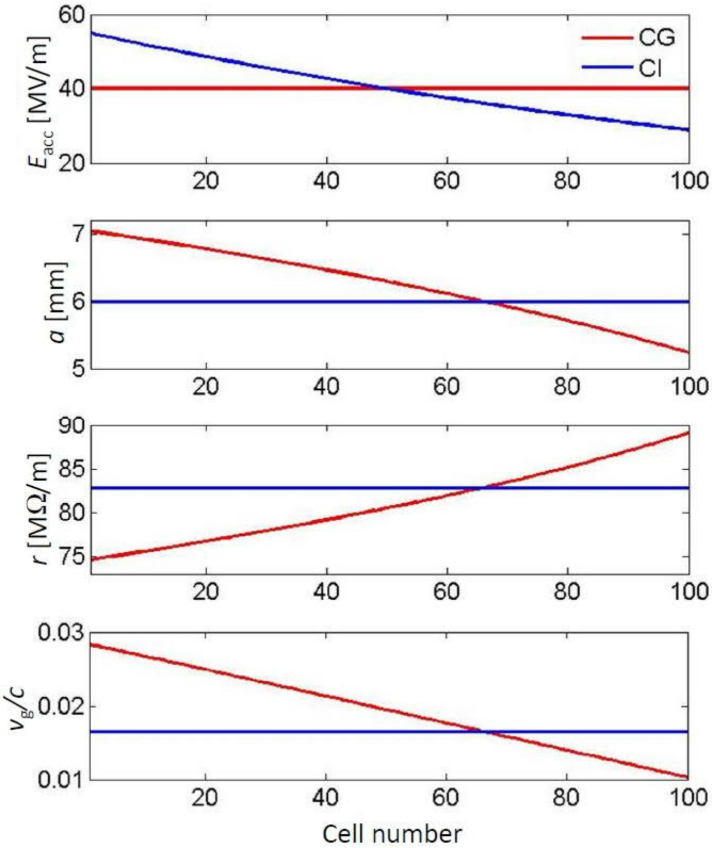


Fig. 14: Iris dimensions and structure parameters for 100-cell C-band constant gradient (red line) and constant impedance (blue line) structures.

Table 2: Parameters of the TW structures of the SLAC linac [8] (used for the LCLS FEL), PSI SwissFEL [9–12] and Spring 8 linacs [13, 14].

Parameter	**LCLS TW structures**	**PSI SwissFEL structures**	**Spring 8**
Operating frequency [GHz]	2.856	5.712	5.712
Operating mode	2π/3	2π/3	3π/4
Structure length [m]	3	2	1.8
Type of structures	CG	CG	Quasi-CG, damped
Number of cells	86	113	91
Accelerating gradient [MV/m]	22	28	38
Filling time [μs]	0.83	0.32	0.3
Cell iris diameter [mm]	26.2 to 19.1	14.4 to 10.9	13.6 to 17.3
Shunt impedance [MΩ/m]	53 to 60	74 to 88	49 to 59
Group velocity [v_g/c]	2.04% to 0.65%	3.08% to 1.21%	2% (av.)
Repetition rate [Hz]	120	100	10 to 60

4 Materials for accelerating structure fabrication

In the previous sections the main characteristics of the accelerating structures with their main figures of merit, properties and geometries have been illustrated. We will now go into the details of linac technology, starting with the material generally used for the structure's realization. As already pointed out the most common alternatives are oxygen-free high conductivity (OFHC) copper for NC cavities, and niobium for the SC ones.

4.1 Oxygen-free high conductivity copper

OFHC copper is the most common material used for NC structures for several reasons:

i) it has a very good electrical (and thermal) conductivity;

ii) it has a low Secondary Emission Yield (SEY) that allows the avoidance of multiple impact electron amplification (multipacting) phenomena [15] during structure power up, conditioning and operation;

iii) it shows good performance at a high accelerating gradient;

iv) it is easy to machine, and a very good roughness (up to the level of a few nm) can be achieved;

v) it can be brazed or welded.

The microwave surface resistance of the copper (as for all metals) is expressed by

$$R_s = \sqrt{\frac{\pi f_{\mathrm{RF}} \mu_0}{\sigma}} = \frac{1}{\sigma\delta} \tag{28}$$

where σ is the conductivity (~5.8 × 10^7 S/m for Cu at 20°C) and δ is the skin depth, which represents the penetration of the EM field and surface currents inside the metal, as sketched in Fig. 15(a), given by

$$\delta = \frac{1}{\sqrt{\pi f_{\mathrm{RF}} \mu_0 \sigma}} \tag{29}$$

The behaviour of the surface resistance and skin depth as a function of frequency is shown in Fig. 15(b). The conductivity increases reducing the temperature. In the DC regime, for instance, it can be more than a factor of 100 higher (depending on the copper purity) at cryogenic temperatures with respect to room temperature. At cryogenic temperatures (<40 K) and in the RF regime, however, a mechanism called the 'anomalous skin effect' [16] takes place. This reduces the gain to a factor of about 20 (depending on the working frequency and copper purity). This also translates into a reduction of the gain in the quality factor, which makes the use of copper at cryogenic temperatures neither practical nor convenient.

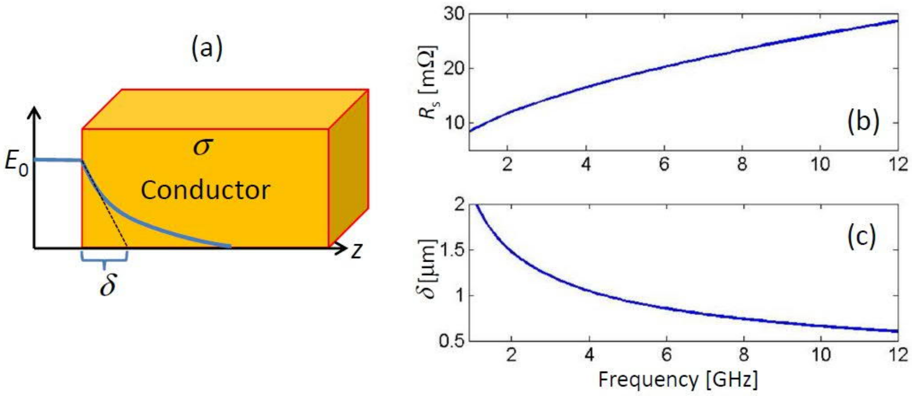


Fig. 15: (a) Sketch of the penetration of RF EM fields and surface currents inside metal; behaviour of (b) copper surface resistance and (c) skin depth as a function of frequency.

4.2 Niobium

Superconductivity was discovered in 1911. For a SC material below its critical temperature T_c, in the DC regime, resistance is zero. In the RF regime, however, the surface resistance is always larger than zero (even if orders of magnitude lower than NC materials), because not all the electrons are in the superconducting state. The residual ones are not completely shielded by the superconducting currents and thus experience a residual electric field, dissipating power. At frequencies below 10 GHz (and temperatures below $T_c/2$), the experimental data available for several materials are well described by the empirical relation [17–21]

$$R_s = \underbrace{A\frac{\omega_{\mathrm{RF}}^{\;2}}{T}e^{-\alpha\frac{T_c}{T}}}_{R_{\mathrm{BCS}}} + R_{\mathrm{res}} \tag{30}$$

The first term, resistance (R_{BCS}), is well explained by a theoretical model of the superconductor. In this term the coefficients A and α depend on the material. The second term is a residual term due to impurities in the material.

The SC state can be destroyed by an external magnetic field larger than a critical field H_c that depends on the material used. In practice, this effect fixes the maximum theoretical field that a SC cavity can sustain.

The most common material for the fabrication of SC cavities is niobium for several reasons [21–23]:

i) it has a relatively high transition temperature (T_c = 9.25 K);

ii) it has a relatively high critical magnetic field, H_c = 170 mT to 180 mT;

iii) it is chemically inert;

iv) it can be machined and deep drawn;

v) it is available either as bulk or sheet in any size, fabricated by forging and rolling;

vi) large grain sizes (often favoured) can be obtained by e-beam melting;

vii) it can also be used as a coating (e.g. by sputtering) on NC materials like Cu;

viii) it has good thermal stability and is of relatively low cost.

The R_{BCS} resistance for Nb is given by

$$R_{\mathrm{BCS}} = 2\times10^{-4}\frac{\left(f_{\mathrm{RF}}\left[\mathrm{MHz}\right]/1500\right)^2}{T}e^{-\frac{17.67_c}{T}} \qquad [\Omega] \tag{31}$$

while the residual resistance can vary between 5 nΩ and 20 nΩ.

The behaviour of niobium resistance as a function of temperature at 700 MHz is shown in Fig. 16.

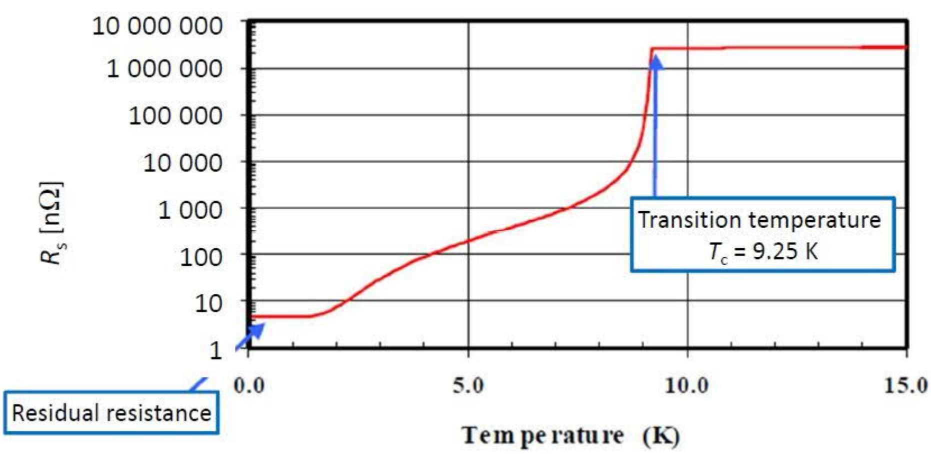


Fig. 16: Surface resistance of Nb at 700 MHz (courtesy [19])

5 Cavity parameters' scaling with frequency

The material properties, described in the previous section, allow understanding the frequency scaling laws of cavity parameters, which have been summarized in Table 3. For the sake of completeness, in the bottom rows of the table the scaling of the wakefield intensities have also been reported. These have an important impact on beam dynamics, and their description is beyond the scope of this paper. For our purposes, it is only important to remark that, the higher the wakefield intensity, the higher is the possible perturbation in single-bunch and multi-bunch beam dynamics.

Table 3: Scaling laws for cavity parameters with frequency

Parameter	NC	SC
Surface resistance (R_s)	$\propto f^{1/2}$	$\propto f^{2}$
Quality factor (Q)	$\propto f^{-1/2}$	$\propto f^{-2}$
Shunt impedance per unit length (r)	$\propto f^{1/2}$	$\propto f^{-1}$
r/Q	$\propto f$	
Longitudinal wakefield ($w_{//}$)	$\propto f^{2}$	
Transverse wakefield ($w_{\perp}$)	$\propto f^{3}$	

r/Q increases at high frequency. For NC structures, r also increases and this push to adopt higher frequencies. For SC structures, since the power losses increases with f^2, r scales with $1/f$ and this push to adopt, in principle, lower frequencies. On the other hand, at higher frequencies the beam–cavity interaction due to wakefields becomes more critical ($w_z \propto f^2$, $w_{\perp} \propto f^3$) and commercial power sources (above 6 GHz) are less commonly available. Cavity fabrication at very high frequency requires higher precision and critical alignment but, on the other hand, at very low frequencies one needs more material and larger machines for the fabrication of components.

These points clearly show that a compromise is required, and for conventional FEL and ERL this basically fixes the operational frequency for SW SC cavities between 500 MHz and 1500 MHz, for TW NC structures between 3 GHz and 6 GHz and for SW NC structures between 0.5 GHz and 3 GHz.

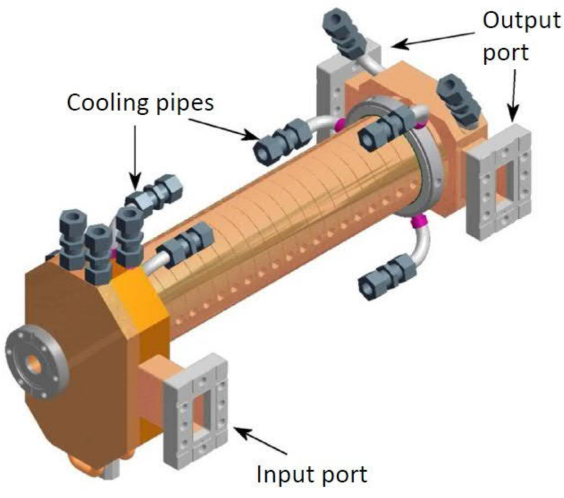


Fig. 17: Mechanical drawing of a small prototype of a C-band TW structure with a reduced number of accelerating cells.

6 Linac technology issues for normally conducting travelling wave structures

TW NC structures are usually a few metres long. They are made of hundreds of cells, and an input and an output coupler (the latter connected to an RF load). As an example, the mechanical drawing of a small prototype of a C-band TW structure with a few accelerating cells is given in Fig. 17. As has already been pointed out, these structures typically operate with short RF pulses (between 0.5 μs and 1.5 μs) in a single bunch or with a few bunches, with high peak power RF pulses (~50 MW). The accelerating field inside such structures is relatively high (ranging from 20 MV/m to 50 MV/m), and the mode of operation is, generally, the 2π/3 mode. The typical frequency bands of operation are S or C (3 GHz and 6 GHz, respectively). TW cells' geometries and, in particular, the irises' dimensions and profiles are optimized to reach a good compromise between a high shunt impedance (which favours small irises) and a short filling time (which favours large irises). As an example, Fig. 18 shows the cell parameters as a function of the iris dimensions for a C-band structure. Peak surface electric fields occur on the irises separating adjacent cells, as shown in Fig. 19(a), where we report the magnitude of the electric field in a C-band accelerating cell with 30 MV/m average accelerating field. Peak field values are typically a factor of 2–2.5 larger than the accelerating one. The magnetic field distribution, for the same average accelerating field, is given in Fig. 19(b). It occurs on the outer walls of the cells (as for TM_{01} mode). Cooling pipes are inserted or brazed around the cells to guarantee temperature stability of the structures, thus avoiding detuning under high power feeding, as shown in Fig. 24(c) below, where we report pictures of a fabricated cell.

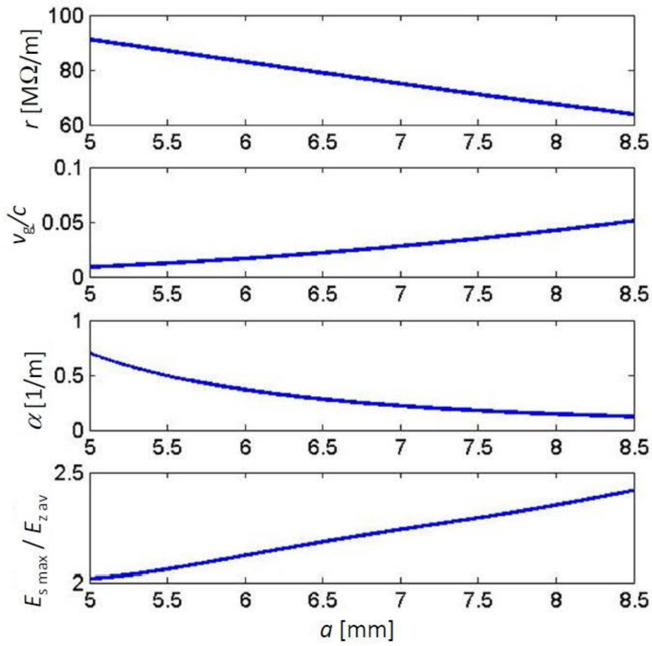


Fig. 18: Cell parameter as a function of the iris dimensions for a C-band structure

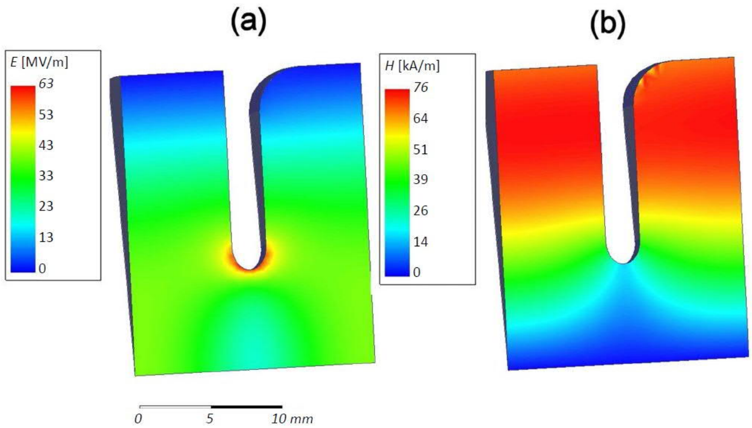


Fig. 19: (a) Electric field and (b) magnetic field intensity for a C-band accelerating cell with 30 MV/m average accelerating field on the axis.

The structures are fed by waveguides coupled by means of slots in a so-called ‘coupling cell’ that matches the TE_{10} mode of the waveguide with the travelling wave mode (TM_{01}-like), as shown in Fig. 20 [3, 24]. Either J-type couplers [13], or couplers with integrated splitters, allow a symmetric feed that compensates for the dipole kicks in the coupling cell, which are present if there is only a single slot, as shown in Fig. 21. In the coupling slot between the waveguides and the cavity we can have a high magnetic field and, as a consequence, high pulsed heating that can increase breakdown phenomena [3]. For this reason, rounded shapes are frequently used (as sketched in Fig. 22). Moreover, racetrack profiles allow for compensation of the quadrupole distortions of the field in the coupling cells introduced by the coupling holes for the waveguides. The typical magnitude and phase of the accelerating field in a short C-band TW structure are plotted, as an example, in Fig. 23. In the plot we can recognize the $2\pi/3$ phase advance from cell to cell and the periodic profile of the accelerating field (where the maximum field occurs in the centre of each cell).

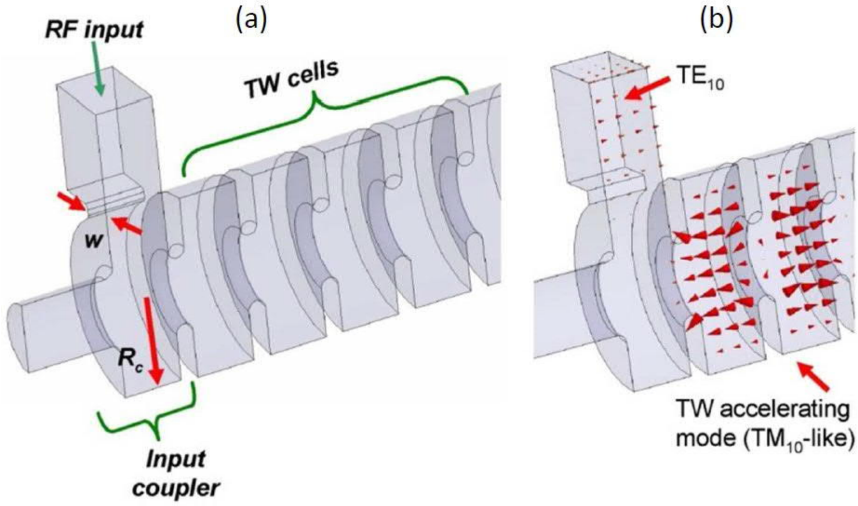


Fig. 20: (a) Coupling of the waveguide network to a TW structure; (b) in the input coupler the TE_{10} mode of the waveguide is matched to the travelling wave mode (TM_{01}-like).

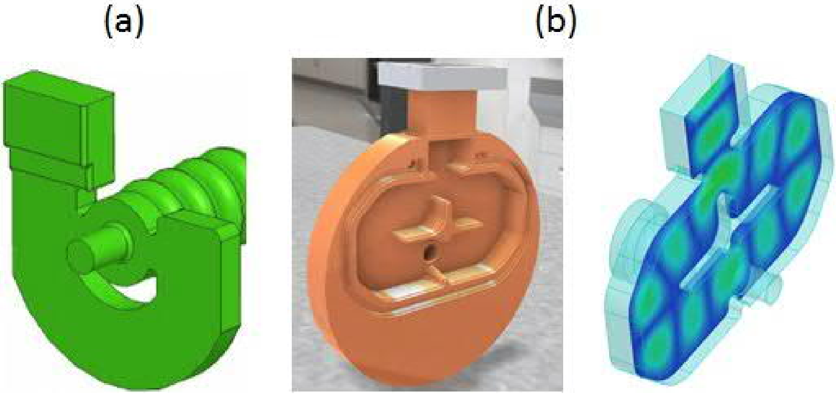


Fig. 21: Example of couplers with symmetric feeding to avoid dipole kicks: (a) J-type couplers (b) couplers with splitters.

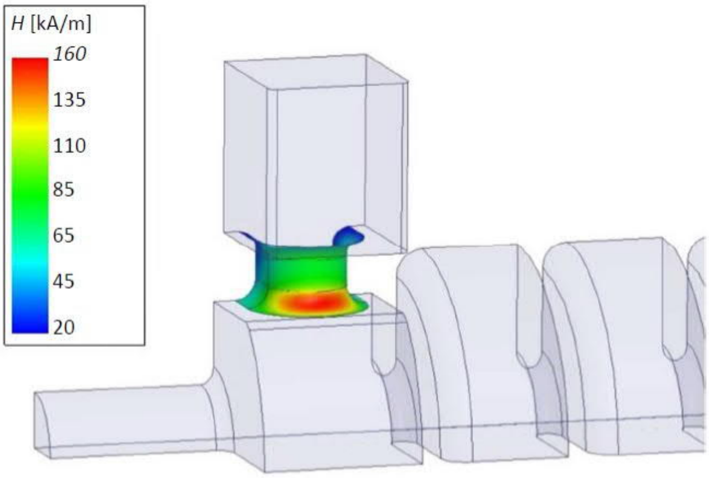


Fig. 22: Magnetic field in the coupler region

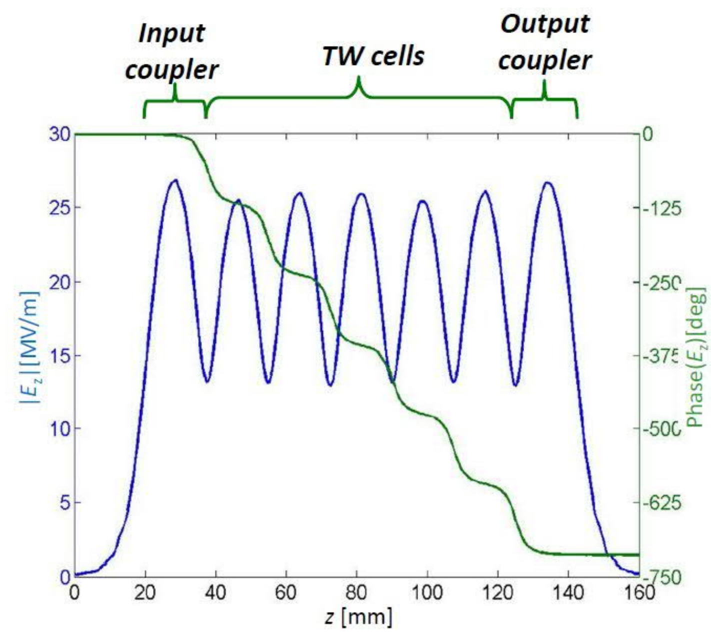


Fig. 23: Accelerating field in magnitude (blue) and phase (green) in a short C-band TW structure with 20 MV/m average accelerating field.

The whole fabrication process sequence is reported in Fig. 24. Cells and couplers are typically realized starting from OFHC forged or laminated copper (Fig. 24(a)) using milling machines and lathes (Fig. 24 (b)), with a precision that can be of the order of a few micrometres and a surface roughness that can reach values below 50 nm. The cells are then cleaned, piled up (Fig. 24(c) and (d)) and brazed together in vacuum or hydrogen furnaces (Fig. 24(e)), using different alloys at different temperatures (700°C to 1000°C).

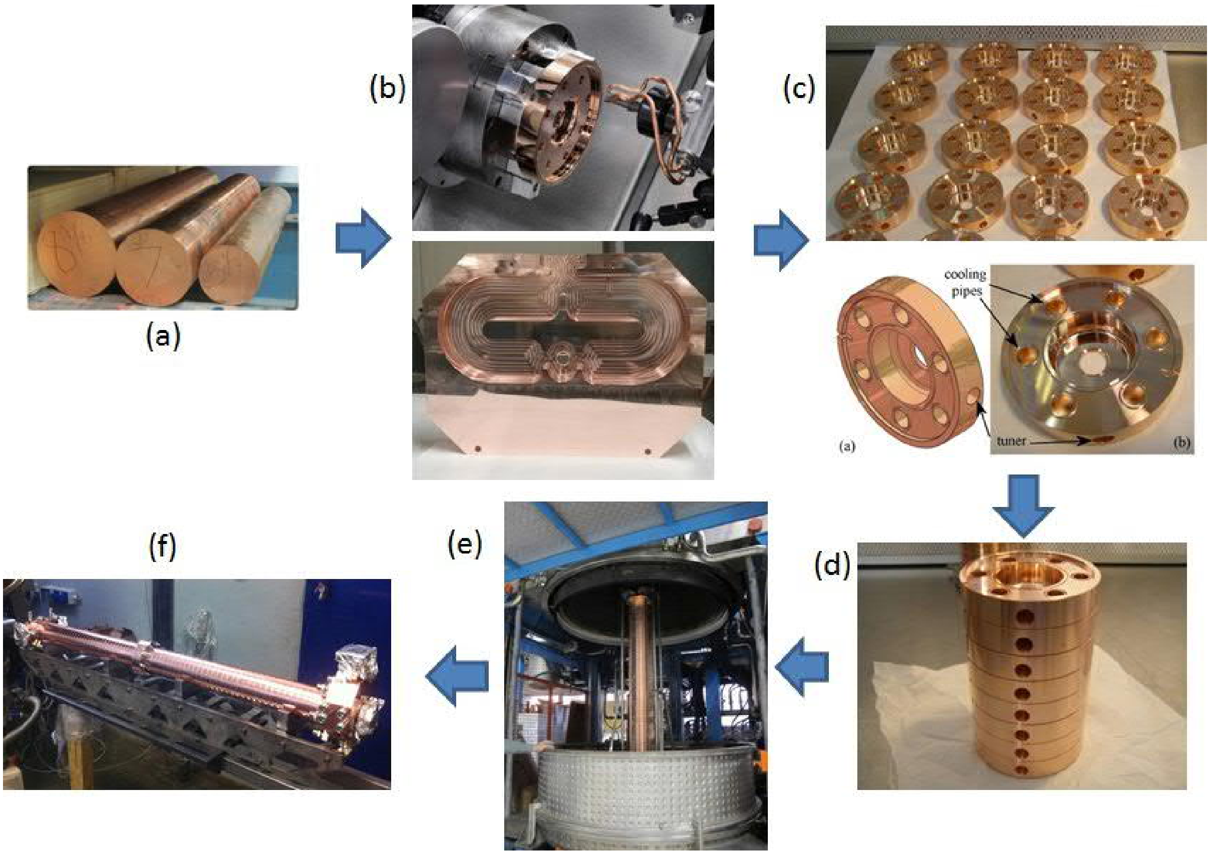


Fig. 24: Sequence of the fabrication process of a TW structure: (a) OFHC forged copper; (b) realization of cells by lathes; (c) single cells machined and ready to be stacked; (d) cells piled up before brazing; (e) the structure in a vacuum or hydrogen furnace; (f) the brazed structure.

To compensate for deformations and imperfections that can occur during the brazing process, tuning of TW structures is often necessary. The standard method is to measure the field inside the structure using a perturbation technique (the Steele method [25]), and adjusting the phase advance per cell to the correct value by deforming the outer volume of the cells with ‘deformation tuners’. A structure during a tuning procedure is shown in Fig. 25, while Fig. 26(a) represents the section of one cell, where deformation tuners have been designed to reduce the thickness of the walls in three points, and allow plastic deformations of the internal surface (shown in Fig. 26(b)). An example of the measured field before and after a tuning procedure is given in Fig. 27 for a 70-cell C-band structure [26]. The same plot also reports the relative phase advance per cell. Tuning algorithms must be used to calculate from the complex field measurements the deformation to be applied to the cells [27]. The final goal of a cell-to-cell phase error is typically ±2°, with a cumulative phase advance in the overall structure within ±5° [28].

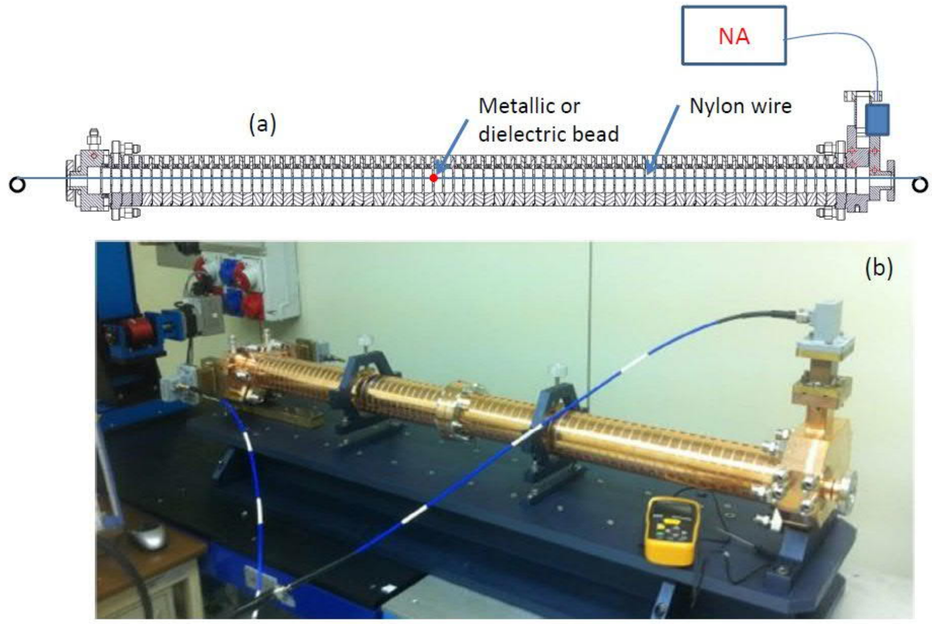


Fig. 25: TW structure during low-power RF measurements and tuning: (a) schematic layout; (b) picture of the cavity.

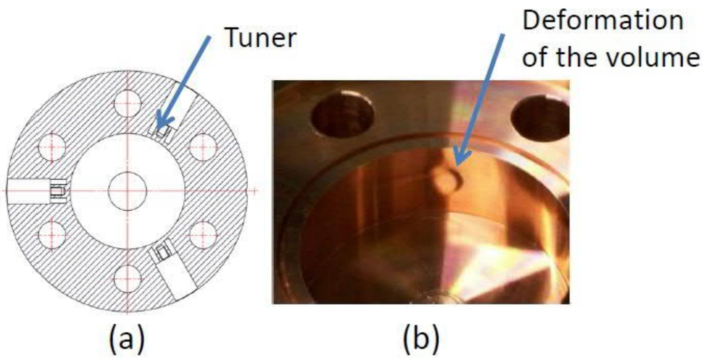


Fig. 26: (a) Section of one cell with deformation tuners reducing the thickness of the walls at three points; (b) deformation of the internal surface after tuning.

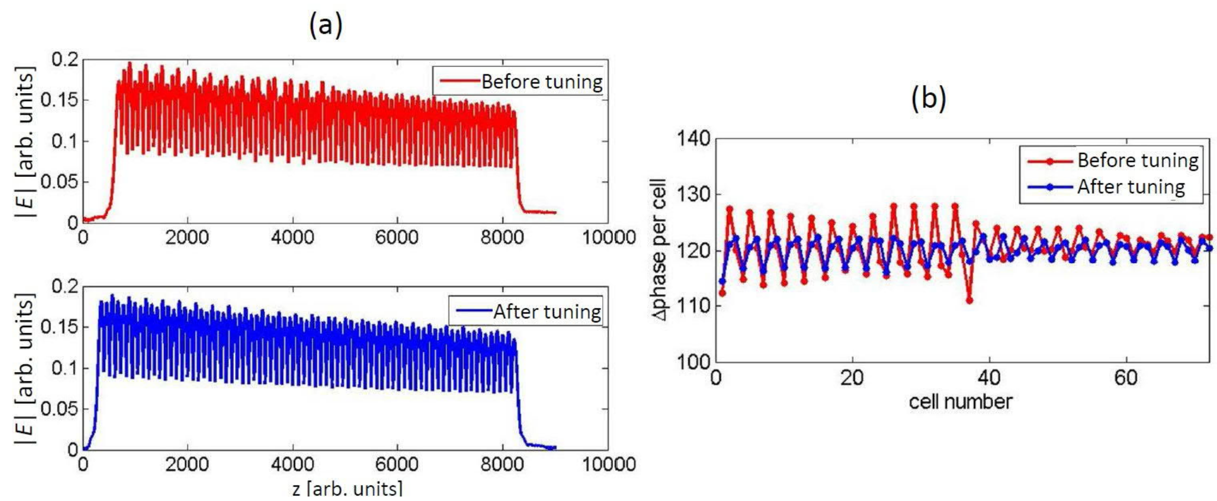


Fig. 27: (a) Measured accelerating field before and after the tuning procedure; (b) phase advance per cell before and after tuning.

TW structures require pulsed sources with a high peak power. For this purpose, klystron and RF compression systems (SLAC Energy Doubler, SLED) are usually adopted. As an example, we report in Fig. 28 a schematic of the RF network of the XFEL at Spring 8 [13]. The TW structures are quasi-constant gradient, working in the C-band with $3/4\pi$ phase advance. Their main parameters are reported in Table 2 [13]. They are fed by klystrons whose 2.5 μs RF pulse is compressed down to 0.5 μs with the use of SLED-type (SLED-SLAC Energy Doubler) cavities. In this way, an accelerating field of 38 MV/m can be reached.

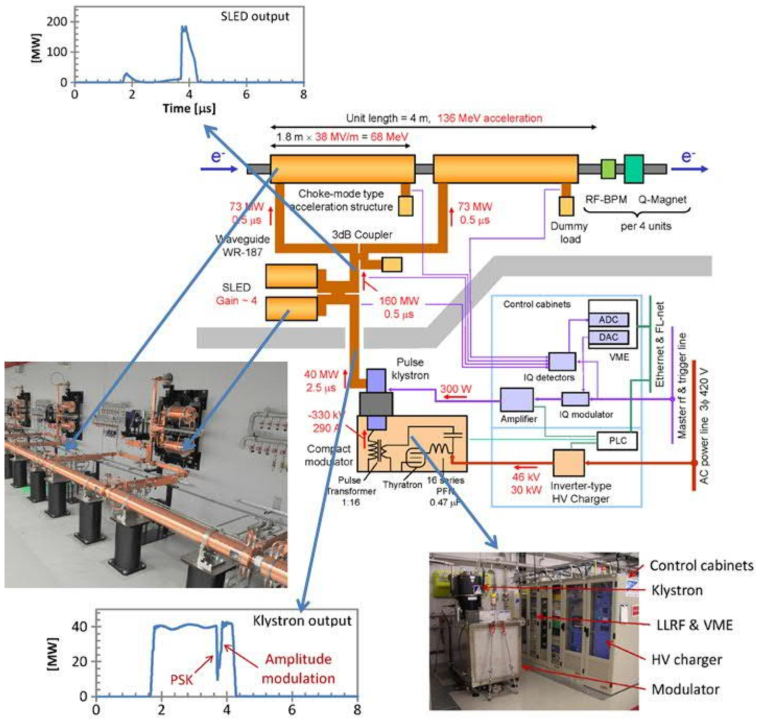


Fig. 28: Schematic of the RF network of the XFEL at Spring 8 (courtesy [13, 14])

7 Linac technology issues for superconducting standing wave structures

SC SW structures are typically multi-cell structures (up to 10 cells) working in the π mode. The cell irises are designed with an elliptical shape to minimize the ratio between the surface electric field and the accelerating one, E_{surf}/E_{acc}. The surface electric field, in fact, has to be minimized to avoid electron field emission processes that can limit cavity performance at a high gradient. The elliptical shape of the cell is chosen to minimize the ratio between the surface magnetic field and the accelerating one, B_{surf}/E_{acc}. As already pointed out, the ultimate limit for the maximum achievable gradient in SC structures is due to the critical magnetic field that for Nb is ~170 mT to 180 mT. For the TESLA cavities [6, 29, 30], since this ratio is $B_{surf}/E_{acc} \cong 4.2\,\mathrm{mT/MV/m}$ the maximum theoretical achievable accelerating field is about 50 MV/m.

The elliptical shape is also useful for suppressing multipacting phenomena that can limit cavity performance at a high gradient, while large irises also increase the machinability and cleanability of the cavities. The sketch of the cell profile of the TESLA cavities is shown in Fig. 29, while the r/Q and surface field are plotted as a function of the irises radius in Fig. 30 [29, 30].

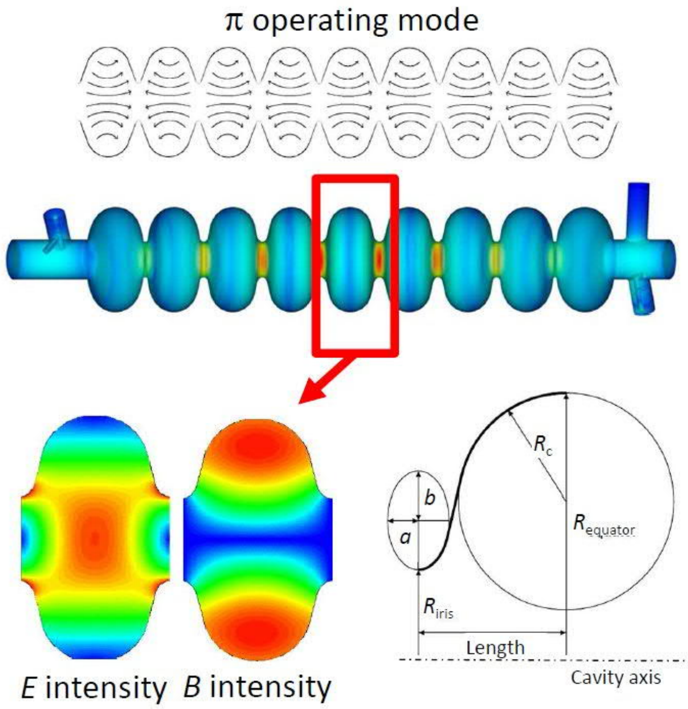


Fig. 29: Sketch of the cell profile of the TESLA cavities (courtesy [6, 29, 30])

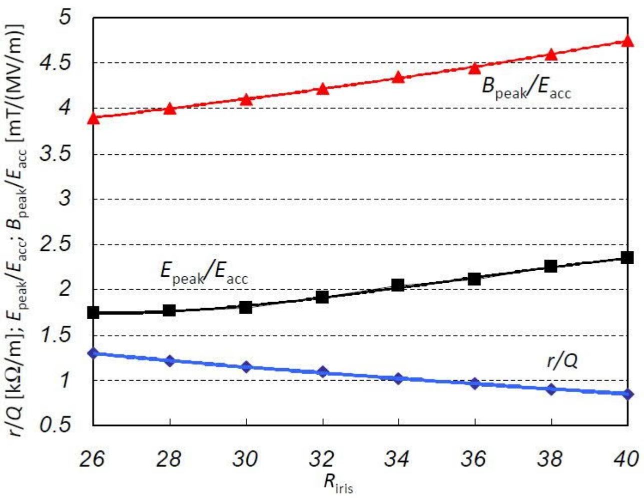


Fig. 30: r/Q, E_{surf}/E_{acc} and B_{surf}/E_{acc} as a function of the irises' radius (TESLA cavities, courtesy [29, 30])

SW SC cavities are typically powered by coaxial-type couplers [3, 31]. Magnetic coupling with waveguides or loops is also possible, but it can create hot spots in the cavities with additional complications in the design. The inner conductor of the coaxial is coupled to the electric field at the end of the structure. Coaxial couplers allow variation of the coupling strength by changing the penetration of the couplers into the cavity itself. For instance, this is necessary for accelerators that operate at different beam currents. The couplers also have vacuum barriers (windows) to prevent contamination of the SC structure. Obviously these barriers are also necessary in normally conducting accelerators, but the demands on the quality of the vacuum and reliability of the windows are less stringent than for SC cavities. The failure of a window in a superconducting accelerator could imply very expensive and time-consuming repairs. These windows are generally made from Al_2O_3. Ceramic material has also a high SEY that stimulates multipacting activity, and to reduce this phenomena Ti coatings on the surface of the ceramic are frequently used. Finally, the couplers must constitute a thermal barrier since they are at the boundary between room temperature and the cryogenic temperature environment. The final drawing of a SC coupler is therefore very complex, as shown in Fig. 31, where the TESLA coupler is presented [6].

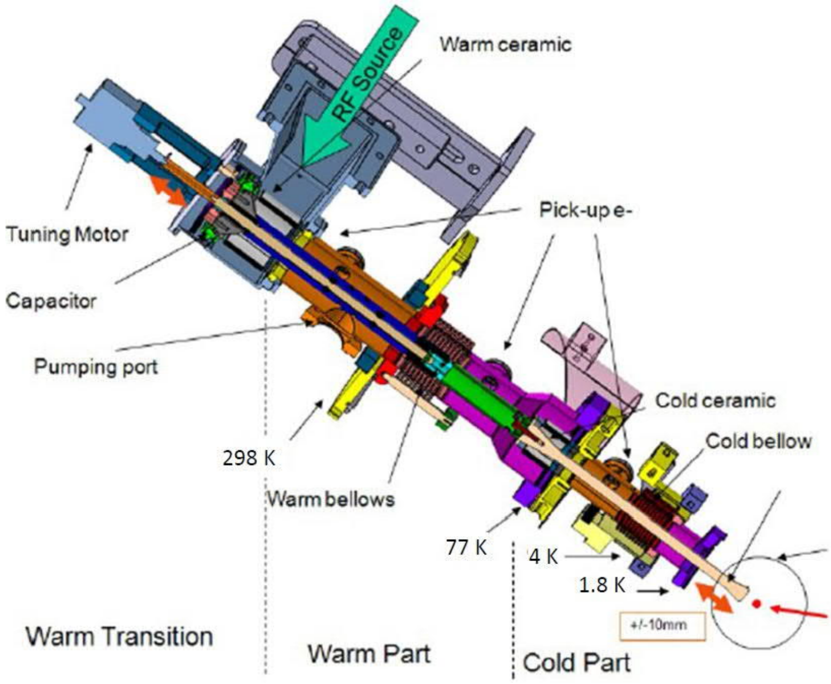


Fig. 31: Mechanical drawing of the TESLA coupler

As already pointed out, SC cavities are suitable for the acceleration of trains of bunches. As a bunch traverses a cavity, however, it deposits EM energy on Higher Order Modes (HOM), described in terms of long-range wakefields. Subsequent bunches (or the same bunch in several turns, as in ERLs) may be affected by these fields, causing beam instabilities and additional heating of accelerator components. As an example, the sketch of the field excited by one bunch passage into a TESLA cavity is reported in Fig. 32. To absorb the excited EM field several approaches can be used, such as loop couplers, waveguide dampers or beam pipe absorbers [31]. In all the above mentioned options, only the fields excited by the beam are absorbed, while the accelerating mode is rejected by EM filters and remains unperturbed. For instance, for a waveguide absorber the cutoff frequency of the waveguide is higher than the working frequency of the operating mode, thus it cannot propagate into the waveguide itself. When loops are used, notch filters are inserted to decouple the working mode from the coaxial. Pictures and drawings of different kind of such devices are shown in Fig. 33.

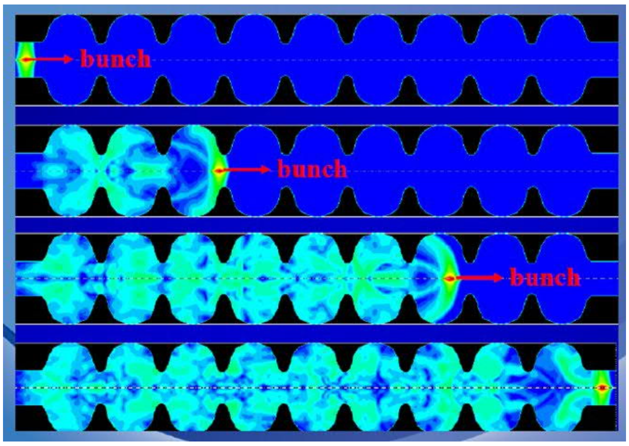


Fig. 32: Sketch of the field excited by the bunch passage into the cells of a TESLA cavity (courtesy [21])

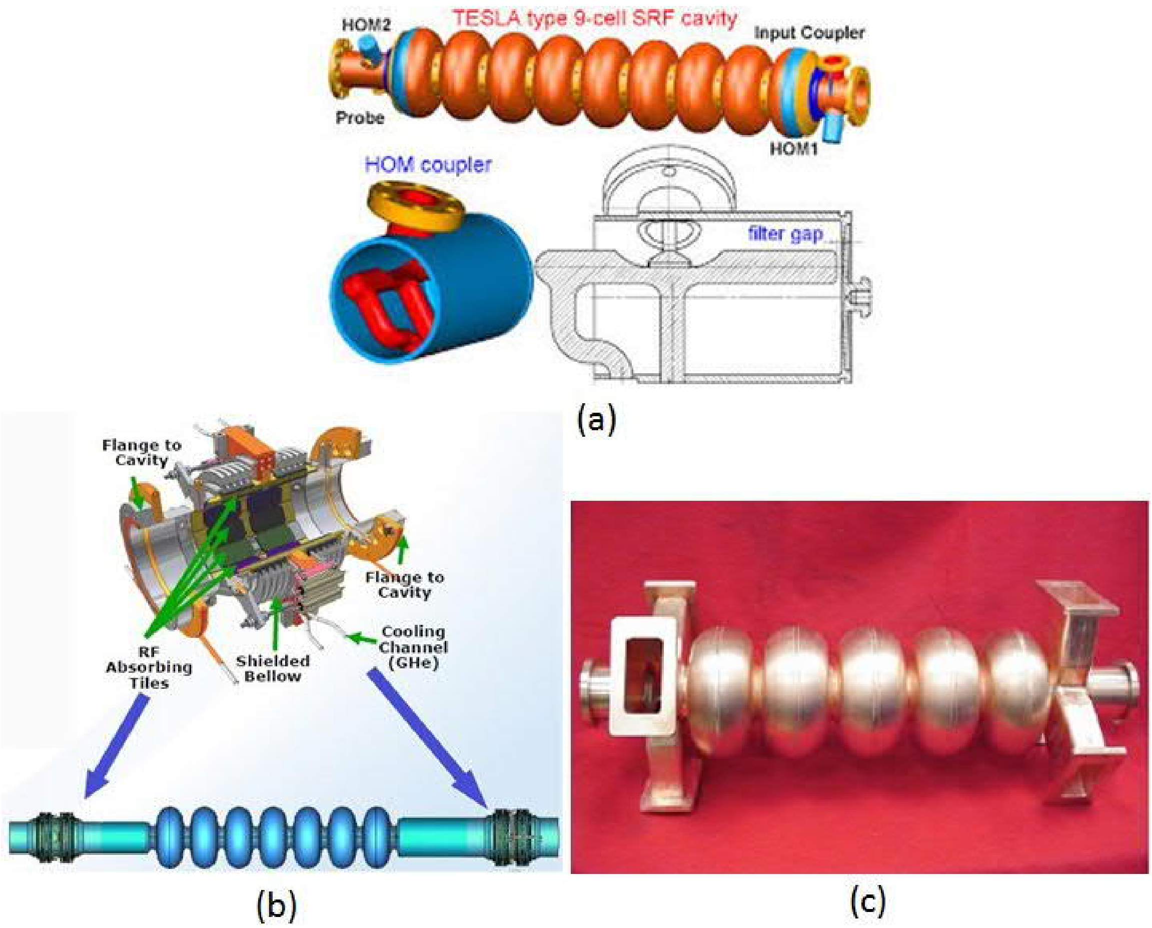


Fig. 33: Pictures and drawings of different types of HOM damping systems: (a) loop couplers, (b) waveguide dampers; (c) beam pipe absorbers (courtesy [21]).

Niobium is available as a bulk and sheet material in any size, fabricated by forging and rolling. High purity Niobium is made by electron beam melting under a good vacuum. The most common fabrication techniques for the cavities are deep drawing or spinning half-cells [21–23]. These processes are schematically represented in Fig. 34.

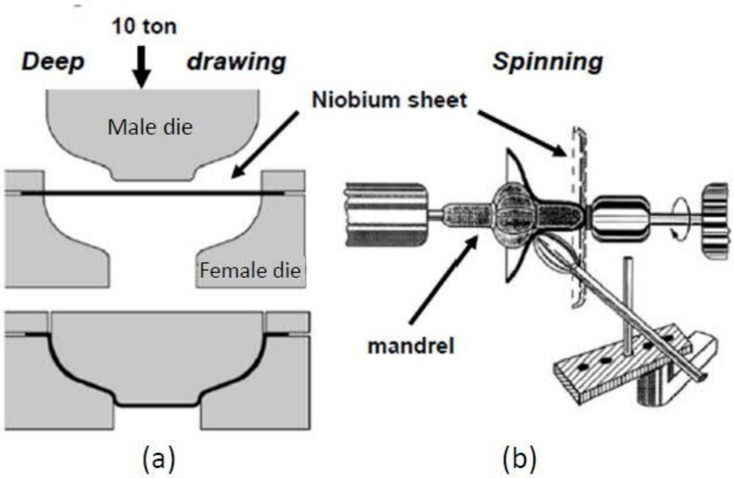


Fig. 34: Most common fabrication techniques for SC cavities: (a) deep drawing; and (b) spinning half-cells

Typically, the cells of the cavity are electron-beam welded and they undergo a long, and very delicate, process of polishing and cleaning [22, 23, 32, 33]. The cavity treatment requires several steps in between, such as buffered chemical polishing (BCP), electro-polishing and etching, which allows the removal of surface damaged layers of the order of ~100 μm. The cavities are rinsed with ultraclean water at high pressure (~100 bar) for several hours to remove the residual acid used for treatment. They also undergo a thermal treatment (with temperatures higher than 1000°C) to diffuse H_2 out of the material, increasing the purity of the niobium. During and after these treatments the cavities have to be maintained in a very clean environment, to avoid contamination that can limit their performances during high-power operation. Before the final cleaning process, the cavities are also characterized and tuned with low power RF. A picture of some cavities during the assembly process in a clean room is shown in Fig. 35. In general, the process of fabrication is not unique, and companies, industries or laboratories adopt their own developed techniques.

Fig. 35: Cavities during the assembly process in a clean room

After their construction and assembly, the cavities under vacuum, or in a controlled N_2 atmosphere, are inserted and assembled into the cryomodules. In the cryomodules the cavity is immersed in a liquid helium bath, which is pumped to remove helium vapour boil-off, as well as to reduce the bath temperature.

The cold portions of the cryomodule need to be extremely well insulated, which is best accomplished by a vacuum vessel surrounding the helium vessel and all ancillary cold components.

A schematic of a cryomodule is given in Fig. 36, while a picture of an XFEL cryomodule is shown in Fig. 37 with its mechanical cross-section. A single cryomodule can incorporate several cavities, as in the XFEL in Fig. 38, where eight cavities are integrated with bellows beam position monitors (BPMs) and quadrupoles.

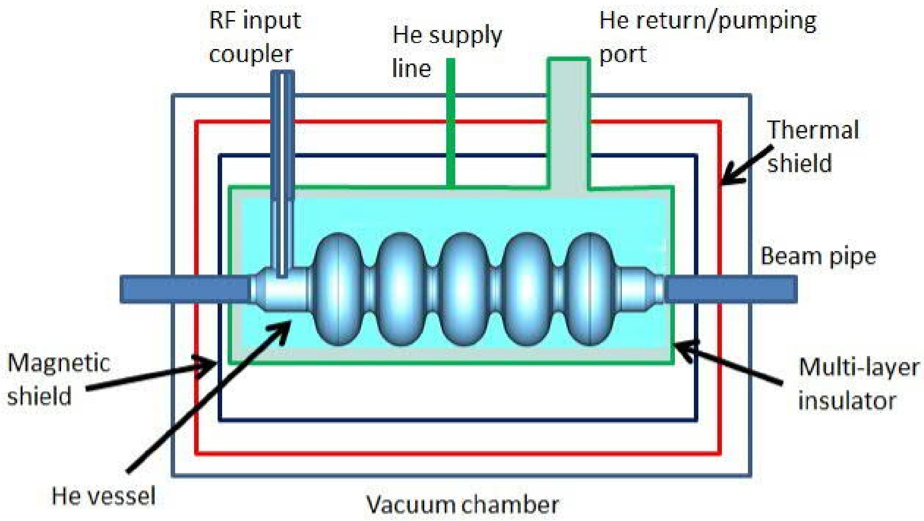


Fig. 36: Schematic of a cryomodule

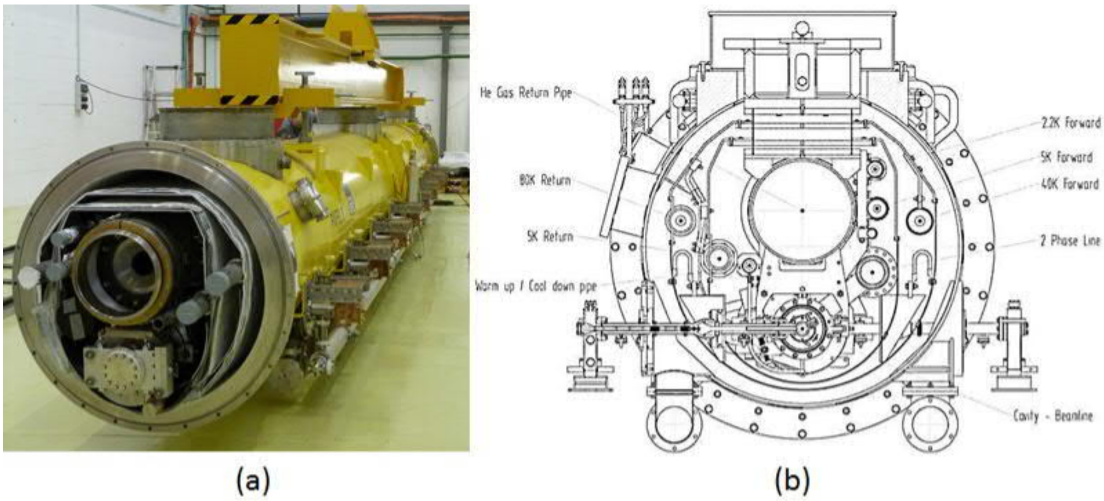


Fig. 37: (a) Picture of the XFEL cryomodule; (b) mechanical cross-section

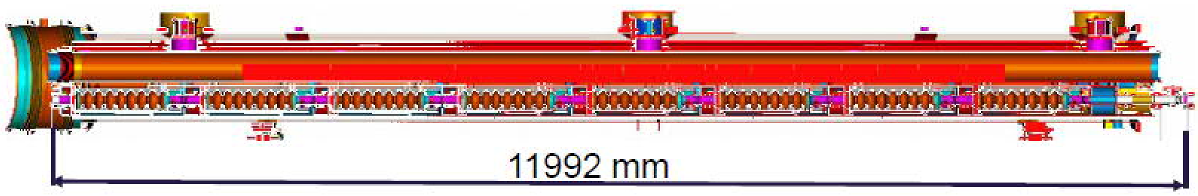


Fig. 38: XFEL cryomodule integrating eight cavities, bellows BPMs and quadrupoles

The requirements for the stability of the accelerating field in a superconducting structure are comparable to those of a normal conducting cavity. The nature and magnitude of the perturbations to be controlled, however, are rather different. Superconducting cavities have a very narrow bandwidth and are, therefore, highly sensitive to mechanical perturbations. Therefore, significant phase and amplitude errors may be induced by frequency variations. Perturbations can be excited by mechanical vibrations (microphonics), changes in helium pressure and level, or Lorentz forces. Slow changes in frequency, on a timescale of minutes or longer, can be corrected by a frequency tuner, while faster changes must be counteracted by an amplitude and phase modulation of the incident RF power. A schematic drawing of the RF system feeding a SC cavity is given in Fig. 39. The signals are measured from the pickup and are the inputs of the low level RF (LLRF) system, which allows the input to be kept linked to the resonant frequency of the cavity. Circulators are also needed to prevent damage to the klystron due to the reflected power from the cavities, as pointed out above.

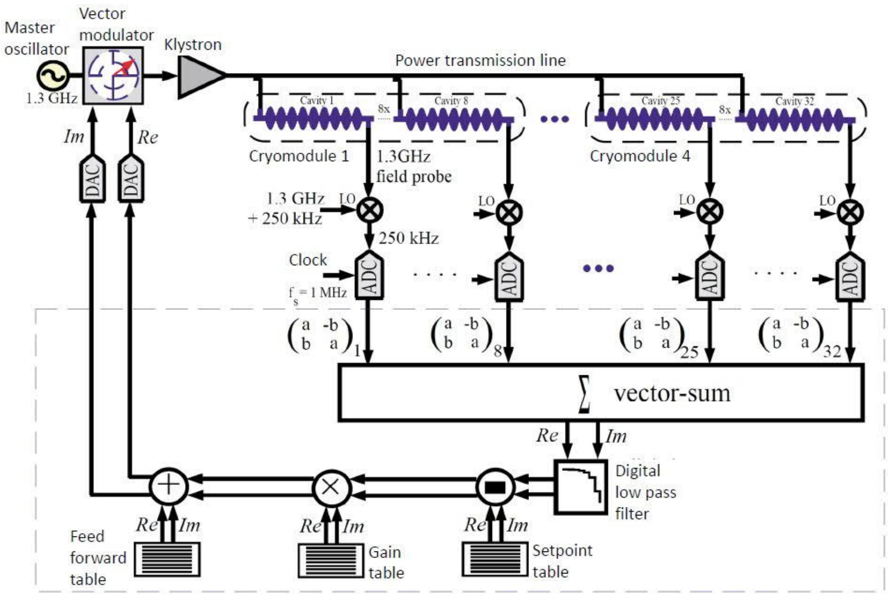


Fig. 39: Schematic drawing of the RF system feeding a SC cavity (courtesy [6])

8 Performances of different type of structures at a high accelerating field

If properly cleaned and fabricated, TW NC cavities can quite easily reach a relatively high gradient (>40 MV/m) without particular limitations. After less than one hundred million pulses (a few hundred hours at 100 Hz repetition rate) of conditioning, in which the power and pulse length are progressively increased, they can usually reach their nominal performance. The main limitation comes from breakdown phenomena, whose physics interpretation and modelling is still under study and is not yet completely understood. The goal of the conditioning process is to expose the internal surface of the component to ramped RF power in order to clean it. Several effects contribute to surface cleaning: induced out-gassing due to RF heating, arcing and multipacting (which enhance local heating) and desorption. To avoid damage to the structure, RF conditioning must be carried out gradually and in a controlled way. Generally, during the process different parameters are monitored, such as vacuum level, and forward, reflected and transmitted RF power. These parameters are used to verify the conditioning progress and to generate interlocks, in order to protect the machine components in case of breakdown.

At full performance, S-band cavities can operate without breakdowns at gradients of up to 25 MV/m to 30 MV/m, while C-band cavities can even reach higher gradients. So far, high gradient tests on C-band structures have been successfully carried out at up to 50 MV/m. In the X-band (12 GHz) higher gradients can be reached (>100 MV/m). Nevertheless, at present, FEL linacs do not operate at these frequencies due to the higher cost of power sources and more critical issues related to fabrication, tolerances, alignments, pumping systems, etc. In principle, a very high gradient (up to the level of the X-band structures) can be reached at lower frequencies (S- or C-band) as well. In these cases, however, the main limitation is the lower shunt impedance, which requires a higher power per unit length to sustain such gradients. A typical behaviour of the conditioning process for a TW structure (Extreme Light Infrastructure-Nuclear Physics, ELI-NP, C-band) is shown in Fig. 40, where the pulse length, the structure input power and the repetition rate are reported as a function of time (hours). As can be noted from Fig. 40, after about 150 hours of conditioning the structure reached its final level of performance [34].

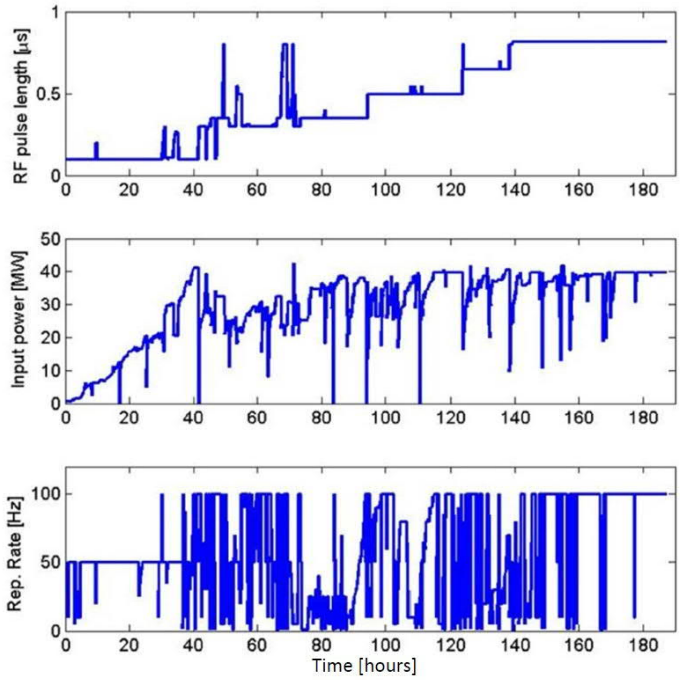


Fig. 40: An example of typical behaviour during the conditioning process for a TW structure (ELI-NP C-band)

SC cavities also need to be conditioned and, their performance is usually analysed by plotting the behaviour of the quality factor as a function of the accelerating field [35]. A typical plot is reported in Fig. 41, where the dots represent the measured points with respect to the ideal case (solid line). There are several phenomena responsible for additional losses under high power, which causes the quality factor to be lower than expected.

First, as already pointed out, the measured surface resistivity is typically larger than that predicted by Bardeen–Cooper–Schrieffer (BCS) theory. This is due to magnetic flux trapped in the cool-down process, dielectric surface contaminations (chemical residues, dust…), defects and inclusions, surface imperfections and hydrogen precipitates. This increase in losses can be limited by very clean and careful cavity fabrication, e.g. using ultra-pure Nb, cleaning processes and thermal cavity treatments.

At an intermediate level of field, multipacting phenomena can occur [15]. Multipacting is a resonant process that occurs when a large number of electrons build up under the influence of an RF field. It can occur either in the regions of input couplers or in cavity cells. Basically, it starts from a process of electron field emission. The emitted electrons start to hit the surfaces, emitting more electrons, and so on. Two main conditions are needed to activate the electron build-up: electron synchronization with the RF field and electron multiplication via secondary emission (SEY). Multipacting was an early limitation of SC cavities performance and it was overcome by adopting spherical or elliptical cell shapes. These allow modification of the electrons' trajectories, reducing surface collisions and the synchronicity with the RF field. RF conditioning can also reduce multipacting. Figure 42(a) shows typical electron trajectories when RF fields are applied, while Fig. 42(b) reports the SEY as a function of the impact energies of electrons for a cavity with several different treatments.

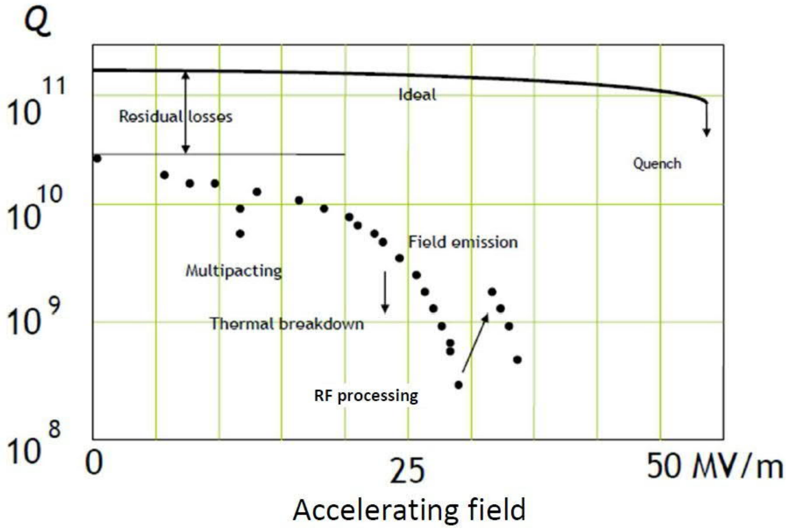


Fig. 41: Typical behaviour of the quality factor of a SC cavity as a function of the accelerating field (courtesy [35])

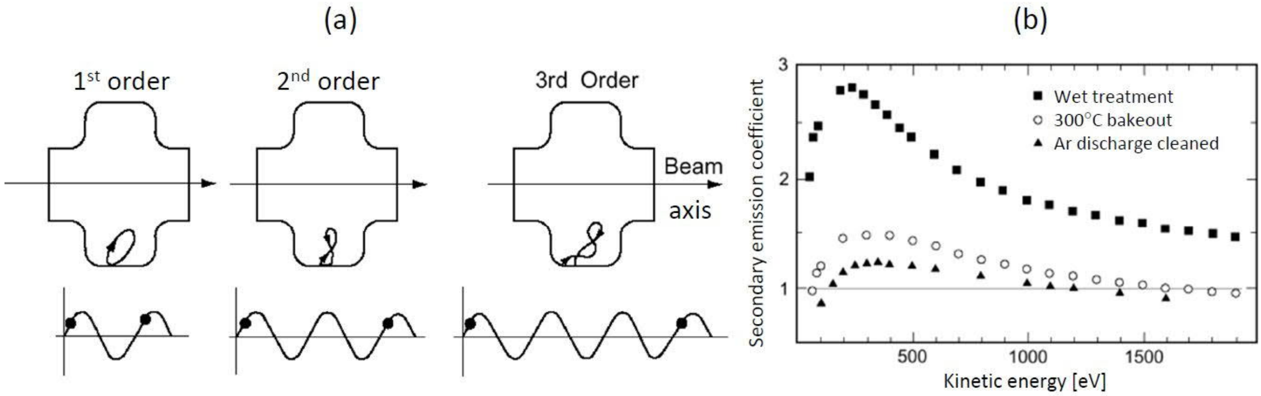


Fig. 42: (a) Typical electron trajectories under RF field that causes hitting of the surface by emitted electrons; (b) SEY as a function of the impact energies of electrons for a cavity with several different treatments (courtesy [35]).

At high fields, thermal breakdown can also occur when the heat generated in one hotspot (e.g. due to impurities or defects) is larger than can be transferred to the helium bath. This causes an increase of the temperature above the critical temperature (T_c) and, consequently, a 'quench' of the superconducting state. The mechanism is schematically illustrated in Fig. 43.

The last phenomenon we want to mention is the exponential increase of losses due to the acceleration of field-emitted (FE) electrons, which is also associated with the production of X-rays and dark current. This mechanism is illustrated in Fig. 44. The main cause of FE is particulate contamination. FE can be prevented by proper surface preparation and contamination control. It is possible to reduce it also using High-power Pulsed Processing (HPP) and/or helium processing.

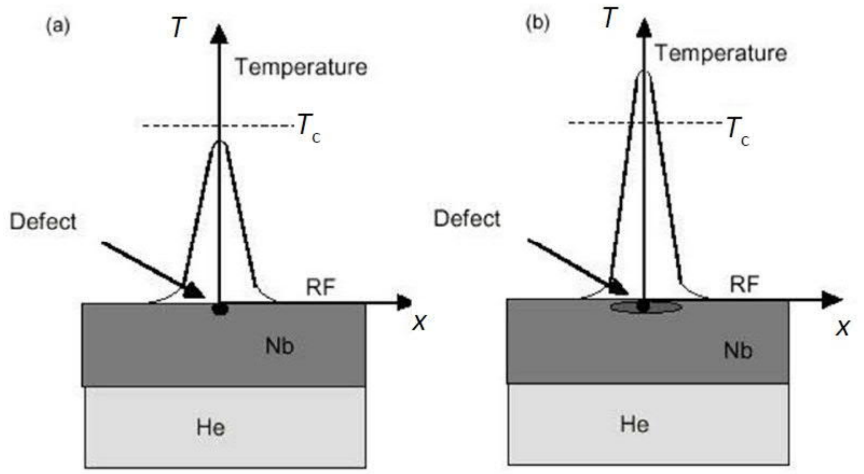


Fig. 43: Thermal breakdown schematic representation (courtesy [35])

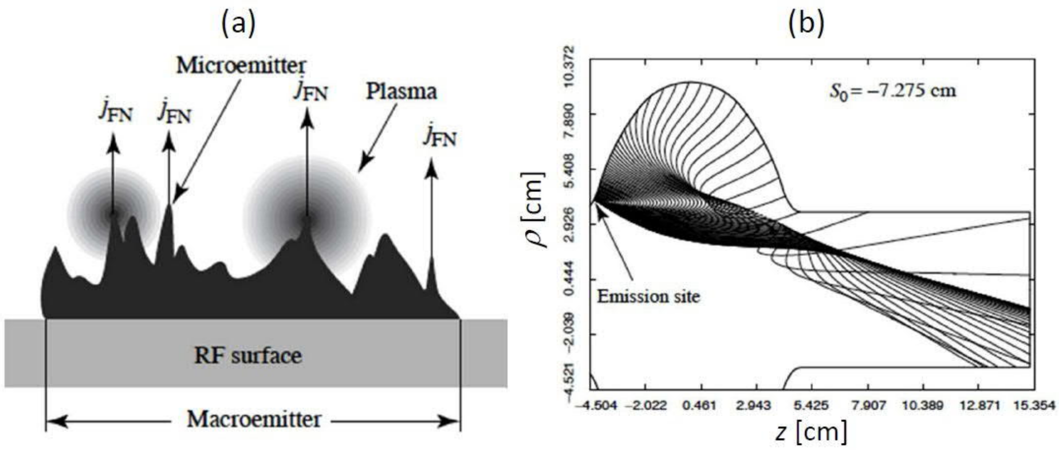


Fig. 44: FE process: (a) the emission of electrons is enhanced by the presence of surface defects of spikes; (b) emitted electrons hit the surface causing X-rays emission and heating (courtesy [21]).

9 Electron sources

There are different types of electron sources. A complete treatment can be found in [36]. Here we want to mention only a couple that may be used for different applications in FEL and ERL.

9.1 Radio frequency photo-guns

Radio frequency photo-guns are used as electron sources and for the first acceleration stages in FEL. They are multi-cell structures, typically made from one to three cells. The geometry of a 1.6 cell RF gun with its electric field lines is given in Fig. 45. The device is powered by a waveguide and a coupling hole on the full cell. These structures typically operate in the π mode, and the electrons are emitted by a cathode, whose surface is hit by a laser (working usually in the UV, e.g. at 240 nm to 260 nm). Electrons are extracted by means of the photo-electric effect, and then accelerated by an electric field that is designed to have a longitudinal component on the gun's axis. They enter the full cell, where they are further accelerated. RF guns can operate in the L- or S-band (i.e. from ~1 GHz to ~3 GHz) at repetition rates upto 100 Hz and above. The cathode peak field is typically of the order of 60 MV/m to 120 MV/m. The RF pulse length can be a few microseconds (S-band) up to hundreds of milliseconds (L-band). The number of emitted electrons is proportional to the laser power and quantum efficiency (QE) of the cathode material [36]. QE is given by the ratio between the number of emitted electrons and the number of incident photons (QE of copper cathodes is in the range 10^{-5} to 10^{-4}). The cathodes can be of different type: copper (as is the bulk of the gun) or other materials with a higher QE (for instance GaAs). The mechanical design of the gun has to allow for the possibility of substituting the cathode itself in case of deterioration or damage.

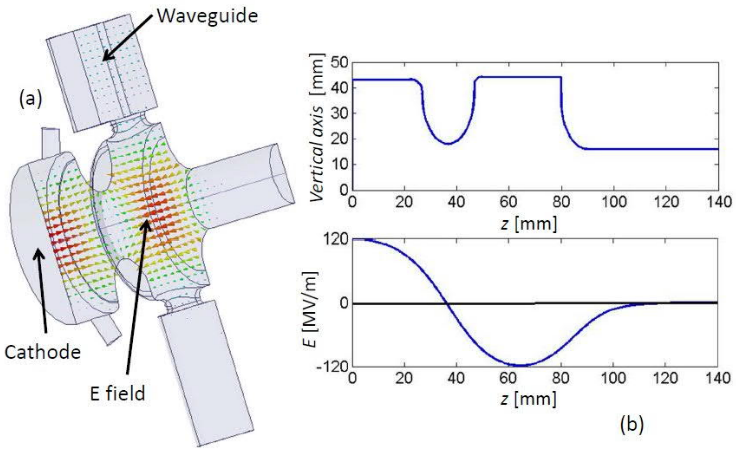


Fig. 45: (a) Geometry of a 1.6 cell RF gun with electric field lines; (b) profile along the longitudinal axis

The mechanical drawing of the ELI-NP RF gun [37] is given in Fig. 46(a) with a picture of the main components before assembly and brazing (Fig. 46(b)). Figure 47 shows the RF gun installed with the solenoid immediately after the accelerating cells for emittance compensation and control. The conceptual scheme of the laser system and photo gun is given in Fig. 48. Figure 49(b) shows a picture of the LCLS photo-gun together with the power splitter, which is needed to symmetrize the power feed to avoid dipole kicks due to the coupler holes. The main LCLS gun parameters are summarized in Table 4 [38–40].

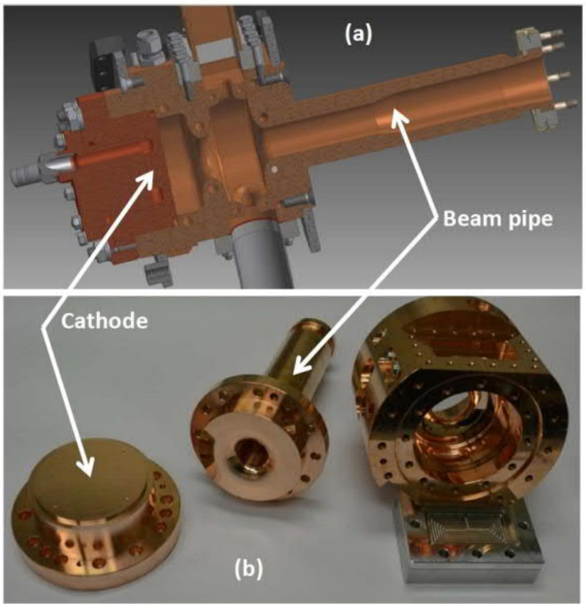


Fig. 46: (a) Mechanical drawing of an RF gun; (b) pictures of the main components before assembly

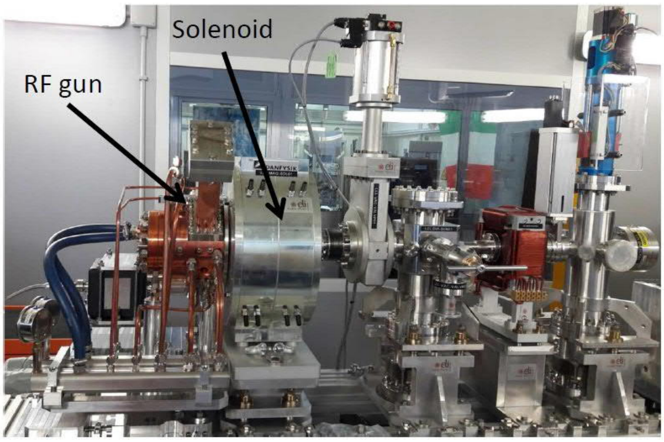


Fig. 47: RF gun installed with the solenoid immediately after the accelerating cells for emittance compensation

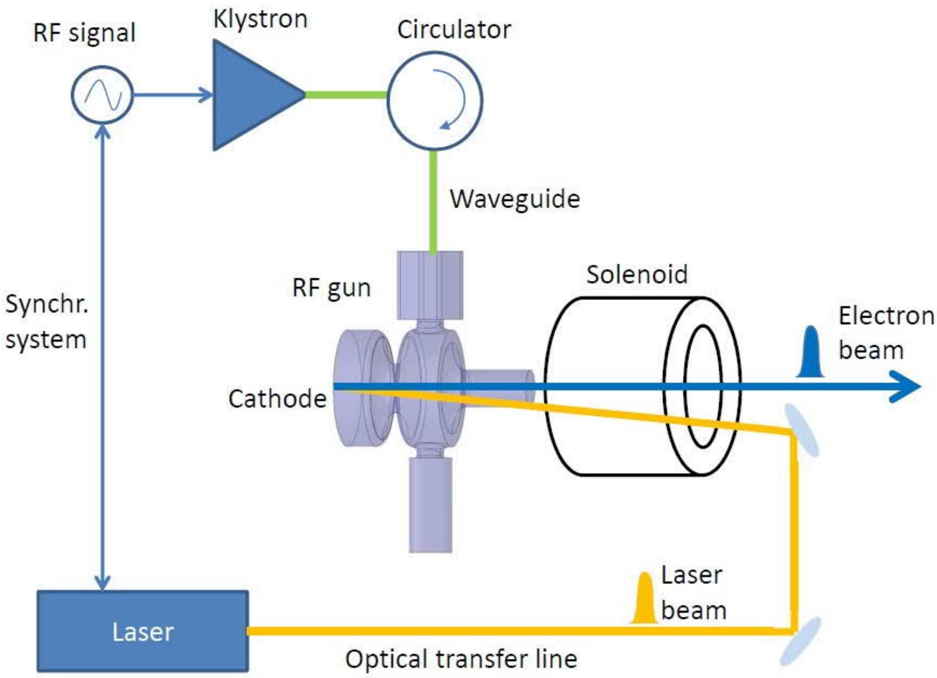


Fig. 48: Conceptual scheme of the laser system and photo-gun

Figure 49(a) shows the PITS L-band gun of the XFEL [5, 41, 42]. Two solenoids are foreseen to completely cancel the magnetic field on the cathode, which can increase the beam emittance. The main parameters are given in Table 4.

Table 4: Main LCLS and PITZ gun parameters

Parameter	LCLS gun	PITZ gun
Frequency [GHz]	2.856	1.3
Number of cells	1.6	1.5
Cathode peak field [MV/m]	120	60
Photocathode type	Cu	Cs_2Te
Repetition rate [Hz]	120	10
Number of bunches per RF pulse	1	<2700
RF pulse length	~2 μs	<700 μs

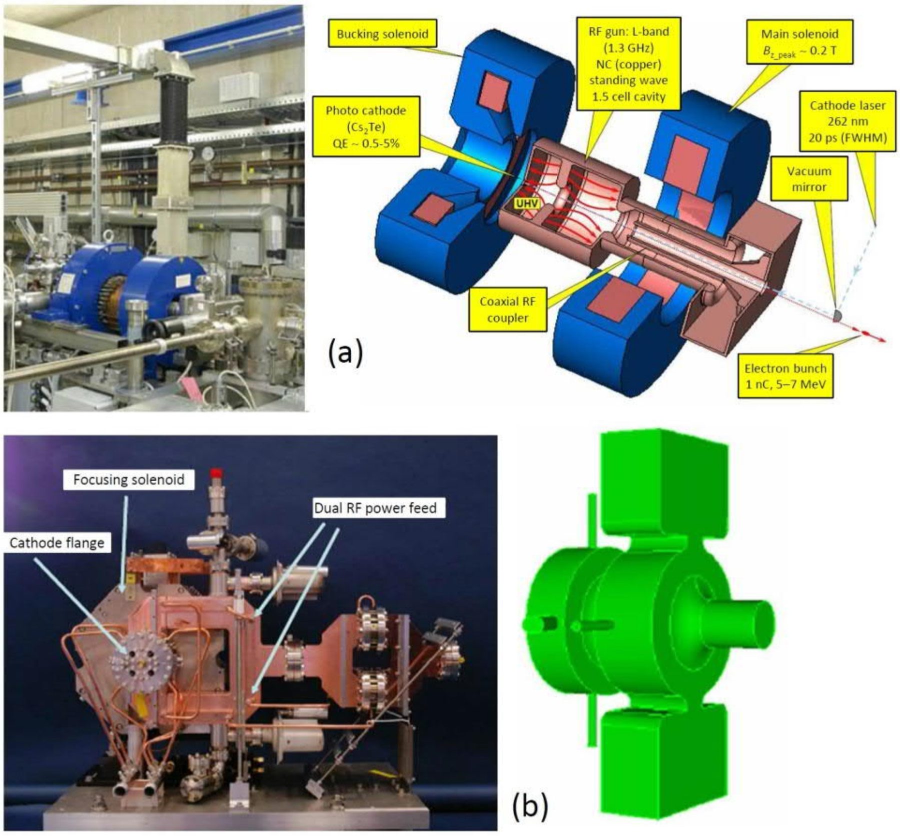


Fig. 49: (a) PITS L-band gun of the XFEL; (b) picture of the LCLS photo-gun with the big splitter to symmetrize the feeding avoiding dipole kick.

9.2 Direct Current photo-guns

Direct current (DC) photo-guns can be used as electron sources for high average current accelerators (CW operation, high Duty Cycle linac, typically ERL). The cathode must have a high QE (as example for GaAs QE is a few percent) and is illuminated by a continuous train of laser pulses. Average currents up to 100 mA can be achieved with this kind of source. A mechanical drawing of the Cornell DC gun [43] is given in Fig. 50, with the main parameters reported in Table 5. DC photo-guns are characterized by an electron energy that is lower than that of RF photo-guns. This is due to the lower accelerating gradient that is possible to achieve in the DC regime; post-acceleration and beam manipulation are then required to reach the desired beam energy and performance. As an example, Fig. 50(b) shows the line after the Cornell DC gun, made from two emittance compensation solenoids and a 1.3 GHz normal conducting bunching cavity. These elements are used to compensate the initial emittance blowup near the cathode, and to compress the bunch longitudinally before further acceleration. The bunches are then accelerated using five superconducting niobium cavities. This multi-cell cavity allows the partial freezing of the emittance, to increase the energy and to perform further emittance compensation and longitudinal compression via time-dependent transverse and longitudinal focusing.

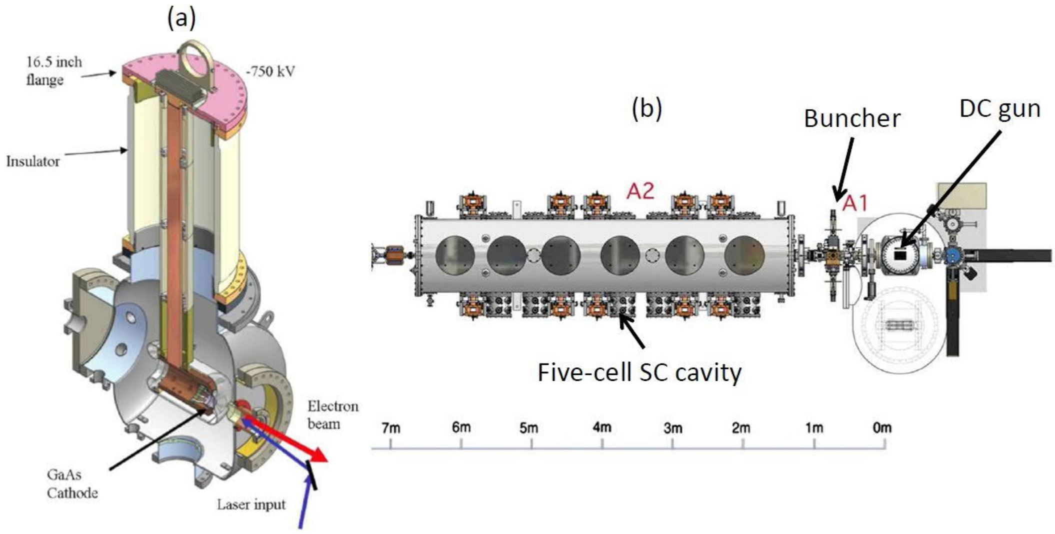


Fig. 50: (a) Mechanical drawing of the Cornell DC gun; (b) layout of the complete Cornell injector

Table 5: Main Cornell injector design parameters

Parameter	Value
Beam energy [MeV]	5 to 15
Normalized emittance [μm]	≤0.3
Bunch length [ps]	≤3 ps
Photocathode type	GaAs
DC accelerating voltage [kV]	500 to 600
Average current [mA]	100
Bunch charge [pC]	77
Bunch frequency [GHz]	1.3

10 Power sources and power distribution systems

RF power sources would require a dedicated treatment. A comprehensive description can be found, for example, in [44–46]. In this paper, for sake of completeness, we report some useful information.

10.1 Klystron, inductive output tube and solid-state amplifiers

The most common RF power sources are klystrons. They allow pulsed or CW operation and can work from hundreds of megahertz up to tens of gigahertz, generating peak power, in the low DC pulsed regime, up to several tens of MW. A sketch of a klystron is given in Fig. 51(a). A high DC voltage (hundreds of kV) is applied between a cathode and an anode through special devices called modulators. The DC current (up to more than 100 A) emitted by the cathode filament is then accelerated by the DC voltage and is bunched through a system of cavities. The first input cavity is fed with a driver signal that starts this bunching process, while the last output cavity is excited by the bunched beam and is coupled to a waveguide that collects the produced RF power. Finally, the collector absorbs the beam. The signal from the RF driver is, in conclusion, amplified more than 40 dB.

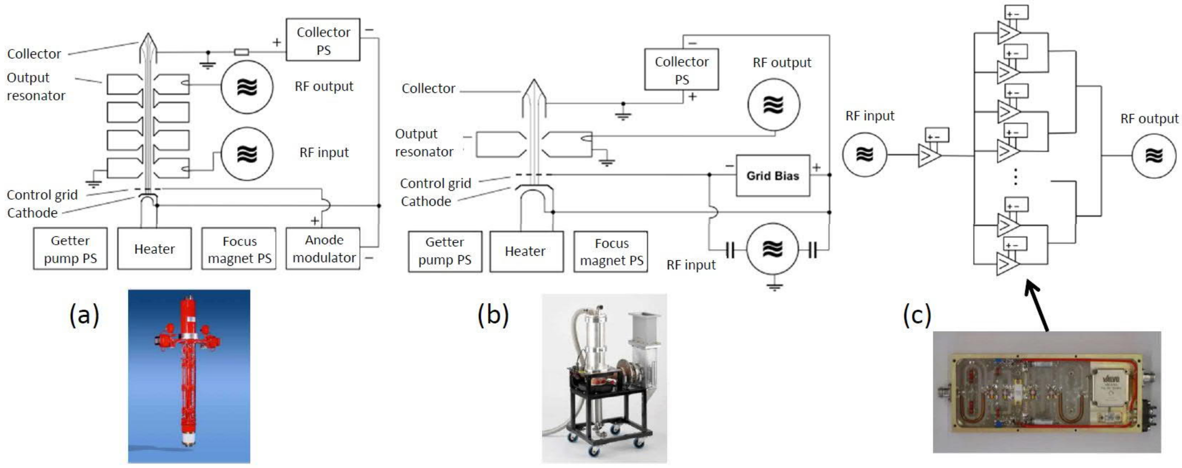


Fig. 51: (a) Sketch of a klystron; (b) schematic of an IOT; (c) schematic principle of a solid-state amplifier

At lower power (up to hundreds of kilowatts) the Inductive Output Tube (IOT) can also be used. A schematic of an IOT is given in Fig. 51(b). The intensity modulation of the DC beam is realized by a control grid. These devices can typically operate at up to 2 GHz and, in general, their efficiency is larger than that of klystrons.

Finally, solid-state amplifiers can also be used at frequencies lower than 2 GHz. A schematic of the principle of such devices is shown in Fig. 51(c). The power of many transistors is combined. The advantage of this system, with respect to vacuum tubes, is the compactness and the possibility to operate even if a single module fails. In fact, we have a reduction just of the output power. Concerning their efficiency, they can also exceed that of vacuum tubes. On the other hand, they allow operation at moderate power. Finally, Fig. 52 gives an overview of the maximum power and frequency for the different types of source.

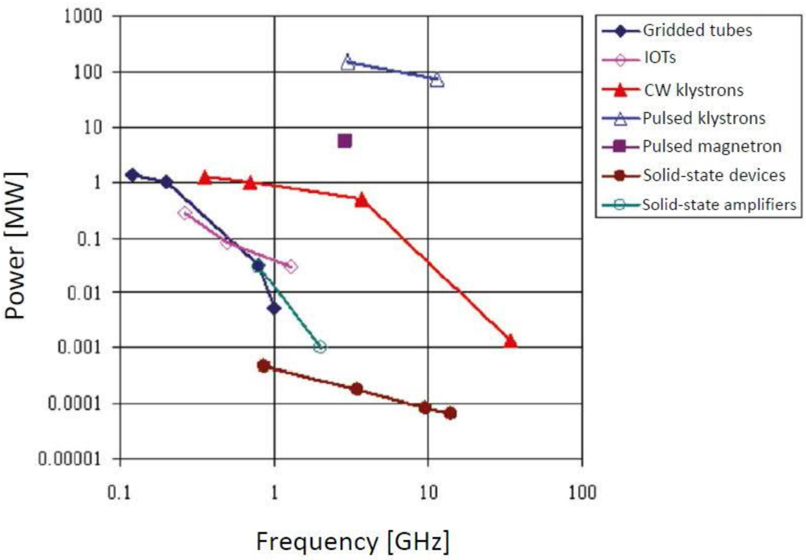


Fig. 52: Overview of the maximum power and frequency for the different type of sources (courtesy [45])

10.2 Pulse compressor systems: SLAC Energy Doubler (SLED)

As pointed out above, TW structures have to be fed by short, high peak power pulses. One way to achieve this is the energy doubler that was invented and implemented for the first time in the SLAC linac [47]. The principle of operation is given in Fig. 53. Basically, the power from the klystron is stored in special cavities and then abruptly released to the accelerator. More precisely, the waveguide system is connected through a 3 dB coupler to a pair of two cavities with a high quality factor (>10^5). At the beginning of the RF pulse, part of the klystron power goes into the cavities, building them up, while the remainder is reflected and through the 3 dB coupler is sent to the accelerator. As the stored cavity fields increase they radiate power in counter-phase with the incident power so that the mean power transmitted to the accelerator decreases. The cavities are over-coupled so that the peak reflected power rises to a level greater than the transmitted power. At a certain time (t_{switch}) the phase of the klystron drive is abruptly reversed and the reflected and transmitted signals are then in phase, causing a fast increase of the actual power transmitted to the accelerator, at the expense of pulse duration. A picture of the C-band SLED implemented in the Spring-8 FEL is given in Fig. 54.

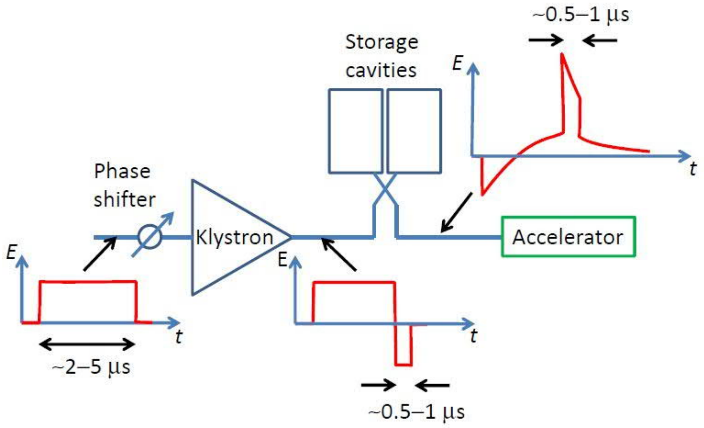


Fig. 53: Schematic layout of a SLED system

Fig. 54: Picture of the C-band SLED system implemented at Spring 8 [13, 14]

10.3 Waveguide power distribution systems

A treatment on linac technology cannot exclude a very quick overview on waveguide components that allow the transport and distribution of RF power from the source to the accelerator devices. A few of these components are shown in Fig. 55. Amongst others, we can mention the following [46].

i) Circulator (or isolator) that allows protection of the RF source from reflections due to the powering of SW structures. The circulator is a passive non-reciprocal device with three ports, and protects (isolates) the RF power sources from microwave power reflected back from non-ideal loads. This is possible due to the unique magnetic properties of ferrites that, when properly magnetized, introduce different phase shift for EM waves travelling in opposite directions.

ii) Attenuator and phase shifters that allow changes, at high power, to the phase or the power level.

iii) Directional couplers that allow the measurement of the power flowing into the waveguide in both directions (forward and reflected).

iv) Pumping ports to evacuate the waveguide and reach a vacuum level of the order of 10^{-8} mbar.

v) Ceramic windows that allow separation of the vacuum of the linac from that of the waveguides (which are generally at a higher pressure than that of the linac). Moreover, in the case of an intervention, one has to vent only a limited part of the accelerator. Several components like phase shifters, attenuators and circulators have to operate in a controlled atmosphere with gases with a high dielectric rigidity (like SF6). Ceramic windows allow the separation of the vacuum of the linac or waveguides from these regions.

vi) Bends and splitters that allow distribution of power from the source.

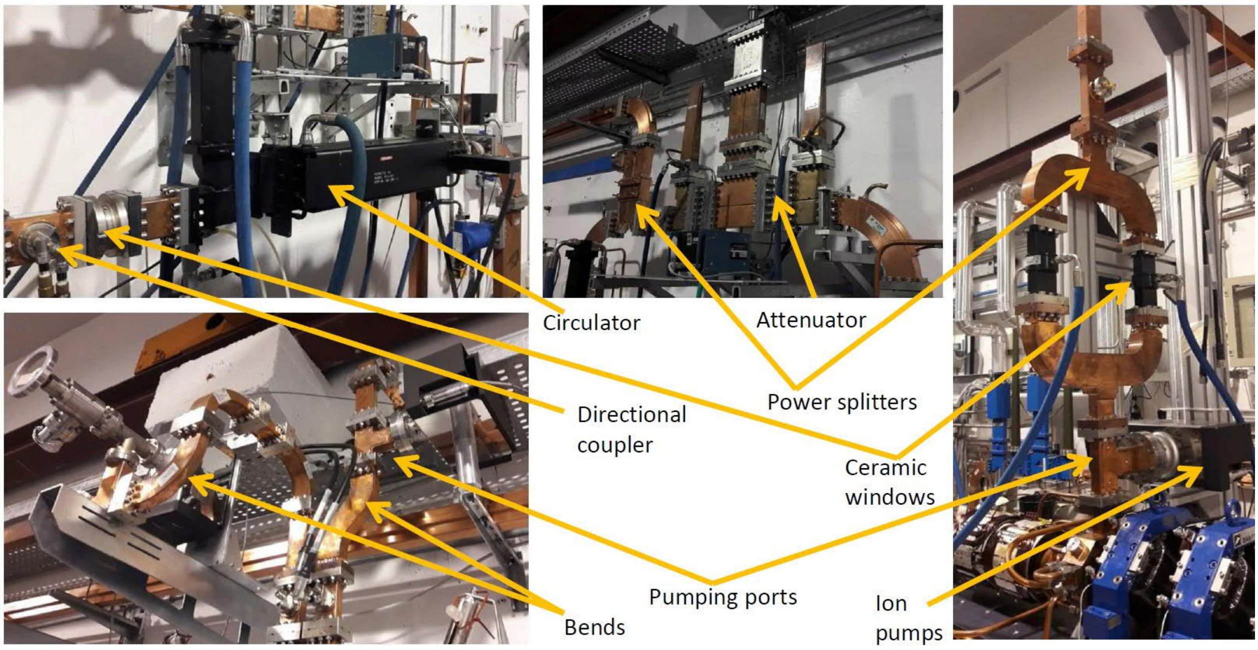


Fig. 55: Pictures of waveguide components

Acknowledgements

I would like to thank Luca Piersanti for the careful reading, corrections and helpful suggestions. Several pictures, schemes, images and plots have been taken from papers and presentations reported in the References: I would like to acknowledge all the authors.

References

[1] E. Jensen, Cavity basics, in Proceedings of the CAS–CERN Accelerator School: RF for Accelerators, Ebeltoft, Denmark, 8–17 June 2010, CERN-2011-007 (CERN, Geneva, 2010) pp. 259–275, arXiv:1201.3202.

[2] F. Gerigk, Cavity types, in Proceedings of the CAS–CERN Accelerator School: RF for Accelerators, Ebeltoft, Denmark, 8–17 June 2010, CERN-2011-007 (CERN, Geneva, 2010), pp. 277–298, arXiv:1111.4897.

[3] D. Alesini, Power coupling, in Proceedings of the CAS–CERN Accelerator School: RF for Accelerators, Ebeltoft, Denmark, 8–17 June 2010, CERN-2011-007 (CERN, Geneva, 2010), pp.125–147, arXiv:1112.3201.

[4] T. Wangler, *RF Linear Accelerators* (Wiley, New York, 1998). https://doi.org/10.1002/9783527618408

[5] DESY, Technical Design Report (TDR) of the European XFEL, 2007, http://xfel.desy.de/technical_information/tdr/tdr/.

[6] B. Aune *et al*., *Phys. Rev. ST Accel. Beams* **3**(9) (2000) 092001. https://doi.org/10.1103/PhysRevSTAB.3.092001

[7] R.E. Collin, *Foundations for Microwave Engineering*, 2nd ed. (McGraw-Hill, Singapore, 1992).

[8] R.B. Neal, *The Stanford Two-Mile Accelerator* (Benjamin, New York, 1968).

[9] T. Garvey, The SwissFEL Linac (2014), http://www-pnp.physics.ox.ac.uk/~jaiweb/slides/2014_Garvey.pdf.

[10] R. Zennaro *et al*., Measurements and high power test of the first C-band accelerating structure for SwissFEL, Proc. LINAC2014, Geneva, Switzerland, (Jacow, CERN, Geneva, 2014) p. 333.

[11] J.-Y. Raguin and M. Bopp, The Swiss FEL C-band accelerating structure: RF design and thermal analysis, Proc. LINAC2012, Tel-Aviv, Israel, (Jacow, CERN, Geneva, 2013), p. 501.

[12] R. Ganter, SwissFEL status report, Proc. FEL2015, Daejeon, Korea 2015, (Jacow, CERN, Geneva, 2015), p. 567.

[13] T. Inagaki *et al*., *Phys. Rev. ST Accel. Beams* **17**(8) (2014) 080702. https://doi.org/10.1103/PhysRevSTAB.17.080702

[14] K. Okihira, Mass production report of C-band choke mode accelerating structure and RF pulse compressor, Proc. IPAC2011, San Sebastián, Spain, 2011, (EPS-AG, Geneva, 2011), p. 668.

[15] R. Parodi, Multipacting, in Proceedings of the CAS–CERN Accelerator School: RF for Accelerators, Ebeltoft, Denmark, 8–17 June 2010, CERN-2011-007 (CERN, Geneva, 2010), pp.447–458, arXiv:1112.2176. http://dx.doi.org/10.5170/CERN-2011-007.

[16] D.C. Mattis and J. Bardeen, *Phys. Rev.* **111**(2) (1958) 412. https://doi.org/10.1103/PhysRev.111.412

[17] H. Padamsee. RF *Superconductivity: Science, Technology, and Applications* (Wiley-VCH, Weinheim, 2009). https://doi.org/10.1002/9783527627172

[18] H. Padamsee, J. Knobloch, and T. Hays. *RF Superconductivity for Accelerators* (Wiley, New York, 1998).

[19] H. Safa, Surface effects in SCRF cavity, in Proceedings of the CAS–CERN Superconductivity and Cryogenics for Accelerators and Detectors, Erice, 2002, CERN 2004-008 (CERN, Geneva, 2004), p. 196. http://dx.doi.org/10.5170/CERN-2004-008

[20] G. Bisoffi, Superconducting cavities, in Proceedings of the CAS School on Radiofrequency Engineering, CERN 2005-003 (CERN, Geneva, 2005), p. 315. http://cdsweb.cern.ch/record/386544/files/CERN-2005-003.pdf.

[21] S. Belomestnykh, Principles of RF superconductivity, lecture at USPAS 2013, Durham, NC, 2013, http://uspas.fnal.gov/materials/materials-table.shtml.

[22] G. Ciovati, Cavity fabrication, lecture at USPAS 2015, http://uspas.fnal.gov/materials/15Rutgers/Cavity_Fabrication.pdf.

[23] D. Proch, RF Cavities fabrication, in Proceedings of the CAS–CERN Accelerator School, Superconductivity and Cryogenics for Accelerators and Detectors, Erice, 2002, CERN 2004-008 (CERN, Geneva, 2004), p. 214. http://cdsweb.cern.ch/record/503603/files/CERN-2004-008.pdf.

[24] D. Alesini *et al*., *Nucl. Instrum. Meth. A* **580**(3) (2007) 1176. https://doi.org/10.1016/j.nima.2007.06.045

[25] C. Steele, *IEEE T. Microw. Theory* **14**(2) (1966) 70. https://doi.org/10.1109/TMTT.1966.1126168

[26] D. Alesini *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **837** (2016) 161. https://doi.org/10.1016/j.nima.2016.09.010

[27] D.Alesini *et al.*, *J. Instrum*. **8** (2013) P10010. https://dx.doi.org/10.1088/ 1748-0221/8/10/P10010.

[28] F. Cardelli, PhD thesis, Sapienza Università di Roma, 2017.

[29] J.K. Sekutowicz, Superconducting elliptical cavities, in Proceedings of the CERN Accelerator School RF for Accelerators, Ebeltoft, Denmark 8–17 June, 2010, CERN–2011–007 (CERN, Geneva, 2011), p. 369. http://cas.web.cern.ch/schools/ebeltoft-2010.

[30] J. Sekutowicz, Superconducting linear accelerator for the European XFEL, presentation at the International Workshop on X-ray Diagnostics and Scientific Application of the European XFEL, Ryn, Poland, 2010.

http://xfel-old.desy.de/sites/site_xfel-gmbh/content/e63594/e63599/e81230/e70177/e83636/ryn-workshop-2010-sekutowicz_eng.pdf

[31] R.F. Parodi, Couplers and HOM dampers, in Proceedings of the CAS–CERN Accelerator School, Superconductivity and Cryogenics for Accelerators and Detectors, Erice, 2002, CERN 2004-008 (CERN, Geneva, 2004), p. 253. http://cdsweb.cern.ch/record/503603/files/CERN-2004-008.pdf.

[32] P. Michelato, Cavity processing: EP/BCP, heat treatments, baking and clean room techniques, presentation at SRF2013, Caen/Paris, September 19-27, 2013. https://indico.in2p3.fr/event/9782/contributions/50504/.

[33] G. Ciovati, Surface preparation, lecture at USPAS Accelerator School, 2015, http://uspas.fnal.gov/materials/15Rutgers/Surface_Preparation.pdf.

[34] D. Alesini *et al*., Realization and high power tests of damped C-band accelerating structures for the ELI-NP linac, in proceeding of IPAC 2016, Busan, Korea, 2016, (Jacow, CERN, Gedneva, 2016) p. 399.

[35] G. Ciovati, Cavity limitations, lecture at USPAS Accelerator School, June 2015, http://uspas.fnal.gov/materials/15Rutgers/Cavity_Limitations.pdf.

[36] D. Dowell, High brightness electron injectors for light sources, lecture at USPAS School, 2010, http://uspas.fnal.gov/materials/10MIT/MIT-High-Brightness.shtml.

[37] D. Alesini *et al.*, High power test results of the ELI-NP S-band gun fabricated with the new clamping technology without brazing, in Proceeding of IPAC 2017, Copenhagen, Denmark, 2017, (Jacow, CERN, Gedneva, 2017), p. 3662.

[38] D.H. Dowell *et al.*, Results of the SLAC LCLS gun high-power RF tests, in proceeding of PAC07, Albuquerque, New Mexico, 2007, (IEEE, 2007) p. 1296. https://doi.org/10.1109/PAC.2007.4441061

[39] C. Limborg *et al.*, RF design of the LCLS Gun, LCLS-TN-05-3, 2005. http://www-ssrl.slac.stanford.edu/lcls/technotes/LCLS-TN-05-3.pdf.

[40] L. Xiao *et al.*, Dual feed RF gun design for the LCLS, in Proceedings of the 2005 Particle Accelerator Conference, Knoxville, Tennessee, 2005, (IEEE 2005) p. 3432.

https://doi.org/10.1109/PAC.2005.1591495

[41] B. Dwersteg *et al.*, *Nucl. Instrum. Meth. A* **393**(1–3) (1997) 93–95. https://doi.org/10.1016/S0168-9002(97)00434-8

[42] F. Stephan *et al.*, *Phys. Rev. ST Accel. Beams* **13**(12) (2010) 020704. https://doi.org/10.1103/PhysRevSTAB.13.020704

[43] C. Gulliford *et al.*, *Phys. Rev. ST Accel. Beams* **16**(7) (2013) 073401. https://doi.org/10.1103/PhysRevSTAB.16.073401

[44] R.G. Carter, Review of RF power sources for particle accelerators, in Proceedings of the CAS–CERN Accelerator School, Radio Frequency Engineering, Seeheim, Germany, 8–16 May 2000, CERN 2005-003 (CERN, Geneva, 2005), p. 107. http://cdsweb.cern.ch/record/386544/files/CERN-2005-003.pdf.

[45] R.G. Carter, RF power generation, in Proceedings of the CERN Accelerator School RF for accelerators, Ebeltoft, Denmark 8–17 June 2010, CERN-2011-007 (CERN, Geneva, 2011), p. 173. http://dx.doi.org/10.5170/CERN-2011-007.

[46] R.K. Cooper and R.G. Carter, High-power RF transmission, in Proceedings of the CAS–CERN Accelerator School: Radio Frequency Engineering, Seeheim, Germany, 8–16 May 2000, CERN 2005-003 (CERN, Geneva, 2005), p. 210. http://cdsweb.cern.ch/record/386544/files/CERN-2005-003.pdf.

[47] Z. Farkas *et al.*, *IEEE Trans. Nucl. Sci.*, **22**(3) (2007) 1299. https://doi.org/10.1109/TNS.1975.4327871

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Wakefields—An Overview

F. Reimann and U. van Rienen
University of Rostock, Rostock, Germany

Abstract
In this paper, we will report on the basic concepts of ultra-relativistic wakefields and their effects on particles while focusing; in this context, particularly on particle accelerators. We will introduce the commonly used terminology, and will derive and explain important quantities, such as the wake potential and wake function, the impedance as the Fourier transform of the wake potential, and loss parameters. To deepen knowledge on wake functions and potentials further, we will illustrate the derived quantities using, as examples, the cylindrical cavity and a rectangular waveguide lined with dielectrics.

Keywords
Wakefields; Green's functions; loss parameters; impedances.

1 Introduction

1.1 The term 'wakefield'

Outside of accelerator physics, the term 'wake' is mostly known from fluid dynamics, referring to the wave pattern behind an object moving in a liquid (e.g., a ship moving in water). This also comes to mind first if the term is used in everyday language.

In accelerator physics, the word 'wakefield' has a different meaning, as it describes an electromagnetic effect created by charged particles (see next paragraph). However, it is not completely wrong to think of the wakefield as a certain field pattern that follows a charged particle, as water waves follow a ship.

Wakefields, in the context of accelerator physics, are generated by a charged particle that travels through a metallic vacuum chamber. The self-field of an ultra-relativistic particle ends perpendicular to the highly conductive walls. On the surface of the walls, image charges are created (electric polarization), which turn into the sources of new fields and act back on the particle. In a metallic vacuum chamber without any geometric variation, the image charges travel together with the ultra-relativistic particle. Any geometric variation forces the field lines to bend, since they still need to stay perpendicular to the conductive walls. Then, some parts of the fields (and the energy stored therein) stay behind and consequently trail behind the particle. These fields are denoted wakefields. If a second particle follows the first closely enough, it will still see the wakefields of the first particle and interact with them. In a bunch of charged particles, the trailing particles of the bunch will see the wakefields of the leading particles and interact with them.

Figure 1 illustrates this process. A Gaussian pulse enters a so-called pillbox cavity. A pillbox cavity consists of a cylindrical cavity resonator with round openings that are attached to the beam pipe. The beam axis equals the symmetry axis of the pillbox cavity. At the transitions between the cylindrical cavity and the beam pipes, the diameter of the vacuum chamber changes so that a wakefield can be generated—a figurative description would say that part of the self-field of the bunch is 'stripped off' by the geometric changes in the structure. The wakefield remains in the structure and oscillates for some time after the bunch has left. A second particle bunch that traverses the structure would consequently be influenced by the generated wakefield. The pillbox cavity used in Fig. 1 has a radius of $R = 5\,\mathrm{cm}$ and a length of $R = 10\,\mathrm{cm}$ and will be used in examples throughout this paper as a model pillbox.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.125

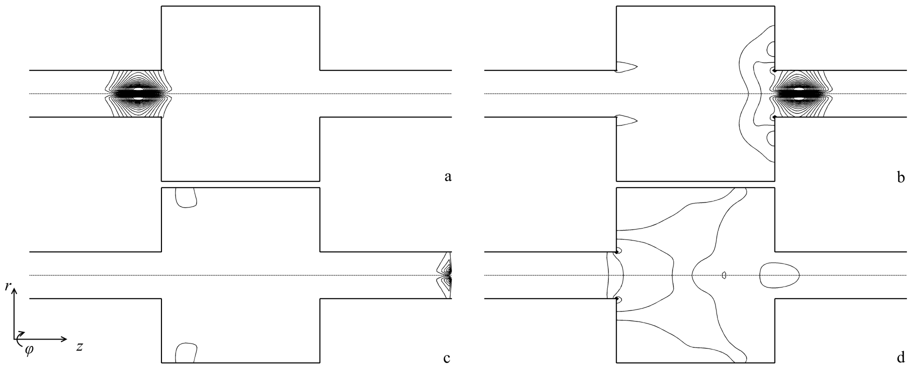


Fig. 1: Electric field lines of a Gaussian pulse travelling through the model pillbox cavity ($R = 5\,\mathrm{cm}, g = 10\,\mathrm{cm}$), calculated using the CST STUDIO SUITE® [1]: (a) while entering the structure; (b) leaving the cavity; (c) leaving the structure; (d) after leaving. It can be seen that a part of the electric field remains in the structure even after the bunch has left.

The word 'wakefield' is usually understood as a general term. More specifically, we speak about 'wake potentials', when considering the wakefield behind a particle bunch, and 'wake functions', when the wakefield behind a point charge is considered. These expressions will be explained in more detail in subsequent sections.

This report follows the lines of Ref. [2]. For further reading on wakefields and linear accelerators, please refer to Refs. [3–5].

1.2 Basic concepts of ultra-relativistic wakefields

To understand the underlying principles of wakefields, we first need to understand what happens to a point charge q that moves in free space with a velocity close to the speed of light, $v \approx c$.

Owing to the Lorentz contraction, the electromagnetic field of the electron will be shrunk to a thin disk perpendicular to its moving direction. The opening angle of the field travelling with the particle is given by $1/\sqrt{1-\beta^2} = 1/\gamma$ with the factor $\beta = v/c$. If the velocity approaches the speed of light, the thickness of the disk shrinks further, to a δ-distribution (see Fig. 2). This field is strictly radial, i.e., there are no components of the field *behind* or *in front* of the charge, which is also a consequence of the *principle of causality*. Accordingly, in this case, there can be no wakefield behind the electron in free space.

To actually achieve a field and a force behind the field-generating charge, additional mechanisms are required. For example, the image charges and fields created on the waveguide walls are only synchronous with the fields generating them if the walls are perfectly conducting. In resistive or imperfectly conducting walls, the image fields will trail behind the field-generating charge. Other possibilities include obstacles in the beam pipe, e.g., geometric variations (cf. Fig. 1), from which the fields are scattered. Another possibility is to introduce dielectric walls because the speed of light will be lower here than in a vacuum. This means that the fields are 'slowed down', in the sense that travelling waves in these media will have a lower phase velocity and thus trail behind the generating fields.

2 Basic definitions

In the following, we will consider a field-generating charge q_1, which is located at the three-dimensional coordinate $\mathbf{r}$. First, we want to examine the electromagnetic force that this field-generating charge exerts on a test charge q_2 that moves at the speed of light along the z-axis, $\mathbf{v} = c\mathbf{e}_z$. This force is simply the

Lorentz force,

$$\mathbf{F}(\mathbf{r},t) = q_2 \left(\mathbf{E}(\mathbf{r},t) + c\mathbf{e}_z \times \mathbf{B}(\mathbf{r},t)\right) . \tag{1}$$

We introduce a new variable for the distance between q_2 and q_1 (see Fig. 3), so that

$$s = ct - z \, ,$$

and

$$\mathbf{F}(s,t) = \mathbf{F}(x,y,z = ct - s, t) \, .$$

The net momentum change δp of the test charge due to the Lorentz force will then be

$$\delta p \sim \int \mathbf{F}(s,t)\,\mathrm{d}t \, . \tag{2}$$

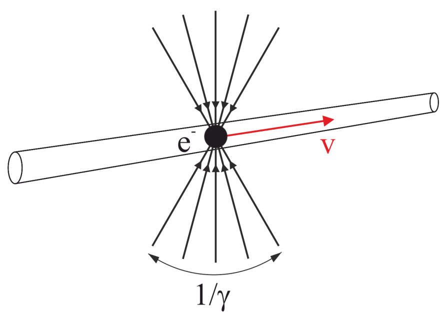


Fig. 2: Radial electric field of an electron moving at the speed of light, contracted to a disc.

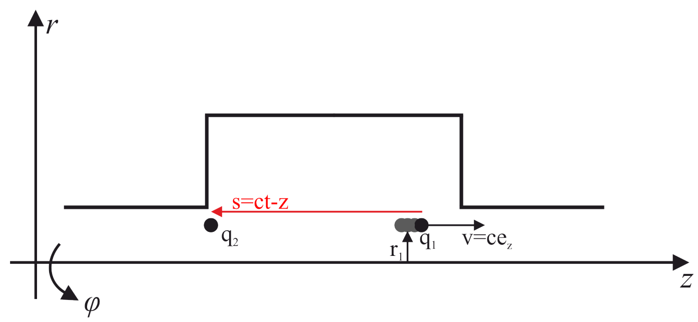


Fig. 3: Field-generating and test charges in a pillbox cavity. The grey dots represent the possibility of a particle bunch generating the wakefield instead of a single charge.

2.1 The wake potential

The concept of wake potentials is related to the concept of the momentum change on a test charge, described before. Consider the situation described in Fig. 3: a pillbox cavity with its transitions between the cylindrical cavity and the beam pipes introduces a radial change of the vacuum chamber and thus wakefields can be generated. Additionally, both charges are assumed to have a transverse offset $\mathbf{r}_1$ from the centre of the beam pipe, while the movement is still parallel to the z-axis. In cylindrical coordinates, the described situation is as shown in Fig. 3. For this case, the *three-dimensional wake potential* is defined as

$$\mathbf{W}(\mathbf{r}_1,s) = \frac{1}{q_1}\int\limits_{-\infty}^{\infty} \left[\mathbf{E}(\mathbf{r}_1,z,t) + c\mathbf{e}_z \times \mathbf{B}(\mathbf{r}_1,z,t)\right]_{t=(z+s)/c} \mathrm{d}z \, , \tag{3}$$

which is basically an integral over the Lorentz force evaluated on the beam axis and normalized to the field-generating charge. Additionally, time is substituted with $t = z + s/c$. The momentum change of the test charge is related to this via

$$\delta p = q_1 q_2 \mathbf{W}(s) \, . \tag{4}$$

Usually, the wake potential is separated into the *longitudinal wake potential* and the *transverse wake potential*.

For the longitudinal wake potential, the projection of the Lorentz force onto the z-axis is used:

$$\mathbf{W}(\mathbf{r}_1,s)\cdot\mathbf{e}_z = \frac{1}{q_1}\int\limits_{-\infty}^{\infty} \left[\mathbf{E}(\mathbf{r}_1,z,t)\cdot\mathbf{e}_z + c\left(\mathbf{e}_z \times \mathbf{B}(\mathbf{r}_1,z,t)\right)\cdot\mathbf{e}_z\right]_{t=(z+s)/c} \mathrm{d}z \, .$$

Here, the second term vanishes because $\mathbf{e}_z \cdot (\mathbf{e}_z \times \mathbf{B}(\mathbf{r}_1, z, t)) = 0$. The longitudinal component of the wake potential is thus only dependent on the electric field:

$$W_{||}(\mathbf{r}_1, s) = \frac{1}{q_1} \int\limits_{-\infty}^{\infty} E_z \left(\mathbf{r}_1, z, \frac{z+s}{c} \right) \mathrm{d}z \,. \tag{5}$$

Consequently, the transverse wake potential is only dependent on the transverse components of the electric and magnetic field:

$$\mathbf{W}_{\perp}(\mathbf{r}_1, s) = \frac{1}{q_1} \int\limits_{-\infty}^{\infty} \left[\mathbf{E}_{\perp}(\mathbf{r}_1, z, t) + c\mathbf{e}_z \times \mathbf{B}_{\perp}(\mathbf{r}_1, z, t) \right]_{t=(z+s)/c} \mathrm{d}z \,. \tag{6}$$

Both wake potentials are dependent on the distance s between the field-generating charges and the test charge. This distance is measured in the negative longitudinal direction (see Fig. 3). This means that a negative distance s corresponds to the case in which the test charge is *in front* of the field-generating charges. Owing to the principle of causality, in this case there can be *no wake potential*. Consequently, this means that:

$$W_{||}(\mathbf{r}_1, s) = 0 \text{ for } s < 0, \text{ and} \tag{7}$$

$$\mathbf{W}_{\perp}(\mathbf{r}_1, s) = 0 \text{ for } s < 0 \,. \tag{8}$$

Example wake potentials of Gaussian pulses inside a pillbox cavity are shown in Figs. 4 and 5.

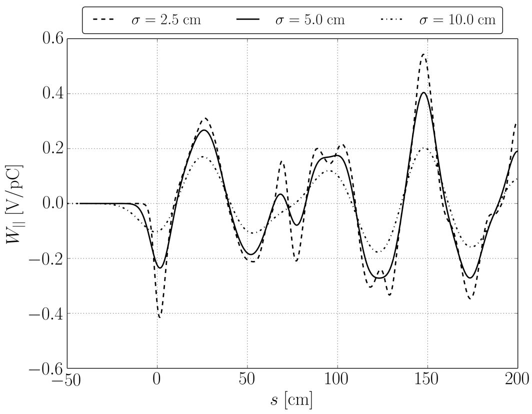


Fig. 4: Longitudinal wake potentials of Gaussian pulses with different pulse width σ inside the model pillbox cavity, computed using CST STUDIO SUITE ® [1].

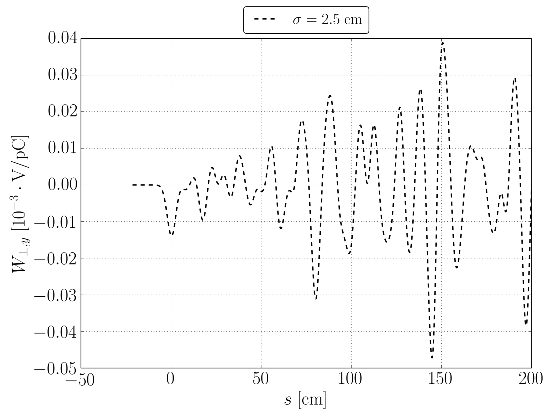


Fig. 5: Transverse wake potential of a Gaussian pulse of width $\sigma = 2.5$ cm inside the model pillbox cavity, computed using CST STUDIO SUITE ®. Note that the transverse wake potential is several orders of magnitude weaker than the longitudinal wake potential of the same pulse, displayed in Fig. 4.

2.2 The Panofsky–Wenzel theorem

The Panofsky–Wenzel theorem connects the longitudinal and transverse wake potentials via

$$\mathbf{W}_{\perp}(x, y, s) = -\nabla_{\perp} \int\limits_{-\infty}^{s} \mathbf{W}_{||}\left(x, y, s'\right) \mathrm{d}s' \,. \tag{9}$$

Therefore, in principle, knowledge of only the longitudinal component of the wake potential is enough, since the transverse component can be constructed from it.

In the following, we want to briefly sketch the proof of this theorem.

We start with the transverse wake potential from Eq. (6). Its derivative with respect to s is

$$\frac{\partial}{\partial s}\mathbf{W}_\perp(\mathbf{r}_1, s) = \frac{1}{q_1}\int_{-\infty}^{\infty}\left[\underbrace{\frac{\partial}{c\partial t}\mathbf{E}_\perp(\mathbf{r}_1, z, t)}_{T_1} + \underbrace{\mathbf{e}_z \times \frac{\partial}{\partial t}\mathbf{B}_\perp(\mathbf{r}_1, z, t)}_{T_2}\right]_{t=(z+s)/c} \mathrm{d}z\,, \tag{10}$$

where we used $s = ct - z$ and $\partial s = c\partial t$.

Now we want to replace T_1 and T_2 with more convenient expressions. For term T_1, we need the total derivative of the transverse electric field with respect to z. Using $s = ct - z$ this reads as

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right) = \left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\,,$$

where $\mathrm{d}/\mathrm{d}z$ is the total differential with respect to z. We reformulate T_1 in Eq. (10) as

$$\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right) = \left(\frac{\mathrm{d}}{\mathrm{d}z} - \frac{\partial}{\partial z}\right)\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\,. \tag{11}$$

For term T_2, we start with Faraday's law of induction,

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}, t)\,.$$

Computing the cross product of this equation with the beam axis leads to

$$\mathbf{e}_z \times \frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}, t) = -\mathbf{e}_z \times (\nabla \times \mathbf{E}(\mathbf{r}, t)) = \frac{\partial}{\partial z}\mathbf{E}_\perp(\mathbf{r}, t) - \nabla_\perp E_z(\mathbf{r}, t)\,. \tag{12}$$

Inserting Eqs. (11) and (12) into Eq. (10) results in

$$\begin{aligned}\frac{\partial}{\partial s}\mathbf{W}_\perp(\mathbf{r}_1, s) = \frac{1}{q_1}\int_{-\infty}^{\infty} &\left(\frac{\mathrm{d}}{\mathrm{d}z} - \frac{\partial}{\partial z}\right)\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right) \\ &+ \left(\frac{\partial}{\partial z}\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right) - \nabla_\perp E_z\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\right)\mathrm{d}z\,.\end{aligned}$$

We reformulate this as

$$\frac{\partial}{\partial s}\mathbf{W}_\perp(\mathbf{r}_1, s) = \frac{1}{q_1}\int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}z}\mathbf{E}_\perp\left(\mathbf{r}_1, z, \frac{z+s}{c}\right) - \nabla_\perp\mathbf{E}_z\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\mathrm{d}z\,.$$

We assume perfect electric conductor boundary conditions at the waveguide walls so that the tangential electric field vanishes there. This simplifies Eq. (10) to

$$\frac{\partial}{\partial s}\mathbf{W}_\perp(\mathbf{r}_1, s) = -\frac{1}{q_1}\int_{-\infty}^{\infty}\nabla_\perp\mathbf{E}_z\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\mathrm{d}z\,,$$

which is equivalent to

$$\frac{\partial}{\partial s}\mathbf{W}_\perp(\mathbf{r}_1, s) = -\nabla_\perp W_{||}(\mathbf{r}_1, s)\,.$$

Integrating the last statement over s leads to Eq. (9).

2.3 The fundamental theorem of beam loading

This section follows the lines of Ref. [5].

Up to this point, we only defined wakefields for $s > 0$. For $s < 0$, we concluded from the principle of causality that there can be no wakefield, and thus $\mathbf{W}(\mathbf{r}_1, s) = 0$.

We now want to consider the case of $s = 0$, which we had excluded before. For this, we will first look at a different example situation: two particles with equal charge q and a distance of a half wavelength, $\lambda/2$, between them are moving along the same axis, at the same speed. The first charge enters a previously empty cavity with no internal electric fields or stored energy (see Fig. 6). The charge will induce surface charges, electric fields, and voltages in the cavity,

$$V_i = -\int_C \mathbf{E}\,\mathrm{d}\mathbf{l}\,,$$

where the voltage is defined as a line integral over the electric field along a closed path C.

We will refer to this induced voltage as $-V_i$. This induced voltage is left in the cavity, even after the first charge left. From the law of energy conservation, we also infer that energy must be left behind in the cavity. However, the first charge will also 'see' a fraction a of its own induced voltage while in the cavity,

$$V_1 = -aV_i\,.$$

This corresponds to an energy loss of the first particle,

$$\Delta W_1 = qV_1 = -qaV_i\,.$$

Thus, with the first particle in the cavity, the net cavity voltage is $V_\mathrm{c} = -V_i$, while the stored energy U will be proportional to this voltage squared, $U \propto V_i^2$. This situation is shown in Fig. 7.

When the second particle arrives in the cavity, the voltage induced by the first particle will have changed phase by π, owing to the distance between the two particles. Thus, the induced voltage from particle 1 is now $+V_i$. The second particle, however, will *also* induce a voltage in the cavity of $-V_i$. The net cavity voltage will be

$$V_\mathrm{c} = +V_i - V_i = 0\,.$$

Particle 2 will also lose energy according to

$$\Delta W_2 = \underbrace{qV_i}_{\text{from particle 1}} - \underbrace{qaV_i}_{\text{from own induced voltage}}\,.$$

Since the net energy of the cavity must remain 0, the energy changes of particle 1 and 2 have to compensate each other (see Fig. 8),

$$\Delta W_1 + \Delta W_2 = 0\,,$$
$$qV_i - qaV_i - qaV_i = 0\,.$$

This leads directly to

$$a = \frac{1}{2}\,.$$

From this, we can directly derive the *fundamental theorem of beam loading*: a moving charge will experience (or 'see') half of its own induced voltage.

For the case of the wake potential, this implies that, for $s = 0$, when the field-generating and test charges are virtually at the same place, the wake potential must be multiplied by $1/2$. Thus, the final definition of the longitudinal wake potential is:

$$W_{||}(\mathbf{r}_1, s) = \frac{1}{q_1}\int\limits_{-\infty}^{\infty} E_z\left(\mathbf{r}_1, z, \frac{z+s}{c}\right)\mathrm{d}z \begin{cases} 0 & \text{for } s < 0 \\ \frac{1}{2} & \text{for } s = 0 \\ 1 & \text{for } s > 0 \end{cases}\,. \tag{13}$$

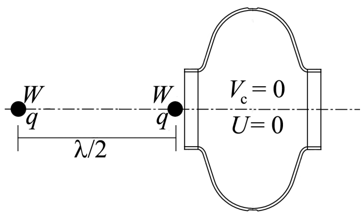


Fig. 6: Two charges separated by a distance $\lambda/2$ travelling along the same beam axis at the same speed are about to enter a previously empty elliptic cavity.

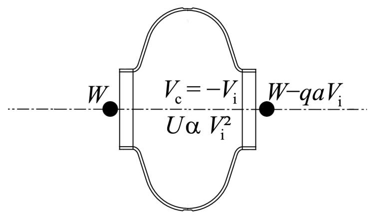


Fig. 7: The first charge has traversed the cavity and induced an image voltage of $-V_i$ in the cavity.

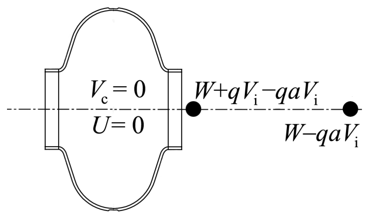


Fig. 8: Both charges have left the cavity. The total energy change in the cavity is 0, while both charges have lost energy, owing to the image fields they experienced. From this, the proportionality factor a can be calculated.

3 Impedances and loss parameters

3.1 Wakefields and impedances

The *impedance* is a physical quantity closely related to the wake potential by the Fourier transform:

$$Z_{||}\left(x,y,\omega\right) = \frac{1}{c}\int\limits_{-\infty}^{\infty} W_{||}\left(x,y,s\right)\exp\left(-\mathrm{i}\frac{\omega}{c}s\right)\mathrm{d}s\;. \tag{14}$$

Physically, the impedance and the wake potential describe one and the same effect; namely, the coupling between the beam and its environment. In contrast with the wake potential, which is described in the time domain, the impedance is described in the frequency domain. The impedance corresponds to a frequency spectrum that shows which of the structure's *eigenmodes* couple with the beam. The amplitude of the impedance in the frequency spectrum can also indicate the mode's coupling strength.

For more information on impedances, see Ref. [2].

3.2 Loss parameters

In the following, we want to have a closer look at the coupling strength of eigenmodes.

In general, it is possible to expand any physical quantity into a complete set of orthonormal functions. While in theory, *any* complete set of orthonormal functions will suffice, the choice of a suitable set of functions will usually decrease the effort for such an expansion while simultaneously increasing the numerical accuracy. For waveguide structures, the eigenmodes of the structure represent a suitable set of functions into which other fields can be expanded, e.g., the electric and magnetic field. This is often done if a straightforward solution of Maxwell's equation to obtain these fields directly is either very difficult or outright impossible.

In this case, let us assume that we expand the electric field inside an arbitrary cavity or waveguide into a set of this structure's spatial eigenmodes $\mathbf{E}_n\left(\mathbf{r}\right)$,

$$\mathbf{E}\left(\mathbf{r},t\right) = \sum_{n=0}^{\infty}\chi_n\left(t\right)\mathbf{E}_n\left(\mathbf{r}\right)\;. \tag{15}$$

This expansion is analytically correct as long as the upper limit of the summation is infinity. Of course, this is not feasible for practical use. For the moment, however, it shall be sufficient to note that if the summation is ceased after a finite number of terms, the expansion will be a mere *approximation* of the analytically correct result, and that the quality of the approximation strongly depends on the number of expansion functions used.

From the spatial electric fields, the energy stored in each of these eigenmodes can be computed according to

$$U_n = \frac{\epsilon_0}{2} \int |\mathbf{E}_n(\mathbf{r})|^2 \, \mathrm{d}^3 r \,. \tag{16}$$

A point charge moving at c along the beam axis will experience a voltage drop (per mode) of

$$V_n = \int_{-\infty}^{\infty} E_{z,n}(z) \exp\left(\mathrm{i}\frac{\omega_n z}{c}\right) \mathrm{d}z, \tag{17}$$

where ω_n is the eigenfrequency of the mode.

The *loss parameter* k_n of an eigenmode is defined using these two quantities,

$$k_n = \frac{|V_n|^2}{4U_n} \,. \tag{18}$$

It is independent of the phase and frequency of the eigenmode and represents the eigenmode's coupling strength to the beam and thus its contribution to the wakefield. Moreover, it describes how much energy a point charge loses into a mode n by

$$\Delta W_n = q_1^2 k_n \,.$$

Another useful aspect of the loss parameters is that they are closely related to the wakefield of a point charge (called the *wake function*),

$$W_{||,0} = \sum_{n=0}^{\infty} 2k_n \cos\left(\frac{\omega_n s}{c}\right) \begin{cases} 0 & \text{for } s < 0 \\ \frac{1}{2} & \text{for } s = 0 \\ 1 & \text{for } s > 0 \end{cases} . \tag{19}$$

The wake function is a sum of the contributions from all modes, and can be calculated from just the knowledge of the loss parameters and the eigenfrequencies of the modes. Its main merit is that it serves as a *Green's function*, i.e., the wake potentials of any arbitrary bunch shape can be derived from it by convolution.

We assume a bunch with the shape function $\psi(s)$, i.e., a normalized distribution of particles measured relative to the field-generating particle. This bunch's wake potential can be obtained from the wake function via a convolution with the bunch shape function,

$$W_{||}(s) = \int_{0}^{\infty} \psi\left(s - s'\right) W_{||,0}\left(s'\right) \mathrm{d}s' \,. \tag{20}$$

This relation makes the wake function a very versatile quantity. For simple geometries, it is possible to calculate the eigenmodes, and thus the wake function, analytically. In general, there exist several error sources, owing to the different kinds of approximation needed. First, there is truncation error in the series expansion. Second, for customary accelerating cavities, it is usually necessary to compute the eigenmodes numerically, thus introducing a numerical error. This can render the numerical determination of the wake function infeasible in a number of cases. However, numerical software, as e.g., CST STUDIO SUITE®, is able to compute the wake potential but not the wake function, as this would require modelling of the idealized point charge. Such software directly computes the wake potential of the studied bunch shape. Should the wake function be needed and not be available analytically, it is then often replaced with the wake potential of a very short Gaussian bunch (since an infinitesimally short Gaussian pulse would represent a point charge). Again, this is an approximation and can lead to systematic errors.

Another quantity related to the wake potential and the bunch shape function is the total loss parameter of the bunch,

$$k_{\mathrm{tot}} = \int_{-\infty}^{\infty} \psi(s)\, W_{||}(s)\, \mathrm{d}s\,. \tag{21}$$

The total loss parameter can give information about the total power lost due to wake potential, P_{tot} via

$$P_{\mathrm{tot}} = Iqk_{\mathrm{tot}}\,, \tag{22}$$

where I denotes the electric current.

4 Example 1: cylindrical cavity

4.1 The eigenmodes

In this section, we want to examine the longitudinal wake potential inside a cylindrical cavity. We assume that the cavity has a radius of R and a width of g, and that all the cavity walls are perfectly conducting. For the sake of simplicity, we also assume that the point charge traverses the cavity on the z-axis so that $\mathbf{r}_1 = \mathbf{0}$.

A pillbox cavity with beam pipes, as shown in Fig. 3, could also be regarded as such a cylindrical structure below the cut-off frequencies of the interesting eigenmodes (a length of the beam pipes above the attenuation length of these modes). For the rest of this subsection, though, we will restrict our considerations to the ideal cylindrical cavity.

First, we regard the eigenmodes of the evacuated cavity. For this example, it is possible to obtain an analytical expression for the eigenmodes by solving Maxwell's equations inside the cavity (for a harmonic time dependence, $\rho = 0$ and $\mathbf{j} = \mathbf{0}$):

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = 0\,, \tag{23}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0\,, \tag{24}$$

$$\nabla \times \mathbf{E}(\mathbf{r}) = -\mathrm{i}\omega \mathbf{B}(\mathbf{r})\,, \tag{25}$$

$$\nabla \times \mathbf{B}(\mathbf{r}) = \mathrm{i}\frac{\omega}{c^2}\mathbf{E}(\mathbf{r})\,. \tag{26}$$

Generally, two types of eigenmode can exist in cavities. *Transverse electric (TE)* modes do not exhibit an electric field component in the longitudinal direction (in most cases, this means $E_z = 0$); for *transverse magnetic (TM)* modes the magnetic field is zero in the longitudinal direction ($B_z = 0$). (The third type, the *transverse electromagnetic (TEM)* mode, cannot exist in a cylindrical cavity. Its occurrences are limited to geometries in which two isolated conductors exist, such as in coaxial cables, or to mode considerations in structures that are not electrically conducting.) Since we want to determine the longitudinal wakefield, which is an integral over the longitudinal electric field according to Eq. (5), we do not need to consider TE modes for this purpose, since their electric field along the z-axis is zero. Therefore, we will restrict our consideration to TM modes.

Choosing a vector potential $\mathbf{A}(\mathbf{r})$ and defining the magnetic field as its rotation,

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})\,, \tag{27}$$

automatically fulfils Eq. (24), since $\nabla \cdot \mathbf{B} = \nabla \cdot (\nabla \times \mathbf{A}) = 0$. Additionally, since for all TM modes $B_z = 0$, it is convenient to choose the vector potential parallel to the z-axis:

$$\mathbf{A}(\mathbf{r}) = A(\mathbf{r})\, \mathbf{e}_z\,. \tag{28}$$

In this way, Eq. (27) automatically results in a magnetic field with $B_z = 0$.

The dependence of the electric field on the vector potential can be determined by plugging Eq. (27) into Eq. (25),

$$\nabla \times \mathbf{E}(\mathbf{r}) = -\mathrm{i}\omega \nabla \times \mathbf{A}(\mathbf{r}) \ .$$

This means that, up to the gradient of a scalar potential ϕ, the electric field and the vector potential are equivalent,

$$\mathbf{E}(\mathbf{r}) = -\mathrm{i}\omega \left(\mathbf{A}(\mathbf{r}) + \nabla\phi\right) \ . \tag{29}$$

Inserting Eq. (27) into Eq. (26), and using fundamental vector algebra, yields:

$$\nabla \times \nabla \times \mathbf{A}(\mathbf{r}) = \nabla \cdot (\nabla \cdot \mathbf{A}(\mathbf{r})) - \nabla^2 \mathbf{A}(\mathbf{r}) = \mathrm{i}\frac{\omega}{c^2}\mathbf{E}(\mathbf{r}) \ .$$

After we insert Eq. (29) into this equation, we can choose the gauge $\phi = \frac{c^2}{\omega^2}\nabla \cdot \mathbf{A}$,

$$\begin{aligned} \nabla \cdot (\nabla \cdot \mathbf{A}(\mathbf{r})) - \nabla^2 \mathbf{A}(\mathbf{r}) &= \frac{\omega^2}{c^2}\left(\mathbf{A}(\mathbf{r}) + \nabla\phi\right) \\ &= \frac{\omega^2}{c^2}\mathbf{A}(\mathbf{r}) - \frac{\omega^2}{c^2}\nabla \cdot \left(\frac{c^2}{\omega^2}\nabla \cdot \mathbf{A}(\mathbf{r})\right) \ , \end{aligned}$$

and eliminate two terms.

Introducing the wavenumber $k^2 = \omega^2/c^2$, this leads to the Helmholtz equation for the vector potential,

$$\nabla^2 \mathbf{A}(\mathbf{r}) + k^2 \mathbf{A}(\mathbf{r}) = 0 \, . \tag{30}$$

Since we chose the vector potential to be parallel to the z-axis and the Laplacian is a scalar operator, we can simplify this equation to

$$\nabla^2 A(\mathbf{r}) + k^2 A(\mathbf{r}) = 0 \, .$$

It is convenient to work in cylindrical coordinates in this case. In cylindrical coordinates, the Helmholtz equation reads

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)A(r,\varphi,z) + \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}A(r,\varphi,z) + \frac{\partial^2}{\partial z^2}A(r,\varphi,z) = -k^2 A(r,\varphi,z) \ . \tag{31}$$

To solve this equation, we employ a separation ansatz for A. We assume that $A(r,\varphi,z) = A_r(r)\,A_\varphi(\varphi)\,A_z(z)$. Using this ansatz and dividing by $A(r,\varphi,z)$ leads to three separate equations:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)A_r(r) = -k_r^2 A_r(r) \ , \tag{32}$$

$$\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}A_\varphi(\varphi) = -k_\varphi^2 A_\varphi(\varphi) \ , \tag{33}$$

$$\frac{\partial^2}{\partial z^2}A_z(z) = -k_z^2 A_z(z) \ , \tag{34}$$

together with the separation equation $k^2 = k_r^2 + k_\varphi^2 + k_z^2$.

The solution of Eq. (34) can be expressed generally as a linear combination of sine and cosine functions,

$$A_z(z) = C_{z,1}\sin(k_z z) + C_{z,2}\cos(k_z z) \ . \tag{35}$$

For the solution of Eq. (33), we want to reformulate the equation first,

$$\frac{\partial^2}{\partial\varphi^2}A_\varphi(\varphi) = -\underbrace{k_\varphi^2 r^2}_{=m^2} A_\varphi(\varphi) \ .$$

The solution of this equation, like the solution of A_z, can be expressed as a linear combination of sine and cosine functions. Moreover, because of the periodicity of the structure regarding the angle φ, the condition $A_\varphi(\varphi) = A_\varphi(\varphi + 2\pi)$ must be met. From this, we can deduce that m has to be an integer. Without any loss of generality, the origin can always be set such that either a sine or a cosine function is sufficient to express A_φ. Since the sine function would be zero for $m = 0$, we choose the cosine, so that

$$A_\varphi(\varphi) = C_\varphi \cos(m\varphi) \;, \text{ with } m = 0, 1, 2, 3 \ldots \tag{36}$$

With these two solutions at hand, we now examine the radial equation, Eq. (32). Using the separation equation to replace k_r, it can be reformulated as

$$\frac{1}{r}\frac{\partial}{\partial r}A_r(r) + \frac{\partial^2}{\partial r^2}A_r(r) = -\left(k^2 - k_z^2 - k_\varphi^2\right)A_r(r) \;.$$

We now multiply the whole equation with r^2 and substitute $k_\varphi^2 = m^2/r^2$, so that

$$r^2\frac{\partial^2}{\partial r^2}A_r(r) + r\frac{\partial}{\partial r}A_r(r) = -r^2\left(k^2 - k_z^2 - \frac{m^2}{r^2}\right)A_r(r) \;. \tag{37}$$

The solutions of this equation are the so-called *Bessel functions* Z_m. Accordingly, the solution can be written as

$$A_r(r) = Z_m\left(r\sqrt{k^2 - k_z^2}\right) \;.$$

For a restricted area, like the cylindrical cavity, only the Bessel functions of the first kind, J_m, and the Bessel functions of the second kind, Y_m, must be considered. Since the functions of the second kind diverge for $r \to 0$, we can restrict the solution to the functions of the first kind,

$$A_r(r) = J_m(Kr) \;, \tag{38}$$

where we introduce $K = \sqrt{k^2 - k_z^2}$.

The complete solution for the vector potential $\mathbf{A}(r)$ is a product of Eqs. (38), (36), and (35):

$$\mathbf{A}(r, \varphi, z) = C_\varphi J_m(Kr)\cos(m\varphi)\left(C_{z,1}\sin(k_z z) + C_{z,2}\cos(k_z z)\right)\mathbf{e}_z \;. \tag{39}$$

The resulting components of the electric and magnetic field can be determined by employing the curl in cylindrical coordinates and using Eqs. (27), (26), and (29):

$$\begin{aligned}
B_r(r, \varphi, z) &= \frac{1}{r}\frac{\partial}{\partial\varphi}A = -C_\varphi\frac{m}{r}J_m(Kr)\sin(m\varphi)\left(C_{z,1}\sin(k_z z) + C_{z,2}\cos(k_z z)\right) \;, \\
B_\varphi(r, \varphi, z) &= -\frac{\partial}{\partial r}A = -C_\varphi K J_m^{'}(Kr)\cos(m\varphi)\left(C_{z,1}\sin(k_z z) + C_{z,2}\cos(k_z z)\right) \;, \\
E_r(r, \varphi, z) &= \mathrm{i}\frac{c^2}{\omega}\frac{\partial}{\partial z}B_\varphi = \mathrm{i}C_\varphi\frac{k_z c^2}{\omega}K J_m^{'}(Kr)\cos(m\varphi)\left(C_{z,1}\cos(k_z z) - C_{z,2}\sin(k_z z)\right) \;, \\
E_\varphi(r, \varphi, z) &= -\mathrm{i}\frac{c^2}{\omega}\frac{\partial}{\partial z}B_r = -\mathrm{i}C_\varphi\frac{k_z c^2}{\omega}\frac{m}{r}J_m(Kr)\sin(m\varphi)\left(C_{z,1}\cos(k_z z) - C_{z,2}\sin(k_z z)\right) \;, \\
E_z(r, \varphi, z) &= -\mathrm{i}\frac{c^2}{\omega}K^2 A = -\mathrm{i}C_\varphi\frac{c^2}{\omega}K^2 J_m(Kr)\cos(m\varphi)\left(C_{z,1}\sin(k_z z) + C_{z,2}\cos(k_z z)\right) \;,
\end{aligned}$$

where we introduce the derivative of the Bessel function,

$$J_m^{'} = \frac{\mathrm{d}}{\mathrm{d}(Kr)}J_m(Kr) \;.$$

To specify this intermediate solution, we also need to employ boundary conditions. We assumed that the cavity is made of perfect electric conductor material, so that the tangential electric field components vanish at the boundaries. In the radial direction, this means that E_z and E_φ must be zero at the boundary at $r = R$. Both components are proportional to J_m, so that we can deduce

$$J_m\left(Kr\right) = 0\,,\ \text{at } r = R\,.$$

From this we can determine the eigenvalue K with

$$K = \frac{j_{m,n}}{R}\,, \tag{40}$$

with $j_{m,n}$ referring to the nth zero of the Bessel function J_m.

In the longitudinal direction, the perfect electric conductor condition means that E_r and E_φ must be zero at the boundaries at $z = 0$ and $z = g$. Both components are proportional to the derivative of A_z. From this, we can deduce that:

$$\left.\frac{\mathrm{d}}{\mathrm{d}z}A_z\left(z\right)\right|_{z=0,z=g} = \left.k_z\left(C_{z,1}\cos\left(k_z z\right) - C_{z,2}\sin\left(k_z z\right)\right)\right|_{z=0,z=g} = 0\,.$$

For $z = 0$, the sine term always vanishes. To meet the boundary condition, the cosine term must also vanish, which can only be fulfilled if $C_{z,1} = 0$. The second condition thus simplifies to

$$-k_z C_{z,2}\sin\left(k_z g\right) = 0\,,$$

which is fulfilled for $k_z = p\pi/g$ and integer mode numbers p.

With these specifications the field components can be finalized as:

$$\begin{aligned}
B_r^{m,n,p}\left(r,\varphi,z\right) &= -C\frac{m}{r}J_m\left(j_{m,n}\frac{r}{R}\right)\sin\left(m\varphi\right)\cos\left(p\pi\frac{z}{g}\right)\,,\\
B_\varphi^{m,n,p}\left(r,\varphi,z\right) &= -C\frac{j_{m,n}}{R}J_m'\left(j_{m,n}\frac{r}{R}\right)\cos\left(m\varphi\right)\cos\left(p\pi\frac{z}{g}\right)\,,\\
E_r^{m,n,p}\left(r,\varphi,z\right) &= -\mathrm{i}C\frac{p\pi}{g}\frac{c^2}{\omega_{m,n,p}}\frac{j_{m,n}}{R}J_m'\left(j_{m,n}\frac{r}{R}\right)\cos\left(m\varphi\right)\sin\left(p\pi\frac{z}{g}\right)\,,\\
E_\varphi^{m,n,p}\left(r,\varphi,z\right) &= \mathrm{i}C\frac{p\pi}{g}\frac{c^2}{\omega_{m,n,p}}\frac{m}{r}J_m\left(j_{m,n}\frac{r}{R}\right)\sin\left(m\varphi\right)\sin\left(p\pi\frac{z}{g}\right)\,,\\
E_z^{m,n,p}\left(r,\varphi,z\right) &= -\mathrm{i}C\frac{c^2}{\omega_{m,n,p}}\left(\frac{j_{m,n}}{R}\right)^2 J_m\left(j_{m,n}\frac{r}{R}\right)\cos\left(m\varphi\right)\cos\left(p\pi\frac{z}{g}\right)\,,
\end{aligned} \tag{41}$$

with the integer mode numbers m, n, and p and the normalization constant $C = C_\varphi C_{z,2}$. The eigenfrequency of the modes is

$$\omega_{m,n,p} = \sqrt{\left(\frac{j_{m,n}}{R}\right)^2 + \left(\frac{p\pi}{g}\right)^2}\,. \tag{42}$$

For more information on the eigenmodes of cylindrical structures, see Ref. [6].

4.2 The loss parameters and the monopole wake function

We want to further limit our considerations to the so-called monopole modes, and the subsequent monopole wake function. In the monopole case, the azimuthal mode number is zero, $m = 0$. The resultant time-dependent field components of the TM modes are:

$$B_r^{0,n,p}\left(r,\varphi,z,t\right) = 0\,,$$

$$
\begin{aligned}
B_{\varphi}^{0,n,p}(r,\varphi,z,t) &= -C\frac{j_{0,n}}{R}J_0'\left(j_{0,n}\frac{r}{R}\right)\cos\left(p\pi\frac{z}{g}\right)\exp\left(\mathrm{i}\omega_{0,n,p}t\right)\,,\\
E_{r}^{0,n,p}(r,\varphi,z,t) &= -\mathrm{i}C\frac{p\pi}{g}\frac{c^2}{\omega_{0,n,p}}\frac{j_{0,n}}{R}J_0'\left(j_{0,n}\frac{r}{R}\right)\sin\left(p\pi\frac{z}{g}\right)\exp\left(\mathrm{i}\omega_{0,n,p}t\right)\,,\\
E_{\varphi}^{0,n,p}(r,\varphi,z,t) &= 0,\\
E_{z}^{0,n,p}(r,\varphi,z,t) &= -\mathrm{i}C\frac{c^2}{\omega_{0,n,p}}\left(\frac{j_{0,n}}{R}\right)^2 J_0\left(j_{0,n}\frac{r}{R}\right)\cos\left(p\pi\frac{z}{g}\right)\exp\left(\mathrm{i}\omega_{0,n,p}t\right)\,,
\end{aligned}
\tag{43}
$$

employing a harmonic time dependence.

To calculate the loss parameters, we first need to compute the voltage drop per mode,

$$
V_{0,n,p} = \int_0^g E_z^{0,n,p}\left(r=0,z,t=\frac{z}{c}\right)\mathrm{d}z\,.
$$

Inserting $E_z^{0,n,p}$ and evaluating the integral yields:

$$
V_{0,n,p} = Cc\left(1-(-1)^p\exp\left(\frac{\mathrm{i}\omega_{0,n,p}g}{c}\right)\right)\,. \tag{44}
$$

We also need to know about the energy stored in each mode,

$$
U_{n,p} = \frac{1}{2\mu_0}\int_0^R\int_0^{2\pi}\int_0^g r\left(B_{\varphi}^{0,n,p}\right)\left(B_{\varphi}^{0,n,p}\right)^*\mathrm{d}z\mathrm{d}\varphi\mathrm{d}r\,,
$$

which is equivalent to Eq. (16), since $\frac{\epsilon_0}{2}\int|\mathbf{E}_n(\mathbf{r})|^2\,\mathrm{d}^3r = \frac{1}{2\mu_0}\int|\mathbf{B}_n(\mathbf{r})|^2\,\mathrm{d}^3r$. We use B_φ from Eq. (43) and solve the integral to get

$$
U_{0,n,p} = C^2\frac{j_{0,n}^2\pi g}{4}J_1^2(j_{0,n})\left(1+\delta_{0,p}\right)\,. \tag{45}
$$

Here, we made use of the properties of the Bessel function to substitute $J_0'(x) = -J_1(x)$, and introduced the Kronecker symbol, $\delta_{0,p}$, i.e., $\delta_{0,p}=1$ if $p=0$ and $\delta_{0,p}=0$ otherwise.

With this, the loss parameters can be calculated via Eq. (18). For the cylindrical cavity, the loss parameters of the monopole eigenmode identified by the radial and axial mode numbers n and p are:

$$
k_{0,n,p} = \frac{1}{\pi\epsilon_0 g}\frac{2}{1+\delta_{0,p}}\frac{1-(-1)^p\cos\left(\frac{\omega_{0,n,p}g}{c}\right)}{j_{0,n}^2J_1^2(j_{0,n})}\,. \tag{46}
$$

Figure 9 demonstrates the loss parameters of eigenmodes inside the given cylindrical cavity. The modes are first distinguished by their radial mode number n and then plotted against their axial mode number p. The graph shows that the loss parameters, and therefore the contribution strength of each mode to the wake potential, strongly vary. The mode with the strongest contribution is the $\mathrm{TM}_{0,1,1}$-mode, visible in the sharp peak displayed in the graph. It can be generally assumed that if we were to increase n even further, the contributions of the modes (the so-called *higher order modes*) would further decrease.

From the loss parameters, the wake function can be calculated similarly to Eq. (19),

$$
W_{||,0}(s) = \sum_{n=1}^{\infty}\sum_{p=0}^{\infty}2k_{0,n,p}\cos\left(\frac{\omega_{0,n,p}s}{c}\right)\,. \tag{47}
$$

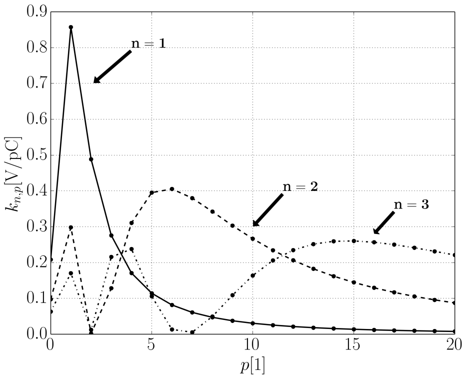


Fig. 9: Loss parameters of different monopole modes, discriminated by their radial mode number n and plotted against their axial mode number p. It is clearly visible that the strengths of the loss parameters, and thus their contributions to the wake potential, vary. The $TM_{0,1,1}$ mode shows the strongest contribution of the displayed modes.

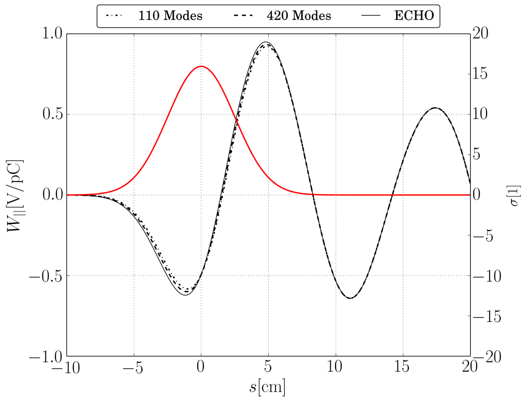


Fig. 10: Wake potential of a Gaussian pulse with $\sigma = 2.5\,\mathrm{cm}$ inside the exemplary pillbox cavity. The dash-dotted line denotes the analytical result for 110 considered modes; the dashed line for 420 modes. The solid line represents a numerical approximation of the wake potential computed with the software ECHO. All wake potentials are in very good agreement. The bunch shape function is plotted in red as a reference.

We now want to compute the wake potential of a Gaussian pulse. The bunch shape function of such a pulse is:

$$\psi(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s - s_0}{2\sigma^2}\right) . \tag{48}$$

Here, we assume that the Gaussian is centred on $s_0 = 0$, and that the width is $\sigma = 2.5$ cm. We will test the accuracy of the expansion for two different numbers of expansion functions, 110 (i.e., $1 \leq n \leq 10$ and $0 \leq p \leq 10$) and 420 modes (i.e., $1 \leq n \leq 20$ and $0 \leq p \leq 20$). The analytical results are compared with a numerical result obtained using the software ECHO [7]; the comparison is shown in Fig. 10. The results are in very good agreement; the difference in the accuracy of the analytical wake potentials is very small. From this, we can conclude that the influence of the higher-order modes on the wake potential is, indeed, comparatively small in this case, though still necessary to increase the accuracy of the analytical approximation of the wake potential.

5 The effects of wakefields

5.1 Ultra-relativistic wakefields

As described before, wakefields that remain in cavities can have large effects on trailing particles and bunches. Effectively, they represent *energy modulations* of the trailing particles—which can already hold true for particles of the same bunch as the wakefield-generating particle. Generally, this phenomenon is hardly predictable in complicated structures. These energy modulations can lead to an increase the emittance of the particle bunch. Usually, this is unwanted since—as soon as the beam has reached its design parameters—it might lead to beam instabilities if no measures are undertaken.

Nevertheless, there are certain cases in which the energy modulating effect of wake potentials can be useful. Devices called *wakefield dechirpers* or *wakefield silencers* are simple, passive accelerator components that are used to counteract the energy spreads of particle beams. Figure 11 compares the effect of a wakefield on a beam with a strong energy spread with that on a beam with no energy spread. In both cases, the modulation of the dechirper is represented by a linear energy gradient over the coordinate s (red line). One beam has a low initial energy spread (dashed line), and so, adding the modulation of the wake potential, the resulting energy width after the wakefield is larger. The second beam, however, has a strong initial energy spread opposing the modulation of the dechirper (solid line). Here, applying the wake potential will effectively *lower* the total energy spread. This effect is currently studied at accelerators all over the world. Section 6 describes the wake potentials in a suitable structure in more detail.

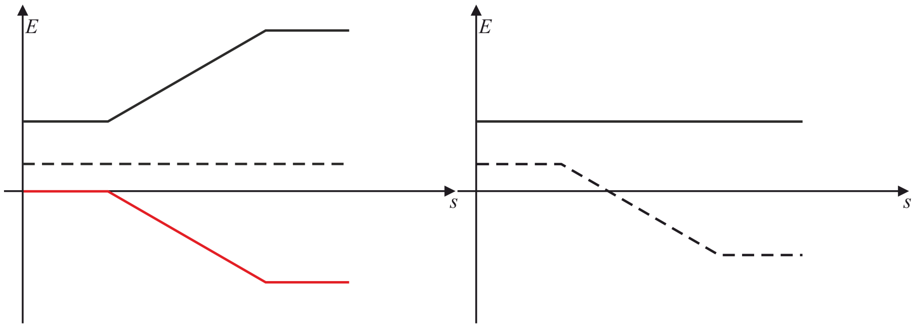


Fig. 11: The energy modulation induced by a wakefield (red) acts on two different beams, one with initial energy spread that is opposed to the modulation (black, solid line), one without (black, dashed line). Effectively, the modulation is added to the phase space of the beams, so that, after the dechirper, the energy width of the initially unchirped beam is *increased*, while the energy width of the chirped beam is *reduced*, owing to the interaction with the wake potential.

5.2 Space charge wakefields

So far, we have only dealt with ultra-relativistic wakefields, which are limited to the case $v \simeq c$. For $v < c$, space charge effects also play a role.

The reason for space charge effects are the Coulomb interactions of the charged particles, which play a role for non-ultra-relativistic beams and need to be taken into account there for every type of simulation or analytical consideration. Usually, these effects have an even larger influence on the trailing particles than the wakefields themselves. The effects of space charges can include the deflection of charged particles, the causing of beam instabilities, the generation of high (unwanted) field intensities, which in turn can lead to material breakdown, etc.

However, space charge effects are not the subject of these considerations. Thus, the reader is referred to Refs. [8] or [9] for more information.

6 Example 2: rectangular waveguide with dielectric linings

6.1 The eigenmodes

Figure 12 shows a rectangular waveguide with dielectric linings, which can be used as a wakefield dechirper ([10], [11]). The outer waveguide is made of a highly conductive material, e.g., copper or aluminium. As discussed before, the phase velocity of electromagnetic waves is smaller in the dielectric regions, which effectively slows down the image fields responsible for wake potentials, so that they can act on trailing particles as well. For this section, we will assume an exemplary waveguide with the parameters $a = 5\,\mathrm{cm}$, $b = 1.3\,\mathrm{cm}$, $L = 30\,\mathrm{cm}$; and dielectric coatings of a thickness $b - d = 1.5\,\mathrm{mm}$ and a relative permittivity, $\epsilon_\mathrm{r} = 4.8$.

As for the pillbox cavity considered before, the wake function is now analysed. First, it is required to take a look at the eigenmodes of these structures. We will consider three-dimensional eigenmodes here. In reality, this would be equivalent to the examination of a structure that is closed in all three dimensions. This is of course unrealistic—for the dechirper to work, it must be passed by a particle beam, and consequently, the structure needs to be *open* in the z-direction. In this special case, however, owing to the properties of the structure, the wake function resulting from the three-dimensional eigenmodes of the close structure is identical to the wake function of the open structure.

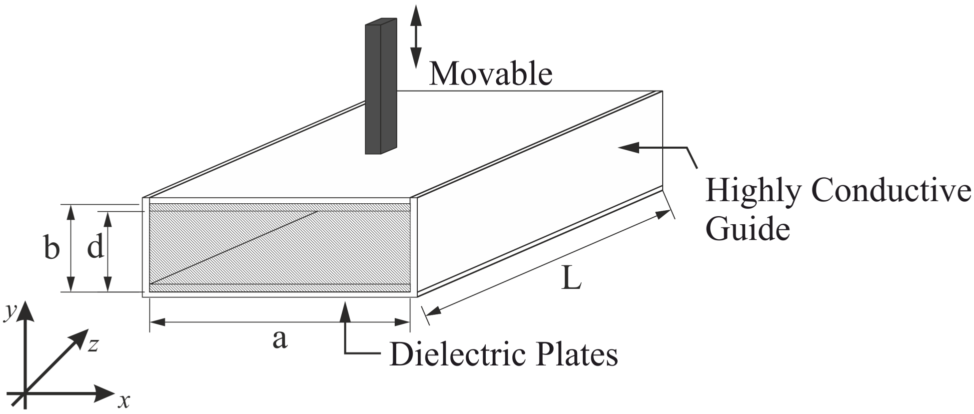


Fig. 12: A rectangular waveguide lined with two dielectric plates (shaded), which can be used as a dechirper. The outer waveguide is made of a highly conductive material. One or both plates (only the upper plate in the case shown) can be left unconnected to the remaining waveguide, so that the distance between the dielectrics can be adjusted.

The similarities between this dielectrically lined waveguide and an empty rectangular waveguide without dielectrics become obvious when the structure is considered for the first time. It would make sense if the eigenmodes reflect these similarities, i.e., if the eigenmodes are similar to TE and TM modes, which are the eigenmodes of conventional empty rectangular waveguides. Indeed, the eigenmodes of the lined waveguide are related to TE and TM modes and can be viewed as a superposition of them. This new set of eigenmodes, however, needs to reflect the dielectrics, which can be effectively described as a change in permittivity in one direction (the y-direction in this case). As a consequence, none of the longitudinal components of the eigenmodes' electric or magnetic fields is zero, as in the case of TE and TM modes but one of the *transverse* components is (the y-component, in this case). These kinds of mode are called *longitudinal-section electric (LSE)* modes (where $E_y = 0$) and *longitudinal-section magnetic (LSM)* modes (where $B_y = 0$) [6].

Their electric and magnetic fields can be derived from Maxwell's equations (assuming a harmonic time dependence and $\rho = 0, \mathbf{J} = \mathbf{0}$). However, the changing relative permittivity ϵ_r must be taken into account, so that:

$$\nabla \cdot \mathbf{D}\left(\mathbf{r}\right) = \epsilon_0 \nabla \cdot \left(\epsilon_\mathrm{r}(y)\mathbf{E}\left(\mathbf{r}\right)\right) = 0\,, \tag{49}$$

$$\nabla \cdot \mathbf{B}\left(\mathbf{r}\right) = 0\,, \tag{50}$$

$$\nabla \times \mathbf{E}\left(\mathbf{r}\right) = -\mathrm{i}\omega \mathbf{B}\left(\mathbf{r}\right)\,, \tag{51}$$

$$\nabla \times \mathbf{B}\left(\mathbf{r}\right) = \mathrm{i}\frac{\omega}{c_0^2}\epsilon_\mathrm{r}(y)\mathbf{E}\left(\mathbf{r}\right)\,. \tag{52}$$

Under the given circumstances, solution of these equations is analytically possible. In the x- and z-directions, the solution procedure follows a similar line as the derivation of the eigenmodes in an empty waveguide, since the phase velocity is the same everywhere in these directions. In the y- direction, the changing permittivity makes a straightforward solution like this impossible, though. To obtain an analytical expression for the eigenmodes, a Fourier expansion can be used to describe the unknown behaviour in the y- direction. Reference [6] provides a detailed example of the solution procedure for the case of only one dielectric lining. The resulting electric fields (for two dielectric slabs) are

$$\mathbf{E}_{\mathrm{LSE}}\left(\mathbf{r}\right) = \omega_{\mathrm{LSE}} \begin{pmatrix} k_z \cos(k_x x) \sum\limits_{m=1}^{N} b_m \sin(k_{ym} y) \sin(k_z z) \\ 0 \\ -k_x \sin(k_x x) \sum\limits_{m=1}^{N} b_m \sin(k_{ym} y) \cos(k_z z) \end{pmatrix} ,$$

for LSE modes and

$$\mathbf{E}_{\mathrm{LSE}}\left(\mathbf{r}\right) = \frac{1}{\varepsilon_r(y)} \begin{pmatrix} -k_x \cos(k_x x) \sum\limits_{m=0}^{N} b_m k_{ym} \sin(k_{ym} y) \sin(k_z z) \\ \left(k_x^2 + k_z^2\right) \sin(k_x x) \sum\limits_{m=0}^{N} b_m \cos(k_{ym} y) \sin(k_z z) \\ -k_z \sin(k_x x) \sum\limits_{m=0}^{N} b_m k_{ym} \sin(k_{ym} y) \cos(k_z z) \end{pmatrix} ,$$

for LSM modes. Here, k_x, k_{ym}, and k_z represent the eigenvalues in the x-, y-, and z- directions. Just as in the case of the empty rectangular waveguide, all eigenvalues indicate the number of nodes or antinodes of the sine and cosine functions in the fields, e.g., $k_x = n\pi/a$, with a being the structure's width. Consequently, $k_{ym} = m\pi/b$ and $k_z = l\pi/L$, with b being the structure's height and L the structure's length. The index m *also* indicates the Fourier expansion; the expansion coefficients are q_m. Note that the summation is ceased after N terms, which makes the analytical expression an approximation of the real result. N is usually determined in a convergence study.

6.2 The electric field and longitudinal wake potential

As a next step, we want to determine the longitudinal wake potential inside a dielectrically lined rectangular waveguide. To achieve this, we first compute the wake function, and subsequently the loss parameters.

Until this point, we have determined the loss parameters using a relation between the stored energy and voltage drop per mode. This time, we want to calculate the wake function by a straightforward integration over the electric field. We still expand the electric field into a series of eigenmodes as shown in Eq. (15). We carry out this summation over both LSE and LSM modes. The time-dependent expansion coefficients are determined by solving Maxwell's equations for a point charge moving along the beam axis in the z-direction:

$$\nabla \cdot \mathbf{D}\left(\mathbf{r}\right) = \varepsilon_0 \nabla \cdot \left(\varepsilon_r(y) \mathbf{E}\left(\mathbf{r}\right)\right) = \rho\left(\mathbf{r}\right) \ , \tag{53}$$

$$\nabla \cdot \mathbf{B}\left(\mathbf{r}\right) = 0 \, , \tag{54}$$

$$\nabla \times \mathbf{E}\left(\mathbf{r}\right) = -\frac{\partial}{\partial t} \mathbf{B}\left(\mathbf{r}\right) \ , \tag{55}$$

$$\nabla \times \mathbf{B}\left(\mathbf{r}\right) = \mu_0 \mathbf{j}\left(\mathbf{r}\right) + \mu_0 \varepsilon_r(y) \varepsilon_0 \frac{\partial}{\partial t} \mathbf{E}\left(\mathbf{r}\right) \ . \tag{56}$$

The beam axis goes straight through the centre of the waveguide, so through the point $(x_{\mathrm{beam}}, y_{\mathrm{beam}}) = \ (a/2, b/2)$.

Once the coefficients have been determined, the longitudinal component of the electric field on the beam axis is integrated following Eq. (5). The process of the integration is tedious but analytically possible. We want to skip it here; Ref. [12] provides a more detailed solution of the wake function integral.

Within the scope of this paper, it is sufficient to discuss the final result of the integration. The integration automatically leads to a description of the wake function similar to Eq. (19). The cosine dependence is a direct result of the integration; the summation is a remnant of the eigenmode expansion of the electric field. From this direct integration, the loss parameters can be read out as:

$$
\begin{aligned}
K_{n,m,l,\mathrm{LSE}} &= -\frac{4}{\varepsilon_0 a L}\frac{k_x^2 k_z^2}{k_x^2+k_z^2}\left(\sum_{\text{even } m} q_m \sin\left(\frac{m\pi}{2}\right)\right)^2 \frac{\left(2-2\mathrm{e}^{\mathrm{i}l\pi}\cos\left(k_{0,\lambda}L\right)\right)}{\left(k_{0,\lambda}^2-k_z^2\right)^2}\,, \\
K_{n,m,l,\mathrm{LSM}} &= -\frac{4}{\varepsilon_0 a L}\frac{k_z^4}{k_{0,\mu}\left(k_x^2+k_z^2\right)}\left(\sum_{\text{even} m} q_m k_{y,m} \sin\left(\frac{m\pi}{2}\right)\right)^2 \frac{\left(2-2\mathrm{e}^{\mathrm{i}l\pi}\cos\left(k_{0,\mu}L\right)\right)}{\left(k_{0,\mu}^2-k_z^2\right)^2}\,.
\end{aligned}
$$

To avoid confusing the loss parameters with the eigenvalues, we refer to them as $K_{n,m,l}$ in this case. The expressions of the loss parameters *appear* very lengthy and complicated; on a closer inspection, however, we see that they only depend on the mode characteristics and geometrical properties of the structure and thus can be determined without previous knowledge of the electric field. The electric field expansion is merely an intermediate analytical step to determine the formula for the loss parameters *once*. After this is done, it does not have to be repeated for every structure; rather the expressions for the loss parameters can be used straight from the formula at hand.

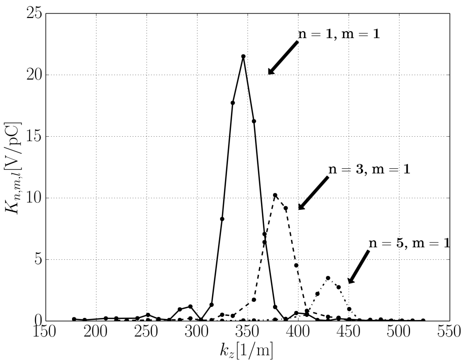


Fig. 13: Loss parameters of LSM modes with different n and m plotted over their longitudinal eigenvalue k_z. It can be seen that several modes exhibit much higher loss parameters than others, resulting in the observed peak structure.

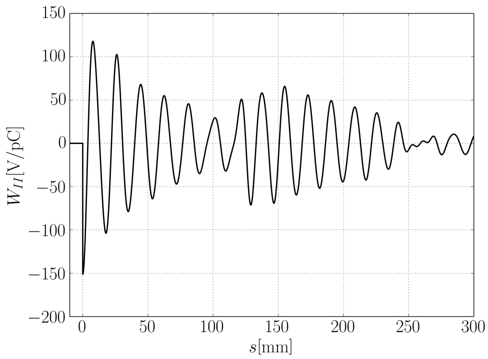


Fig. 14: Wake function in the exemplary dielectrically lined rectangular waveguide.

In Fig. 13, the loss parameters (of LSM modes of an arbitrary rectangular waveguide with an arbitrary dielectric lining) are again sorted (according to the numbers of nodes or antinodes in the transverse direction, n and m) and plotted against their longitudinal eigenvalue. As in the exemplary cylindrical cavity in Fig. 9, we observe several distinguished peaks. These eigenmodes obviously have the largest contribution to the wake function, while the loss parameters of other modes can be so small that their contributions to the wake function are negligible.

Following Eq. (19), the wake function can be calculated from the loss parameters. The wake function resulting from the modes displayed in Fig. 13 is displayed in Fig. 14.

At this point, we want to have a closer look at the influence of the bunch shape function on the resulting wake potential. According to Eq. (20), all that is needed to obtain the wake potential from the wake function is a convolution with the bunch shape function. This can be carried out easily, either numerically or analytically.

Figure 15 compares the short range wake potentials (i.e., the wake potentials in the vicinity of the bunch) of a Gaussian bunch (cf. Eq. (48)) and a so-called flat top pulse, i.e., a pulse with a constant

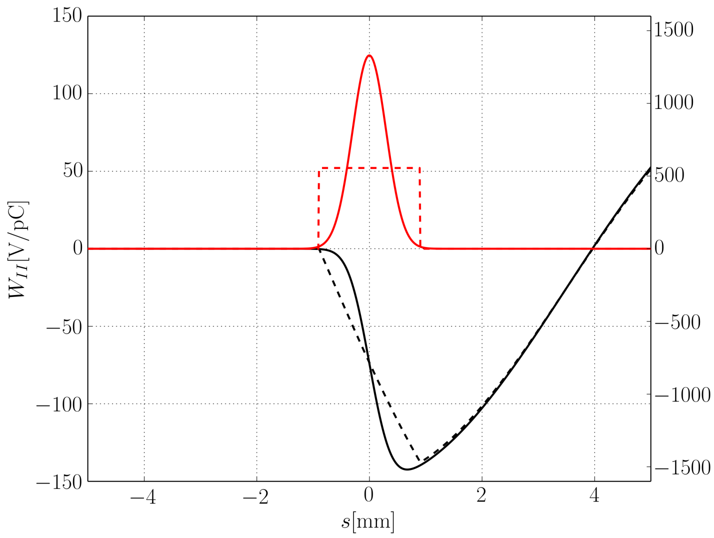


Fig. 15: Wake potentials of a Gaussian pulse (solid line, $\sigma = 0.3\,\mathrm{mm}$) and a flat top pulse (dashed line, pulse length $1.8\,\mathrm{mm}$) inside the exemplary rectangular waveguide with dielectric linings. The bunch shape functions are displayed in red for comparison.

particle distribution over the bunch length. Both bunches are designed to have the same length and both bunch shape functions are normalized to 1. It can be seen that the wake potentials over both bunch shapes refer to an overall energy loss. The maximum loss over the bunch shape is similar in both cases. This results from the normalization of the pulses: the convolution can generally be imagined as 'moving' the bunch shape function over the wake function and measuring the area enclosed by both. If the area under both used bunches is the same, owing to the normalization, it follows that the maximum energy loss should be nearly equal for both bunches. The gradient of the energy loss, however, shows significant differences when comparing both bunch shapes. The gradient for the flat top pulse is perfectly linear with sharp edges, while the gradient for the Gaussian pulse is steeper in the middle section and shows softened edges. This is a direct result of the different bunch shapes: the gradient for the flat top pulse is linear because of the equal distribution of particles over the bunch length and the gradient for the Gaussian pulse shows softened edges, owing to the Gaussian pulse's smooth behaviour.

For the total *energy loss* due to the wake potential, the bunch shape thus plays only a minor role, i.e., Fig. 15 clearly shows that the maximum energy loss over the particle bunch is nearly identical, and only dependent on the normalization of the bunch shape function (which should be 1 in any case). For the *phase space* of the bunch, however, the bunch shape function can have a significant influence. Here, it plays a role in determining 'how many' particles are subjected to a certain energy reduction. The dechirper can significantly alter the phase space of the bunch, depending on its bunch shape.

7 Summary and conclusions

In this paper, we introduced the basic quantities that are necessary to understand the concept of wakefields. We discussed the basic structural requirements for the generation of wakefields. We derived the longitudinal and transverse wake potential from the Lorentz force and introduced the Panofsky–Wenzel theorem, which links both quantities to each other. In addition, we briefly discussed the impedance as the Fourier transform of the wake potential.

When the eigenmodes of a structure are known, it is possible to derive the wake function, the wakefield of a point charge as a sum over each eigenmode's contribution. These contributions are called loss factors. The wake function then serves as the Green's function for the calculation of the wake potential of an arbitrary bunch shape, and the total loss factor can give an insight into the total power loss due to the wakefield.

The effects of wakefields represent energy modulations, which consequently result in a modulation of the longitudinal and possibly also transverse phase space of a particle bunch. This is why wakefields, at least most of the time, are considered unwanted effects that have to be taken into consideration during the design process of accelerators, to mitigate their negative influence on the functionality of the accelerator. However, special structures called 'wakefield dechirper' can be used to utilize the energy modulating effect of wakefields to reduce the energy spread of particle beams.

Further reading

As complementary further reading, the following article for beginners on the topic might be helpful: P. Tenenbaum, *Fields in Waveguides—A Guide for Pedestrians*, 2003, available at `http://www.desy.de/~njwalker/uspas/coursemat/notes/unit_2_notes.pdf`.

References

[1] CST, *STUDIO SUITE®*, CST AG, Darmstadt, Germany.

[2] T. Weiland and R. Wanzenberg, *Wake Fields and Impedances* 100 (Springer, Berlin, 1992) p. 39. https://doi.org/10.1007/3-540-55250-226

[3] U. van Rienen, *Numerical Methods in Computational Electrodynamics: Linear Systems in Practical Applications* 12 (Springer, Berlin, 2001). https://doi.org/10.1007/978-3-642-56802-2

[4] P.B. Wilson, Introduction to wakefields and wake potentials, SLAC-PUB 4547 (1989).

[5] T.P. Wangler, *RF Linear Accelerators* (Wiley, Weinheim, 2008). https://doi.org/10.1002/9783527623426

[6] R.E. Collin, *Field Theory of Guided Waves* (IEEE Press, New York, 1991).

[7] I. Zagorodnov and T. Weiland, TE/TM field solver for particle beam simulations without numerical Cherenkov radiation, PRST-AB 8 (2005).

[8] A. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).

[9] D.A. Edwards and M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (Wiley, New York, 1993). https://doi.org/10.1002/9783527617272

[10] F. Reimann et al., *A dielectrically lined rectangular waveguide as a wakefield dechirper for ELBE*, In G. Addamo et al. *Proc. of ICEAA 2015*, Torino, Italy, 2015, pp. 827, `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7297229`.

[11] S. Antipov et al., *Passive Momentum Spread Compensation by a 'Wakefield Silencer'*, In J. Corbett et al. *Proc. of IPAC 2012*, New Orleans, USA, 2012, pp.598, `http://accelconf.web.cern.ch/accelconf/ipac2012/papers/moppp013.pdf` .

[12] K.L.F. Bane *et al.*, Wakefields and wakefield acceleration, SLAC-Pub. 3528 (1984).

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Transverse Beam Dynamics

B.J. Holzer
CERN, Geneva, Switzerland

Abstract
This paper gives an overview of the transverse dynamics in particle accelerators. The main emphasis is on giving an introduction to the basic concepts, described in linear approximation, and allowing the reader to deduce the main parameters of a machine, based on some simple scaling laws.

Keywords
Accelerator physics; transverse dynamics.

1 Introduction

We would like to start this little essay with some kind of definition of 'what we are talking about' when we mention transverse beam dynamics. While the middle term, *beam*, might be clear enough, *transverse* and *dynamics* deserve some clarification. So in a desperate attempt to summarize in a few lines the key issues of this paper, we formulate the following questions. What do the particles do, while they are travelling along a linac or while we accelerate them in a circular machine? How do we manage to keep them on – or close to – a trajectory in a linear accelerator (even a linear collider in some cases) with deviations from this ideal path of only a fraction of a millimetre? How do we manage to convince our particle ensemble to travel distances that, in a storage ring, turn after turn, easily sum up to many millions of kilometres, without being lost?

Before we start to answer these questions, we would like to emphasize that we have to act with caution. We will use a language that has been developed for 'periodic structures', i.e. where the situation seen by the particle repeats itself after a certain time or distance. The reason lies at the bottom of the mathematics involved. Now, while this is easily fulfilled in a circular accelerator, namely in the case of a synchrotron, a linear accelerator does not have such a built-in periodicity: the particles pass through the machine only once and that's it. However, we like to make use of the expressions derived for the periodic machines, as they present a powerful and elegant tool to design the accelerator as well as to express the most relevant beam parameters. But in doing so we have to exercise due care. In the following sections, we will therefore describe this language and – wherever needed – make a clear point when we have to be careful and make a distinction between linacs and circular machines.

For the time being, let us state that an accelerator usually needs:

- a system of magnetic fields that create focusing forces to keep the particles together, and that ultimately lead to a well-defined beam size;
- in the case of a circular machine, magnetic bending fields to keep the particles on a closed, more or less circular orbit;
- a mechanism to lock these B-fields to the changing particle energy and thus keep the particles on, or close to, this design orbit over the complete energy range of the machine;
- if the particles are successfully kept in both transverse planes, we need a radio frequency (RF) structure to accelerate the particles and create the necessary energy gain via longitudinal electric fields.

By definition of the title, we will neglect the last item in this paper. Here we just assume that our colleagues from the RF systems will do a good job. Those who are interested in the longitudinal

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.145

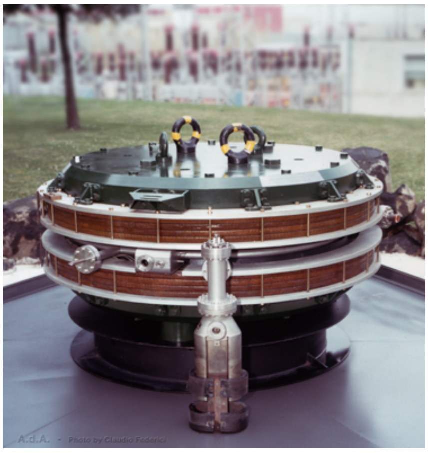


Fig. 1: Annelli de Accumulatione; the first electron–positron collider ring

Fig. 2: The tunnel of the LHC proton–proton collider at CERN, Geneva

dynamics are cordially invited to have a look at two brilliant papers, [1], [2]. For the time being we can concentrate on the issues related with the focusing properties of our accelerator.

So much for the definition. As the basic tools and so the language were developed for circular machines – synchrotrons in most cases – we will follow for a moment this concept, and further along the line we will include in our contemplations linear accelerators and transfer lines.

Two examples of synchrotrons to start with: the Annelli de Accumulatione (Fig. 1), as far as we know, the very first particle collider and certainly one of the smallest synchrotrons, built in Frascati by Bruno Touschek in 1944 [3]; and the large hadron collider (LHC) [4], at present the largest storage ring ever built, running at the highest achievable particle energies at CERN (Figs. 2 and 3).

2 Transverse beam dynamics

The transverse beam dynamics of charged particles in an accelerator describes the movement of single particles under the influence of the external transverse bending and focusing fields. It includes the detailed arrangement (for example, their positions in the machine and their strength) of the accelerator magnets used to obtain well-defined, predictable parameters of the stored particle beam, and it describes methods to optimize the trajectories of single particles, as well as the dimensions of the beam, considered as an ensemble of many particles. A treatment of this field in full mathematical detail, including sophisticated lattice optimizations, such as the right choice of the basic lattice cells and the design of dispersion

Fig. 3: The LHC proton–proton collider

suppressors or chromaticity compensation schemes, is beyond of the scope of this overview. For further reading and for more detailed descriptions, we therefore refer to the more complete explanations in Refs. [5–7]. For the time being, we will just give a basic introduction into the topic and explain – more or less hand-waving – how the trick goes.

2.1 Geometry of the ring

In general, magnetic fields are used in circular accelerators to provide the bending force and to focus the particle beam. In principle, the use of electrostatic fields would also be possible, but at high momenta (i.e., if the particle velocity is close to the speed of light), magnetic fields are much more efficient. The force acting on the particles, the Lorentz force, is given by

$$\mathbf{F} = q \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B})\,. \tag{1}$$

For high-energy particle beams, the velocity $\mathbf{v}$ is close to the speed of light and so represents a nice amplification factor whenever we apply a magnetic field. As a consequence, it is much more convenient to use magnetic fields for bending and focusing the particles.

Therefore, neglecting electric fields for the moment, we write the Lorentz force and the centrifugal force on the particle on its circular path as

$$F_{\mathrm{Lorentz}} = e \cdot v \cdot B\,, \tag{2}$$

$$F_{\mathrm{centrifugal}} = \frac{\gamma m_0 v^2}{\rho}\,. \tag{3}$$

Assuming an idealized homogeneous dipole magnet along the particle orbit, having pure vertical field lines, we define the condition for a perfect circular orbit as equality between these two forces. This yields the following condition for the idealized ring:

$$\frac{p}{e} = B \cdot \rho\,, \tag{4}$$

where we are referring to protons and have accordingly set $q = e$. This condition relates the so-called beam rigidity $B\rho$ to the momentum of a particle that can be carried in the storage ring, and it ultimately defines, for a given magnetic field of the dipole magnets, the size of the storage ring.

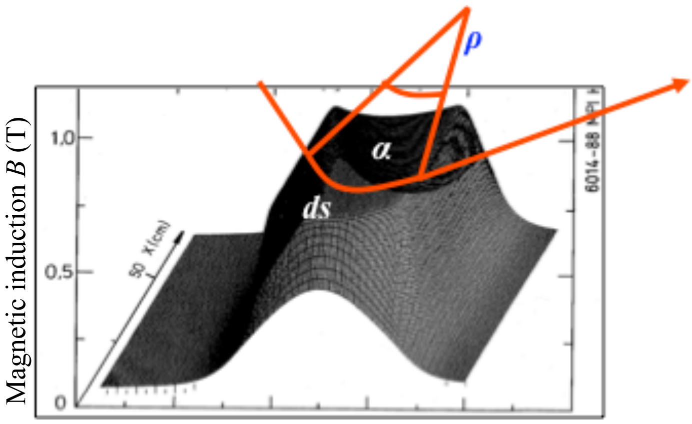


Fig. 4: Field map of a storage ring dipole magnet, and schematic path of a particle

In reality, instead of having a continuous dipole field, the storage ring will be built with several dipole magnets, powered in series to define the geometry of the ring. For a single magnet, the trajectory of a particle is shown schematically in Fig. 4. In the free space outside the dipole magnet, the particle trajectory follows a straight line. As soon as the particle enters the magnet, it is bent onto a circular path until it leaves the magnet at the other side.

The overall effect of the main bending (or 'dipole') magnets in the ring is to define this more or less circular path, which we call the 'design orbit'. By definition, this design orbit has to be a closed loop, and so the main dipole magnets in the ring have to define a full bending angle of exactly 2π. If α denotes the bending angle of a single magnet, then

$$\alpha = \frac{\mathrm{d}s}{\rho} = \frac{B\,\mathrm{d}s}{B\cdot\rho}\,. \tag{5}$$

We therefore require that integrating over all dipole magnets we get

$$\frac{\int B\,\mathrm{d}s}{B\cdot\rho} = 2\pi\,. \tag{6}$$

Thus, a storage ring or synchrotron is not a 'ring' in the true sense of the word but more a polygon, where 'poly' means the discrete number of dipole magnets installed in the 'ring'.

In the case of the LHC, the dipole field has been pushed to the highest achievable values: 1232 superconducting dipole magnets, each 15 m long, define the geometry of the ring (or better 1232-gon, whatever the Greek expression for that might be) and thus, via Eq. (6), the maximum momentum for the stored proton beam. Using these equations, for a maximum momentum $p = 7$ TeV/c, we obtain a required magnetic field of

$$B = \frac{2\pi\cdot 7000\cdot 10^9\ \mathrm{eV}}{1232\cdot 14.3\ \mathrm{m}\cdot 2.99792\cdot 10^8\ \mathrm{m\ s^{-1}}}\,, \tag{7}$$

or

$$B = 8.33\ \mathrm{T}, \tag{8}$$

to bend the LHC beams. For convenience, we have expressed the particle momentum in units of GeV/c here. Figure 5 shows a photograph of one of the LHC dipole magnets, built with superconducting NbTi filaments, which are operated at a temperature $T = 1.9$ K.

Fig. 5: Superconducting dipole magnet in the LHC storage ring

2.2 Focusing properties

In addition to the main bending magnets that guide the beam onto a closed orbit, focusing fields are needed to keep the particles close together. In modern storage rings and light sources, we have to keep more than 10^{12} particles in the machine, distributed over a number of bunches, and these particles have to be focused to keep their trajectories close to the design orbit. Furthermore, these particles are stored in the machine for many hours, and a carefully designed focusing structure is needed to maintain the necessary beam size at different locations in the ring and to guarantee stability of the transverse motion.

Following classical mechanics, linear restoring forces are used, just as in the case of a harmonic pendulum. Quadrupole magnets provide the corresponding field property: they create a magnetic field that depends linearly on the amplitude of the particle, i.e., the distance of the particle from the design orbit:

$$B_x = -g \cdot y\,, \qquad B_y = -g \cdot x\,. \tag{9}$$

The constant g is called the gradient of the magnetic field and characterizes the focusing strength of the quadrupole lens in both transverse planes. The minus sign is a convention that follows the fact that for a positive amplitude, the field configuration of a focusing quadrupole will lead to a Lorentz force that reduces this amplitude, according to Fig. 6. As in the case of the dipole field, the quadrupole gradient is usually normalized to the particle momentum to obtain expressions that are valid for any particle momentum or energy. This normalized gradient is denoted by k and defined as

$$k = \frac{g}{p/e} = \frac{g}{B\rho}\,. \tag{10}$$

The technical layout of such a quadrupole is depicted in Fig. 7. As in the case of the dipoles, the LHC quadrupole magnets were built using superconducting technology to achieve the highest possible focusing forces.

Now that we have defined the two basic building blocks of a storage ring, we need to arrange them in a so-called magnet lattice and optimize the field strengths in such a way as to obtain the required beam parameters. An example of such a magnet lattice is shown in Fig. 8. This photograph shows the dipole (orange) and quadrupole (red) magnets in the TSR storage ring in Heidelberg [8]. Eight dipoles are used to bend the beam into a 'circle', and the quadrupole lenses between them provide the focusing to keep the particles within the aperture limits of the vacuum chamber.

A general design principle of modern synchrotrons and storage rings should be pointed out here. In general, these machines are built following a so-called separate-function scheme: every magnet is

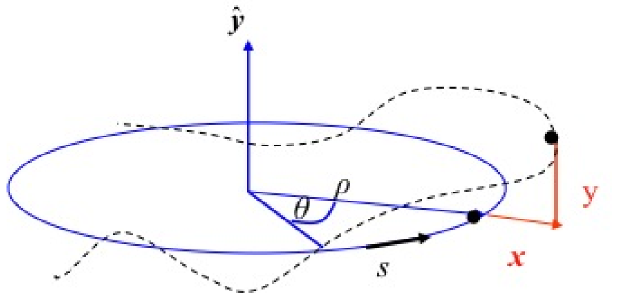


Fig. 6: Co-ordinate system used in particle beam dynamics: the longitudinal co-ordinate s moves around the ring with the particle considered.

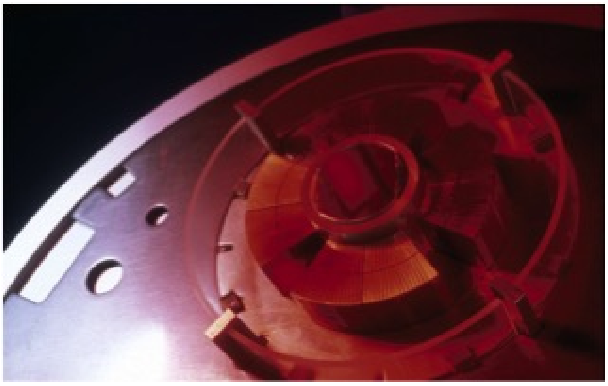


Fig. 7: Superconducting quadrupole magnet in the LHC storage ring

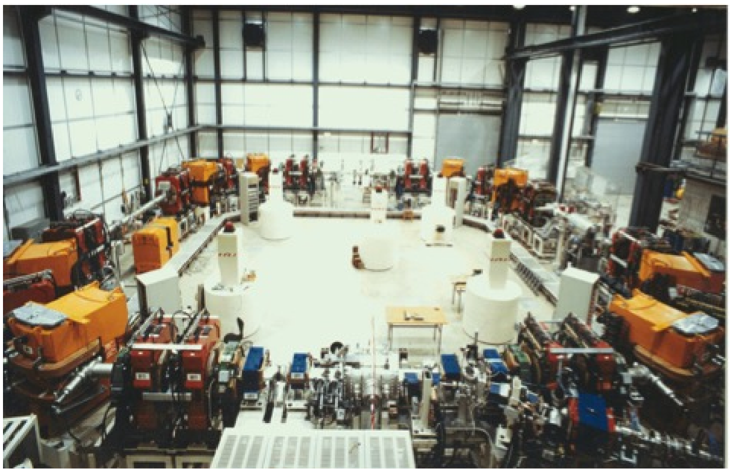


Fig. 8: The TSR storage ring, Heidelberg, is a typical example of a separate-function strong focusing storage ring [8].

designed and optimized for a certain task, such as bending, focusing, or chromatic correction. We separate the magnets in the design according to the job they are supposed to do; only in rare cases is a combined-function scheme chosen nowadays, where different magnet properties are combined in one piece of hardware. To express this principle mathematically, we use the general Taylor expansion of the normalized magnetic field,

$$\frac{B(x)}{p/e} = \frac{1}{\rho} + k \cdot x + \frac{1}{2!} m x^2 + \frac{1}{3!} n x^3 + \cdots . \tag{11}$$

Following these arguments, for the moment we take only constant (dipole-) or linear (quadrupole-) terms into account. The higher-order contributions to the field will be treated later as (hopefully) small perturbations.

Under these assumptions, we can derive – in linear approximation – the equation of motion of the transverse particle movement. We start with a general expression for the radial acceleration, known from classical mechanics (see, e.g., Ref. [9]):

$$a_r = \frac{\mathrm{d}^2 \rho}{\mathrm{d}t^2} - \rho \left(\frac{\mathrm{d}\theta}{\mathrm{d}t} \right)^2 . \tag{12}$$

The first term refers to an explicit change in the bending radius, and the second to the centrifugal acceleration. Referring to our co-ordinate system, and replacing the ideal radius ρ with $\rho + x$ for the general case (Fig. 6), we obtain the relation for the balance between the radial force and the counteracting Lorentz force:

$$F = m \frac{\mathrm{d}^2}{\mathrm{d}t^2}(x + \rho) - \frac{m v^2}{x + \rho} = evB . \tag{13}$$

On the right-hand side of the equation, we take only linear terms of the magnetic field into account,

$$B_y = B_0 + x \frac{\mathrm{d}B_y}{\mathrm{d}x} , \tag{14}$$

and for convenience we replace the independent variable t with the co-ordinate s,

$$x' = \frac{\mathrm{d}x}{\mathrm{d}s} = \frac{\mathrm{d}x}{\mathrm{d}t} \frac{\mathrm{d}t}{\mathrm{d}s} , \tag{15}$$

'Convenience' in this context means that we are more interested in the amplitude x and angle x' of the particle trajectory and therefore prefer the derivative of x with respect to s. Thus, we obtain an expression for the particle trajectories under the influence of the focusing properties of the quadrupole and dipole fields in the ring, described by a differential equation. This equation is derived in its full beauty elsewhere [7], so we shall just state it here:

$$x'' - x \cdot \left(k - \frac{1}{\rho^2} \right) = 0 , \tag{16}$$

where k is the normalized gradient introduced above and the $1/\rho^2$ term represents the so-called weak focusing, which is a property of the bending magnets. Depending on the actual sign of k, the quadrupole will focus (negative sign) or de-focus (positive sign) the beam in the corresponding plane. The situation is shown schematically in Fig. 6. An ideal particle will follow the design orbit represented by the circle in the diagram. Any other particle will perform transverse oscillations under the influence of the external focusing fields, and the amplitude of these oscillations will ultimately define the beam size. To be brief, and referring to the horizontal plane for a moment, we can make the statement that under the influence of the focusing fields from the quadrupoles k and dipoles $1/\rho^2$, the transverse movement of the particles inside the single lattice elements looks like a harmonic oscillation.

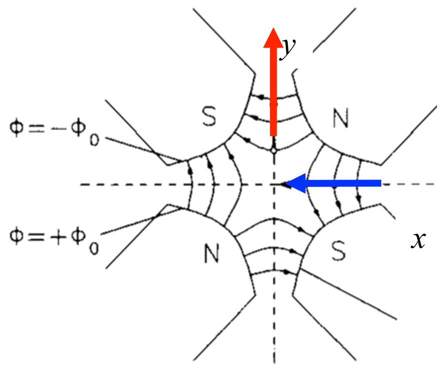


Fig. 9: Field configuration in a quadrupole magnet and the direction of the focusing and defocusing forces in the horizontal and vertical planes.

Unlike the case of a classical harmonic oscillator, however, the equations of motion in the horizontal and vertical planes differ somewhat. Assuming a horizontal focusing magnet, the equation of motion is as shown in Eq. (16). In the vertical plane, however, because of the orientation of the field lines and thus – in the end – by Maxwell's equations, the forces instead have a defocusing effect. Also, the weak focusing term disappears in general:

$$y'' + y \cdot k = 0\,. \tag{17}$$

The principal problem arising from the different directions of the Lorentz force in the two transverse planes of a quadrupole field is sketched in Fig. 9. As a consequence, to overcome this uncomfortable situation, we have to explicitly introduce quadrupole lenses that focus the beam in the horizontal and vertical directions in some alternating order. It is the task of the machine designer to find an adequate solution to this problem and to define a magnet pattern that will provide an overall focusing effect in both transverse planes. The rest is easy, in the sense of A. Wolski's statement: "...in principle, there are only two steps in the analysis of any dynamical system. The first step is to write down the equations of motion; and the second step is to solve them " [10].

Now, closely following the example of the classical harmonic oscillator, we can write down the solutions of the equations of motion. For simplicity, we focus on the horizontal plane; a 'focusing' magnet is therefore focusing in this horizontal plane and at the same time defocusing in the vertical plane. Starting with the initial conditions for the particle amplitude x_0 and angle x'_0 in front of the magnet element, we obtain the following relations for the trajectory inside the magnet:

$$x(s) = x_0 \cdot \cos\left(\sqrt{|K|}\,s\right) + x'_0 \cdot \frac{1}{\sqrt{|K|}} \sin\left(\sqrt{|K|}\,s\right)\,, \tag{18}$$

$$x'(s) = -x_0 \cdot \sqrt{|K|} \sin\left(\sqrt{|K|}\,s\right) + x'_0 \cdot \cos\left(\sqrt{|K|}\,s\right)\,. \tag{19}$$

Here, the parameter K combines the quadrupole gradient and the weak focusing effect: $K := (1/\rho^2) - k$. Usually, these two equations are combined into a more elegant and convenient matrix form,

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{foc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0\,, \tag{20}$$

where the matrix $\mathbf{M}_{\text{foc}}$ contains all the relevant information about the magnet element:

$$\mathbf{M}_{\text{foc}} = \begin{pmatrix} \cos(\sqrt{|K|}\,s) & \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|}\,s) \\ -\sqrt{|K|} \sin(\sqrt{|K|}\,s) & \cos(\sqrt{|K|}\,s) \end{pmatrix}\,. \tag{21}$$

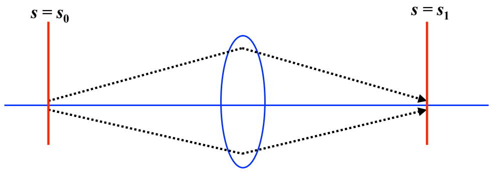


Fig. 10: Schematic illustration of the effect of a focusing quadrupole magnet

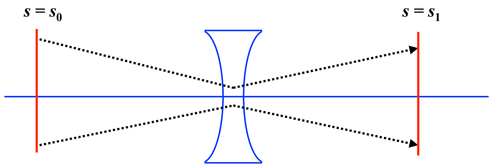


Fig. 11: Effect of a defocusing quadrupole magnet

The situation is illustrated in Fig. 10.

In the case of a defocusing magnet (or to be quite clear, also, in the case of the vertical plane, of a horizontal focusing magnet), we obtain analogously that

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{defoc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0 , \tag{22}$$

with

$$\mathbf{M}_{\text{defoc}} = \begin{pmatrix} \cosh(\sqrt{|K|}\, s) & \frac{1}{\sqrt{|K|}} \sinh(\sqrt{|K|}\, s) \\ \sqrt{|K|} \sinh(\sqrt{|K|}\, s) & \cosh(\sqrt{|K|}\, s) \end{pmatrix} ; \tag{23}$$

see Fig. 11.

For completeness, we also include the situation of field-free drift. In this trivial case with $K = 0$ we obtain

$$\mathbf{M}_{\text{drift}} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} . \tag{24}$$

This matrix formalism allows us to combine the elements of a storage ring in an elegant way, and so it is straightforward to calculate particle trajectories. In this context, we would like to emphasize a few issues:

- a certain quadrupole lens will always have two opposing effects: focusing in one plane and defocusing in the other;
- et vice versa (and the other way round, for the non-Latin-speaking community);
- in linear approximation and without explicit coupling fields, such as roll angles of the quadrupoles or solenoids, the motion in the two transverse planes is uncoupled. An amplitude in the horizontal direction, e.g., will not have any influence on the vertical motion and therefore the corresponding non-diagonal elements of the matrix $M_{1,3}$, $M_{1,4}$, etc., are zero.
- It is therefore convenient to describe this simultaneous effect in the two planes in a single 4×4 matrix and define a vector for both transverse amplitudes and angles.

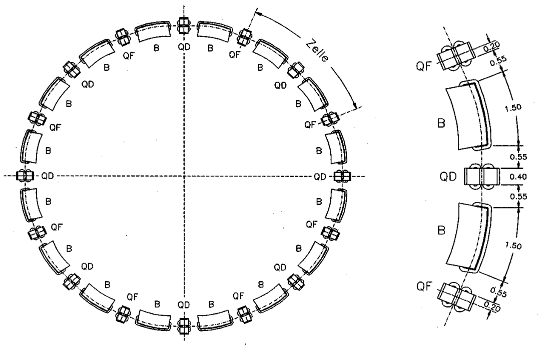


Fig. 12: A simple periodic chain of bending magnets (B) and focusing (QF) or defocusing (QD) quadrupoles forming the basic structure of a storage ring (court. [5]).

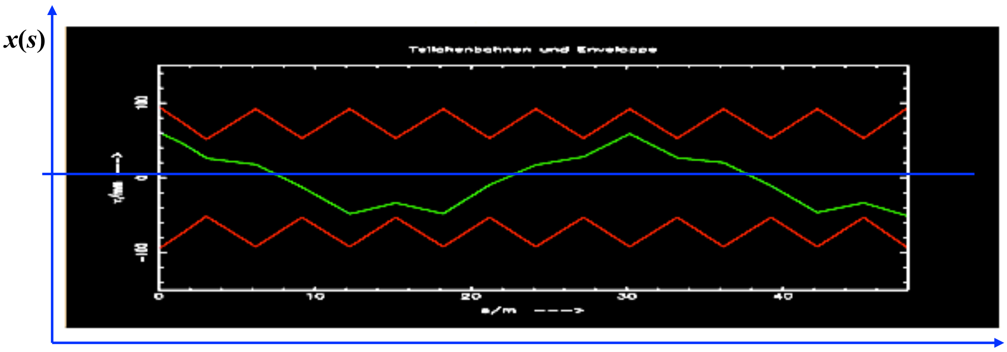


Fig. 13: Calculated particle trajectory in a simple storage ring

$$\begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix}_s = \begin{pmatrix} \cos(\sqrt{|K|}\,s) & \frac{1}{\sqrt{|K|}}\sin(\sqrt{|K|}\,s) & 0 & 0 \\ -\sqrt{|K|}\sin(\sqrt{|K|}\,s) & \cos(\sqrt{|K|}\,s) & 0 & 0 \\ 0 & 0 & \cosh(\sqrt{|K|}\,s) & \frac{1}{\sqrt{|K|}}\sinh(\sqrt{|K|}\,s) \\ 0 & 0 & \sqrt{|K|}\sinh(\sqrt{|K|}\,s) & \cosh(\sqrt{|K|}\,s) \end{pmatrix} \cdot \begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix}_0 . \tag{25}$$

As an example of a larger structure, we consider the simple case of an alternating focusing and defocusing lattice, a so-called FODO lattice [5]; see Fig. 12.

As we know the properties of each and every element in the accelerator, we can construct the corresponding matrices and calculate, step by step, the amplitude and angle of a single-particle trajectory around the ring. Even more conveniently, we can multiply out the different matrices and, given initial conditions x_0 and x_0' at a certain position in the storage ring, directly obtain the trajectory at any location in the ring:

$$\mathbf{M}_{\text{total}} = \mathbf{M}_{\text{foc}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{dipole}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{defoc}} \cdots . \tag{26}$$

The trajectory thus obtained is shown schematically in Fig. 13.

We have to point out the following facts in this context.

– At each moment, which means inside each lattice element, the trajectory is a part of a harmonic oscillation.

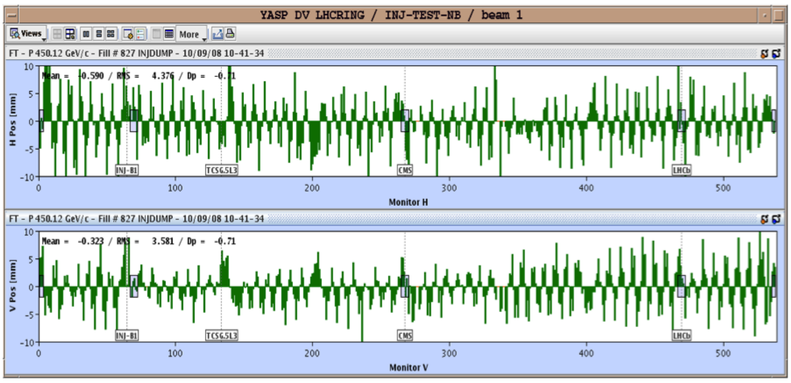


Fig. 14: Beam position measured in LHC on the first turn around the machine, during one of the very first beam injections into the LHC storage ring.

- However, because of the different restoring or defocusing forces, the solution will look different at each location.
- In the linear approximation that we have used in this context, all particles experience the same external fields, and their trajectories will differ only because of their different initial conditions.
- There seems to be an overall oscillation in both transverse planes while the particle is travelling around the ring. Its amplitude stays well within the boundaries set by the vacuum chamber, and its frequency in the example of Fig. 13 is roughly 1.4 transverse oscillations per revolution, which corresponds to the eigenfrequency of the particle under the influence of the external fields.

Coming closer to a real, existing machine, we see in Fig. 14 an orbit, measured during one of the first injections into the LHC storage ring. The horizontal oscillations are plotted in the upper half of the figure and the vertical oscillations in the lower half, on a scale of ± 10 mm. Each histogram bar indicates the value recorded during the first turn of the beam by a beam position monitor at a certain location in the ring; the orbit oscillations are clearly visible. During these first injections, a beam screen had been introduced right after the injection point. In Fig. 15, the spot of the injected beam on this screen is clearly visible as well as the one after the first turn. In both transverse planes, these spots are not yet lying on top of each other and so the orbit is not yet closed. However, this can be achieved after a straightforward orbit correction and we finally obtain what we call a 'closed orbit'.

By counting (or, better, fitting) the number of oscillations in both transverse planes, we obtain, in the case of the LHC, values of

$$Q_x = 64.31\,, \qquad Q_y = 59.32\,. \tag{27}$$

These values, which describe the eigenfrequencies of the particles, are called the horizontal and vertical *tunes*, respectively. Knowing the revolution frequency, we can easily calculate the corresponding transverse oscillation frequencies, which for this type of machine usually lie in the range of several hundred kilohertz.

As the tune characterizes the particle oscillations under the influence of all external fields, it is one of the most important parameters of a storage ring. Therefore, it is usually displayed and controlled at all times by the control system of such a machine. As an example, Fig. 16 shows the tune diagram of the HERA proton ring [11]; this was obtained via a Fourier analysis of a spectrum measured from the signal of the complete particle ensemble. The peaks indicate the two tunes in the horizontal and vertical planes of the machine; in a sufficiently linear machine, a fairly narrow spectrum is obtained.

Briefly referring back to Fig. 13, the question is what the trajectory of the particle will look like in the second turn, or the third, or after an arbitrary number of turns. Now, as we are dealing with a circular

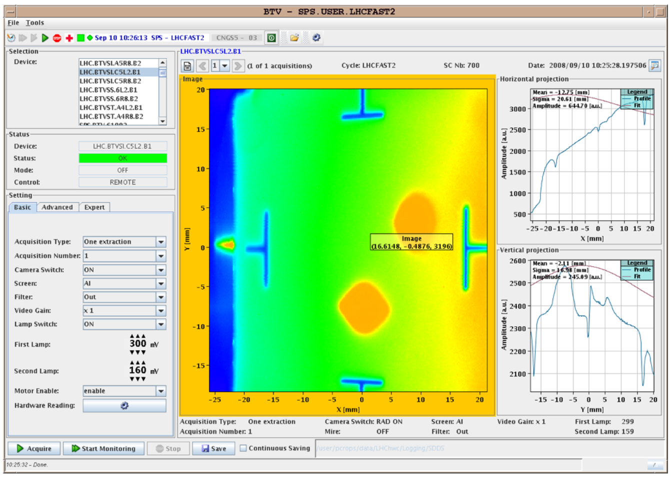


Fig. 15: Measured position of the first turn in LHC during the commissioning of the machine. The beam screen is located immediately after the injection septum and shows the spot at injection and after one full turn around the machine.

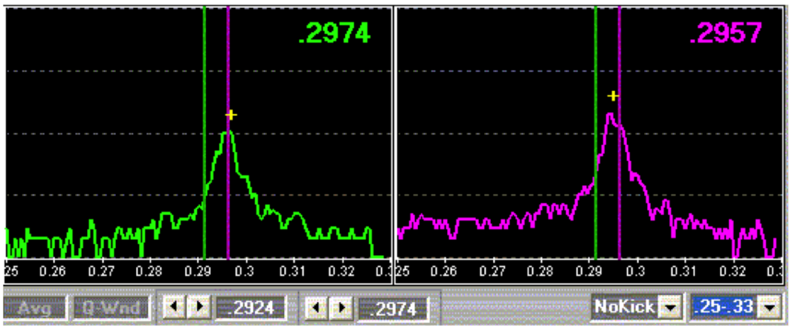


Fig. 16: Tune signal of a proton storage ring (HERA-p)

machine, the amplitude x and angle x' at the end of the first turn will be the initial conditions for the second turn, and so on. After many turns, the overlapping trajectories begin to form a pattern, such as that shown in Fig. 17, which indeed looks like a beam that here and there has a larger and a smaller size but still remains well defined in its amplitude by the external focusing forces.

3 The Twiss parameters α, β, and γ

As explained in the last section, repeating the calculations that lead to the orbit of the first turn will result in a large number of single-particle trajectories that overlap in some way and form the beam envelope. Figure 17 shows the result for 50 turns. Clearly, as soon as we are talking about many turns or many particles, the use of the single-trajectory approach is quite limited and we need a description of the beam as an ensemble of many particles. Fortunately, in the case of periodic conditions in the accelerator, there is another way to describe the particle trajectories and, in many cases, it is more convenient than the aforementioned formalism. It is important to note that, in a circular accelerator, the focusing elements are necessarily periodic in the orbit co-ordinate s after one revolution. Furthermore, storage ring lattices have an internal periodicity in most cases: they are often constructed, at least partly, from sequences in which identical magnetic structures, the lattice cells, are repeated several times in the ring and lead to periodically repeated focusing properties. In this case, the equation of motion can now be written in a

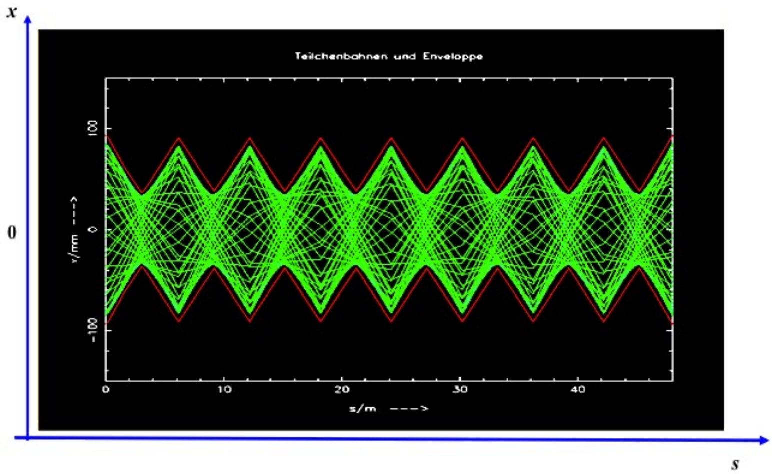


Fig. 17: Many single-particle trajectories together form a pattern that corresponds to the beam size in the ring

slightly different form:

$$x''(s) - k(s) \cdot x(s) = 0\,, \tag{28}$$

where, for simplicity, we refer to a pure quadrupole magnet and so the $1/\rho^2$ term does not appear. The main issue, however, is that unlike the previous treatment, the focusing parameters (or restoring forces) are no longer constant but are functions of the co-ordinate s. However, they are periodic in the sense that, at least after one full turn, they repeat themselves, i.e., $k(s + L) = k(s)$, leading to the so-called Hill differential equation. Following Floquet's theorem [6], the solution of this equation can be written in its general form as

$$x(s) = \sqrt{\varepsilon}\sqrt{\beta(s)}\cos(\psi(s) - \phi)\,, \tag{29}$$

where ψ is the phase of the oscillation, ϕ is its initial condition, and ε is a characteristic parameter of a single particle or, if we are considering a complete beam, of the ensemble of particles. Taking the derivative with respect to s, we get the trajectory angle x':

$$x'(s) = \sqrt{\frac{\varepsilon}{\beta(s)}}\left(\frac{1}{2}\beta'(s)\cos\left(\psi(s) - \phi\right) - \sin(\psi(s) - \phi)\right)\,. \tag{30}$$

The position and angle of the transverse oscillation of a particle at a point s are given by the value of a special amplitude function, the β-function, at that location; ε and ϕ are constants of the particular trajectory. The β-function depends in a rather complicated manner on the overall focusing properties of the storage ring. It cannot be calculated directly by an analytical approach, but instead must be either determined numerically or deduced from properties of the single-element matrices (see, e.g., Ref. [7]). In any case, like the lattice itself, it must fulfil the periodicity condition

$$\beta(s + L) = \beta(s)\,. \tag{31}$$

Inserting the solution (Eq. (29)) into the Hill equation and rearranging slightly, we get

$$\psi(s) = \int_0^s \frac{\mathrm{d}s}{\beta(s)}\,, \tag{32}$$

which describes the phase advance of the oscillation. It should be emphasized that ψ depends on the particle's oscillation amplitude. At locations where β reaches large values, i.e., the beam has a large transverse dimension, the corresponding phase advance is small; conversely, at locations where we create

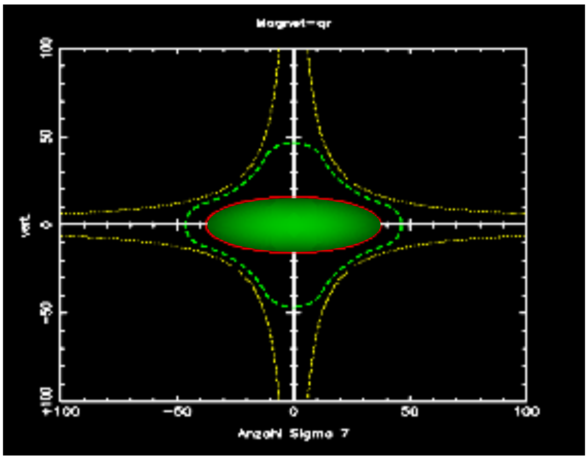


Fig. 18: Transverse beam shape inside a quadrupole magnet: plotted are 7σ of a Gaussian particle density distribution inside the vacuum chamber and magnet aperture.

a small β in the lattice, we obtain a large phase advance. In the context of Fig. 13, we introduced the tune as the number of oscillations per turn, which is nothing else than the overall phase advance of the transverse oscillation per revolution in units of 2π. So, by integrating Eq. (32) around the ring, we get, for the tune, the expression

$$Q = \frac{1}{2\pi} \oint \frac{\mathrm{d}s}{\beta(s)} \,. \tag{33}$$

The practical significance of the β-function is shown in Figs. 17 and 18. Whereas in Fig. 17 the single-particle trajectories are plotted turn by turn, Fig. 18 shows schematically a section through the transverse shape of the beam and indicates the beam size inside the vacuum chamber. The hyperbolic profile of the pole shoes of the quadrupole lens is sketched as a yellow dashed line, and the envelope of the overlapping trajectories, given by $\hat{x} = \sqrt{\varepsilon\beta(s)}$, is marked in red and is used to define the beam size in the sense of a Gaussian density distribution.

3.1 β, ε, and the phase space ellipses

Although the β-function is a somewhat abstract parameter that results from all focusing and defocusing elements in the ring, the integration constant ε has a well-defined physical interpretation. Given the solution of Hill's equation, Eq. (29), and its derivative, Eq. (30), we can transform the first equation to

$$\cos(\psi(s)) = \frac{x(s)}{\sqrt{\varepsilon\beta(s)}} \tag{34}$$

and insert the expression into Eq. (30) to get an expression for the integration constant ε:

$$\varepsilon = \gamma(s)x^2(s) + 2\alpha x(s)x'(s) + \beta(s)x'^2(s) \,. \tag{35}$$

Here, we have followed the usual convention in the literature and introduced the two parameters

$$\alpha(s) = -\frac{1}{2}\beta'(s) \tag{36}$$

and

$$\gamma(s) = \frac{1+\alpha^2(s)}{\beta(s)} \,. \tag{37}$$

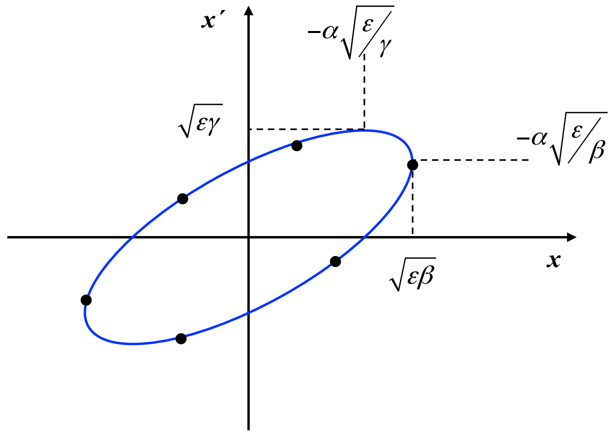


Fig. 19: Ellipse in (x, x') phase space

We obtain for ε a parametric representation of an ellipse in the (x, x') 'phase space'. The mathematical integration constant thus gains physical meaning. In fact, ε describes the space occupied by the particle in the transverse (x, x') phase space (simplified here to a two-dimensional space). More specifically, the area in the (x, x') space that is covered by the particle is given by

$$A = \pi \cdot \varepsilon\,, \tag{38}$$

and, as long as we consider only conservative forces acting on the particle, this area is constant according to Liouville's theorem. Here we take these facts as given, but we should point out that, as a direct consequence, the so-called emittance ε cannot be influenced by any external fields; it is a property of the beam, and we have to take it as given and handle it with care.

To be more precise, and following the usual textbook treatment of accelerators, we can draw the ellipse of the particle's transverse motion in phase space; see, for example, Fig. 19. Although the shape and orientation are determined by the optics function β and its derivative, $\alpha = -\frac{1}{2}\beta'$, and so change as a function of the position s, the area covered in phase space is constant.

In Fig. 19, expressions for the dependence of the beam size and divergence and, as a consequence, the shape and orientation of the phase space ellipse are included. For the sake of simplicity, we shall not derive these expressions here; instead, see Ref. [7].

Referring again to the single-particle trajectory (see Fig. 13), but now plotting the co-ordinates x and x' for a given position s in the ring, turn by turn, we obtain the phase space co-ordinates of the particle as shown in Fig. 19 (marked as dots in the figure). These co-ordinates follow the form of an ellipse, whose shape and orientation are defined by the optical parameters at the reference position s in the ring. Each point in Fig. 19 represents the transverse co-ordinates for a certain turn at that position in the ring, and the particle performs, from one turn to the next, a number of revolutions in phase space that corresponds to its tune. We have already emphasized that, as long as only conservative forces are considered (i.e., no interaction between the particles in a bunch, no collisions with remaining gas molecules, no radiation effects, etc.), the size of the ellipse in (x, x') space is constant and can be considered as a quality factor of a single particle. Large areas in (x, x') space mean large amplitudes and angles of transverse particle motion, and we would consider this as indicating a low particle 'quality'.

Let us now talk a little more about the beam as an ensemble of many (typically 10^{11}) particles. Referring to Eq. (29), at a given position in the ring the single particle amplitude is defined by the emittance ε and the function β. Thus, at a certain moment in time, the cosine term in Eq. (29) will be equal to one and the amplitude of the trajectory will reach its maximum value. Now, if we consider

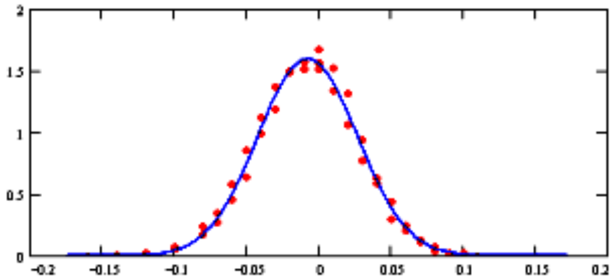


Fig. 20: Transverse particle distribution in a storage ring. The dots correspond to the measurement, the line is a Gaussian fit. A particle at 1σ from the beam centre is used to represent the beam size.

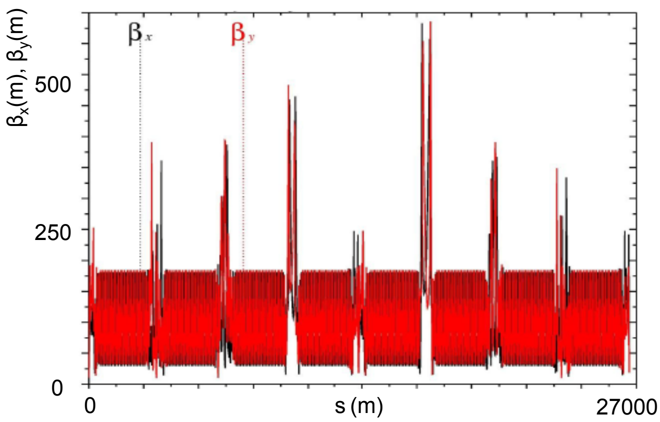


Fig. 21: LHC injection beam optics: owing to the larger beam emittance at low energy, the β-function has to be limited to values of about 600 m.

a particle at one standard deviation (sigma) of the transverse density distribution, then by using the emittance of this reference particle we can calculate the size of the complete beam, in the sense that the complete area (within one sigma) of all particles in the (x, x') phase space is surrounded (and so defined) by our one-sigma candidate. Thus, the value $\sqrt{\varepsilon \cdot \beta(s)}$ defines the one-sigma beam size in the transverse plane.

An example of such a particle density distribution is shown in Fig. 20. The dots correspond to the measured values of the particle distribution at the collision point and the blue curve represents a Gaussian fit. The emittance (usually referred to as 'Courant – Snyder invariant') of the single particle at 1σ from the centre can be used as representative emittance of the beam ensemble.

It is the task of the lattice designer to establish a beam optics that guarantees – for a given emittance – values of the β-function that lead to tolerable beam sizes at every location in the machine. As an example, we shall use the values for the LHC proton beam (Fig. 21). In the periodic pattern of the arc, the β-function is equal to 180 m and the emittance ε at the flat-top energy is roughly 5×10^{-10} rad m. The resulting typical beam size is, therefore, 0.3 mm. Now, clearly, we would not design a vacuum aperture for the machine based on a one-sigma beam size; typically, an aperture requirement corresponding to 12σ is a good rule to guarantee a sufficient beam lifetime, allowing for tolerances arising from magnet misalignment, optics errors, orbit fluctuations, and operational flexibility. In Fig. 22, part of the LHC vacuum chamber is shown, including the beam screen used to protect the cold bore from synchrotron radiation; this corresponds to a minimum beam size of 18σ.

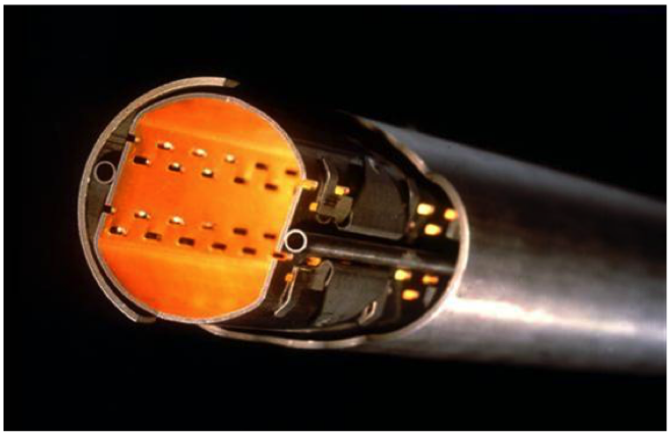


Fig. 22: LHC vacuum chamber with beam screen to shield the bore of the superconducting magnet from synchrotron radiation.

3.2 Adiabatic shrinking

The definition of the beam emittance described in the previous section bears a certain problem. Strictly speaking, Liouville's theorem states that – given conservative forces – the particle density in the phase space x, p_x is constant. Now in accelerator physics we are talking about particle amplitudes and angles, x, x' and a co-ordinate system defined by these variables is sometimes called the *trace space* to make a clear distinction from the *phase space*.

The main issue is related to the particle acceleration. The angle x' of a particle is given by the ratio between longitudinal and transverse momentum:

$$x' = \frac{\mathrm{d}x}{\mathrm{d}s} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{p_x}{p_s} \propto \frac{1}{m_0 c\beta\gamma}\,, \tag{39}$$

where we express the relativistic momentum as a function of the rest mass m_0, and the relativistic parameters $\beta = v/c$ and the Lorentz factor $\gamma = \{1/\sqrt{(1-\beta^2)}\}$. p_s describes the longitudinal component of the particle's momentum; and it is this longitudinal component that increases during the acceleration. Now, Liouville's theorem states that for the canonical conjugate variables x and p_x the phase space area is constant:

$$\int p_x \mathrm{d}x = \text{constant}\,, \tag{40}$$

and we will not argue about that.

However, in the rather sloppy interpretation that we have used until now, we refer to a co-ordinate system x, x' and so in reality we get

$$\int x' \mathrm{d}x = \frac{\int p_x}{p_s} \mathrm{dx} \tag{41}$$

and as, during acceleration, the longitudinal momentum is obviously increasing, our x–x' ellipse will shrink proportionally to $1/\beta\gamma$. We conclude, therefore, that the beam emittance and, as a consequence, the beam dimension in both transverse planes will shrink during acceleration and this is indeed what we observe. As a consequence, a proton beam in a synchrotron or an electron beam in a linac will have the largest emittance at injection energy and it is here, where the beam optics – expressed as a β-function – will have to be optimized for sufficient free aperture. The effect can be quite impressive: in Figs. 23 and 24, the 7σ envelope of a proton beam is shown inside the vacuum chamber (dashed line) of a mini-beta quadrupole magnet. Figure 23 shows the situation at 40 GeV injection energy. Figure 24 – for the same beam optics, and at the same location again – shows the 7σ envelope, but now at flat-top energy

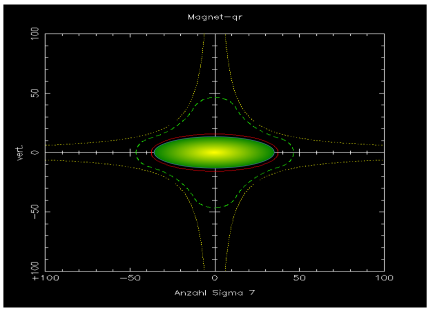


Fig. 23: Beam envelope of HERA proton ring at 40 GeV injection energy. The plot refers to 7σ at the mini-beta quadrupoles.

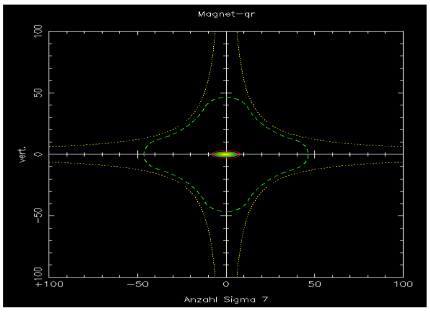


Fig. 24: Beam envelope of the HERA proton ring at 920 GeV flat-top energy. The plot refers to 7σ at the mini-beta quadrupoles.

of 920 GeV. Owing to the much higher energy, the beam size is smaller by a factor of $\sqrt{920/40}$ and the beam lifetime was considerably increased on the energy ramp of this machine.

As a direct consequence, we conclude that beam optics that lead to large beta functions in the ring only can be applied at highest energy, where, due to the reduced emittance, the overall beam size still can be limited. As an example, we refer again to the LHC situation. In direct comparison with the low-energy optics shown in Fig. 21, we now present the optics applied for high-energy collisions, Fig. 25. Here we can afford values of β of up to 4.5 km.

For completeness, we have to mention that as soon as synchrotron light effects must be considered, the situation changes drastically. In electron synchrotrons, the beam dynamics is determined by the equilibrium between synchrotron radiation damping and excitation due to the emitted photon quanta. Therefore, in those cases, we observe a quadratic increase of the emittance with energy [14]

4 Errors in field and gradient

So far, we have treated the beam and the equation of motion as a mono-energetic problem. Unfortunately, in the case of a realistic beam, we have to deal with a considerable distribution of the particles with respect to energy or momentum. A typical value is

$$\frac{\Delta p}{p} \approx 1.0 \cdot 10^{-3} \, . \tag{42}$$

This momentum spread leads to several effects concerning the bending of the dipole magnets and the focusing strength of the quadrupoles. It turns out that the equation of motion, which has been a homogeneous differential equation until now, acquires a non-vanishing term on the right-hand side.

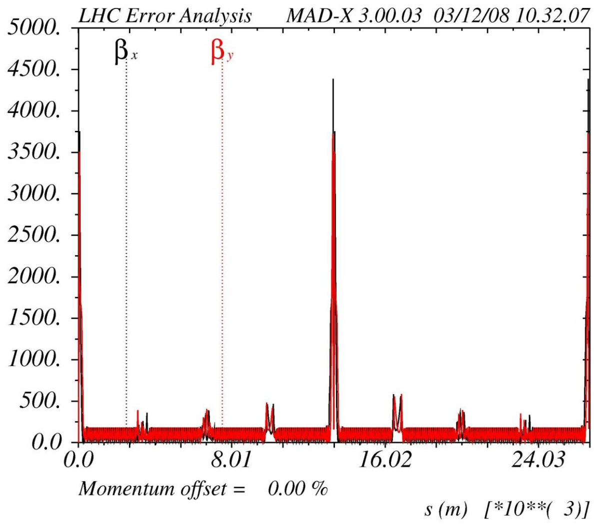


Fig. 25: LHC beam optics at high energy: Due to the small beam emittance at high energy, large values of the amplitude function β can be accepted; to be compared with the situation at injection energy, Fig.21.

4.1 Dispersive effects

Replacing the ideal momentum p in Eq. (10) with $p_0 + \Delta p$, we obtain in approximation of small Δp, instead of Eq. (16):

$$x'' + x \cdot \left(\frac{1}{\rho^2} - k\right) = \frac{\Delta p}{p_0} \cdot \frac{1}{\rho} . \tag{43}$$

The general solution of our now inhomogeneous differential equation is, therefore, the sum of the solution of the homogeneous equation of motion and a particular solution of the inhomogeneous equation:

$$x(s) = x_\beta(s) + x_\mathrm{i}(s) . \tag{44}$$

Here, x_β is the solution that we have discussed up to now and x_i is an additional contribution that still has to be determined. For convenience, we usually normalize this second term and define a special function, the so-called dispersion:

$$D(s) = \frac{x_\mathrm{i}(s)}{\Delta p/p_0} . \tag{45}$$

This describes the dependence of the additional amplitude of the transverse oscillation on the momentum error of the particle. In other words, it fulfils the condition

$$x_\mathrm{i}''(s) + K(s) \cdot x_\mathrm{i}(s) = \frac{1}{\rho} \cdot \frac{\Delta p}{p_0} . \tag{46}$$

As before, we have combined the weak and strong focusing effects in the parameter $K := (1/\rho^2) - k$. The dispersion function is defined by the magnet lattice and is usually calculated by optics programs in the context of the calculation of the usual optical parameters. Analytically, it can be determined for single elements via the expression

$$D(s) = S(s) \cdot \int \frac{1}{\rho(\bar{s})} C(\bar{s})\, \mathrm{d}\bar{s} - C(s) \cdot \int \frac{1}{\rho(\bar{s})} S(\bar{s})\, \mathrm{d}\bar{s} , \tag{47}$$

where $S(s)$ and $C(s)$ correspond to the sine-like and cosine-like elements of the single-element matrices or of the corresponding product matrix if there are several elements considered in the lattice.

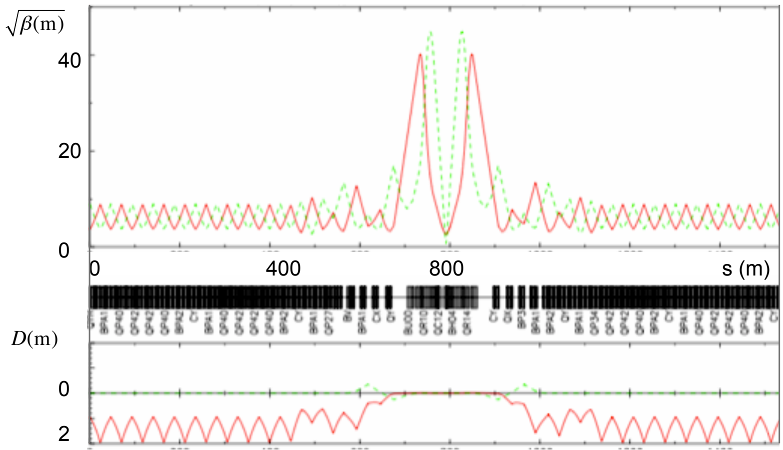


Fig. 26: β-function (top) and dispersion (bottom) of a typical high-energy collider ring

Although this all sounds somewhat theoretical, we would like to stress that typical values for the beam size and dispersive effect in the case of a high-energy storage ring are

$$x_\beta \approx 1\text{–}2\ \text{mm}\,, \qquad\qquad D(s) \approx 1\text{–}2\ \text{m}\,. \tag{48}$$

Thus, for a typical momentum spread of $\Delta p/p = 1 \cdot 10^{-3}$, we obtain an additional contribution to the beam size from the dispersion function that is of the same order as that from the betatron oscillations, x_β. An example of a high-energy beam optics system including the dispersion function is shown in Fig. 26. It should be pointed out that the dispersion describes the special orbit that an ideal particle would have in the absence of betatron oscillations ($x_\beta = x'_\beta = 0$) for a momentum deviation of $\Delta p/p = 1$. Nevertheless, it describes 'just another particle orbit' and so it is subject to the focusing forces of the lattice elements, as seen in the figure.

4.2 Chromaticity

Whereas dispersion is a problem that describes the non-ideal bending effect of dipoles for the case of a momentum error (or spread) in the particles, the careful reader will not be surprised to learn that a similar effect exists for the quadrupole focusing. We call this *chromaticity*. The chromaticity Q' describes an optical error of a quadrupole lens in an accelerator: for a given magnetic field, i.e., gradient of the quadrupole magnet, particles with a smaller momentum will feel a stronger focusing force, and particles with a larger momentum will feel a weaker force. The situation is shown schematically in Fig. 27. As a consequence, the tune of an individual particle will change, and the chromaticity Q' relates the resulting tune shift to the relative momentum error of the particle. In linear approximation, we write

$$\Delta Q = Q' \cdot \frac{\Delta p}{p_0}\,. \tag{49}$$

Q' is a consequence of the focusing properties of the quadrupole magnets and is thus given by the characteristics of the lattice. For small momentum errors $\Delta p/p_0$, the focusing parameter k can be written as

$$k(p) = \frac{g}{p/e} = \frac{ge}{p_0 + \Delta p}\,, \tag{50}$$

where g denotes the gradient of the quadrupole lens, p_0 the design momentum, and the term Δp refers to the momentum error. If Δp is small, as we have assumed, we can write in a first-order approximation,

$$k(p) \approx \frac{ge}{p_0}\left(1 - \frac{\Delta p}{p_0}\right)\,. \tag{51}$$

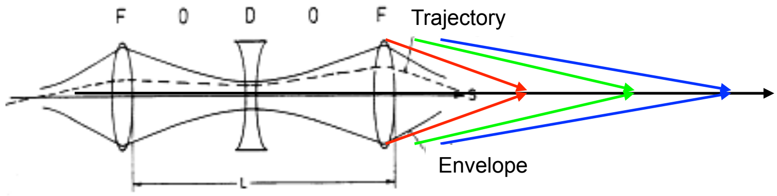


Fig. 27: Chromaticity effect in a quadrupole lens

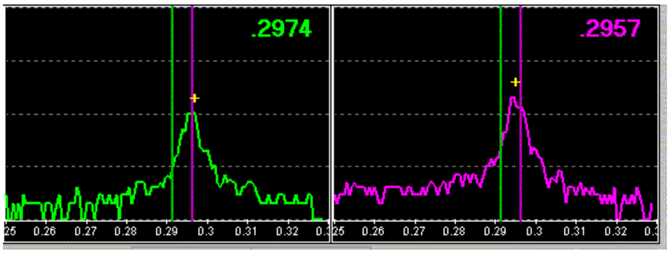


Fig. 28: Tune spectrum of a proton beam with a well-corrected chromaticity $Q' \approx 1$

This describes a quadrupole error

$$\Delta k = -k_0 \cdot \frac{\Delta p}{p}\,, \tag{52}$$

and so

$$\Delta Q = \frac{1}{4\pi} \int \Delta k\, \beta(s)\, \mathrm{d}s\,, \tag{53}$$

$$\Delta Q = -\frac{1}{4\pi} \frac{\Delta p}{p} \int k_0 \beta(s)\, \mathrm{d}s\,. \tag{54}$$

The negative sign indicates that a positive momentum deviation leads to a weaker focusing strength and, accordingly, to a lower tune. By definition, the linear chromaticity Q' of a lattice is therefore given by

$$Q' = -\frac{1}{4\pi} \int k(s)\beta(s)\, \mathrm{d}s\,. \tag{55}$$

Now, unfortunately, although the dispersion created in the dipole magnets requires nothing more than some bigger aperture in the vacuum chamber, the chromaticity of the quadrupoles has an influence on the tune of the particles and so can lead to dangerous resonance conditions. Particles with a particular momentum error will be pushed into resonances and will be lost within a very short time. A look at the tune spectrum visualizes the problem. Whereas an ideal situation leads to a well-compensated chromaticity and the particles oscillate with basically the same frequency (Fig. 28), a non-corrected chromaticity (Q' = 20 units in the case of Fig. 29) broadens the tune spectrum and a number of particles are pushed towards dangerous resonance lines.

In large synchrotrons, and storage rings in particular, this problem is crucial and represents one of the major factors that limit machine performance: because of the strong focusing of the quadrupoles and the large values of the $\beta-$function obtained, the chromaticity can reach considerable values. A chromaticity correction scheme is therefore indispensable. The trick goes in three steps.

– We sort the particles in the horizontal plane according to their momentum. This is done whenever we have a non-vanishing dispersion, for example, close to the focusing quadrupoles in the arc,

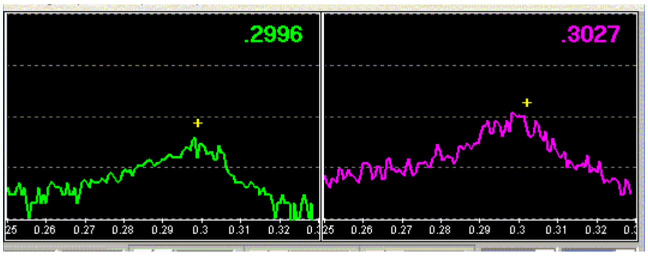


Fig. 29: Tune spectrum of a proton beam with a poorly matched chromaticity $Q' \approx 20$

where both the dispersion and the β-function reach high values and the particle trajectories are determined by the well-known relation $x_\mathrm{d}(s) = D(s) \cdot \Delta p/p$.

– At these places, we create magnetic fields that have a position-dependent focusing strength, in other words, fields that represent a position-dependent gradient. Sextupole magnets have exactly this property: if g' describes the strength of the sextupole field, we get

$$B_x = g' \cdot xy \tag{56}$$

for the actual horizontal field component and

$$B_y = g' \frac{1}{2} \cdot (x^2 - y^2) \tag{57}$$

for the vertical component. The resulting effective gradient in both planes is obtained as

$$\frac{\mathrm{d}B_x}{\mathrm{d}y} = \frac{\mathrm{dB_y}}{\mathrm{d}x} = g' \cdot x \,. \tag{58}$$

– We now only have to adjust the strengths of two sextupole families (one each to compensate the horizontal and vertical chromaticities, respectively) to get an overall correction in both planes.

In a little more detail, and referring again to normalized gradients, we can write for the normalized quadrupole gradient of an off-centre particle in a sextupole magnet

$$k_\mathrm{sext} = \frac{e}{p} g' \cdot x_d = m \cdot x_d \,, \tag{59}$$

Here we have explicitly written x_d to point out that it is the dispersive amplitude that is creating the position dependent quadrupole effect and so leads, for a given particle amplitude

$$x_\mathrm{d} = D \cdot \frac{\Delta p}{p} \,, \tag{60}$$

to the momentum-dependent focusing strength of the sextupole magnet:

$$k_\mathrm{sext} = m \cdot D \frac{\Delta p}{p} \,. \tag{61}$$

The combined effect of the so-called natural chromaticity created by the quadrupole lenses (Eq. (55)) and the compensation by the sextupoles leads to an overall chromaticity

$$Q' = -\frac{1}{4\pi} \oint (K(s) - m(s) \cdot D(s)) \beta(s) \, \mathrm{d}s \tag{62}$$

and needs to be compensated to zero in both transverse planes.

To summarize and make things as crystal clear as possible, the focusing properties of the magnet lattice lead to restoring forces in both transverse planes. The transverse motion of a particle is therefore a quasi-harmonic oscillation as the particle moves through the synchrotron, and the tune describes the frequency of these oscillations. As we cannot assume that all particles have exactly the same momentum, we have to take into account the effect of the momentum spread in the beam: the restoring forces are a function of the momentum of each individual particle and so the tune of each particle is different. We have to correct for this effect, and we do so by applying sextupole fields in regions where a non-vanishing dispersion distributes the off-momentum particles in the horizontal plane.

As easy as that!

5 Transformation of the Twiss parameters α, β, γ

"Once more unto the breach, dear friends," [12].

While it is straightforward to develop the rules for transformation of the trajectory amplitudes and angles via the single-element matrices of the lattice elements, a similar formulation can be deduced for the optical functions α, β, and γ. The derivation is closely related to the fact that – for a given energy – the beam emittance ε is constant.

Starting from the usual transformation of a trajectory amplitude x and angle x' between two locations in the lattice

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s2} = \mathbf{M} \begin{pmatrix} x \\ x' \end{pmatrix}_{s1} , \tag{63}$$

where the matrix $\mathbf{M}$ describes the focusing properties of a single lattice element, or in case of several lattice elements, it represents the product matrix, as described in Eq. (26). In the general case of a sequence of lattice elements we write

$$\mathbf{M} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} . \tag{64}$$

where the matrix elements $C, S, \ldots$ refer to the elements of the product matrix and in the trivial case of, e.g., a single focusing quadrupole are the usual descriptions that we introduced before:

$$\mathbf{M}_{\text{foc}} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \cos(\sqrt{|K|}\, s) & \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|}\, s) \\ -\sqrt{|K|} \sin(\sqrt{|K|}\, s) & \cos(\sqrt{|K|}\, s) \end{pmatrix} . \tag{65}$$

Now we consider two locations s_1 and s_2 in the storage ring, as shown schematically in Fig. 30. At both positions, the emittance can be expressed as a function of the Twiss parameters at these positions:

$$\begin{aligned} \varepsilon &= \gamma_1 x^2(s1) + 2\alpha_1 x(s1)x'(s1) + \beta_1 x'^2(s1) , \\ \varepsilon &= \gamma_2 x^2(s2) + 2\alpha_2 x(s2)x'(s2) + \beta_2 x'^2(s2) , \end{aligned} \tag{66}$$

keeping in mind that the numerical value of the emittance at both positions has to be the same, as long as Mr. Liouville is fulfilled.

Knowing the amplitude and angle of the trajectory at position s_2, we can deduce these values at position s_1:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s_1} = \mathbf{M}^{-1} \begin{pmatrix} x \\ x' \end{pmatrix}_{s_2} , \tag{67}$$

where the matrix from s_2 to s_1 is the inverse transformation matrix,

$$\mathbf{M}^{-1} = \begin{pmatrix} S' & -S \\ -C' & C \end{pmatrix} \tag{68}$$

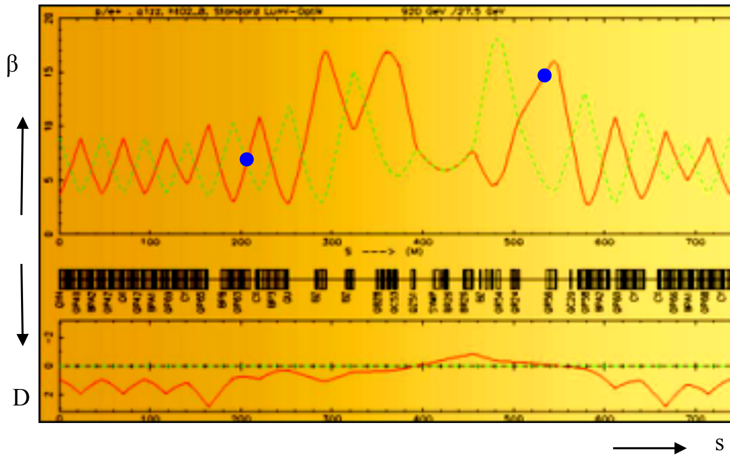


Fig. 30: The optical functions at two positions in a ring are related to each other via the constant beam emittance

and we used the fact that, for all matrices in a storage ring, the determinant has to be equal to one:

$$\det(M) = CS' - SC' = 1\,. \tag{69}$$

Thus, we can write for our trajectory co-ordinates:

$$x_1 = S' x_2 - S x'_2\,, \tag{70}$$
$$x'_1 = -C' x_2 + C x'_2\,. \tag{71}$$

Inserting these into Eq. (66), we express the emittance at position s_1 as a function of the trajectory co-ordinates at position s_2:

$$\varepsilon = \beta_1 (C x'_2 - C' x_2)^2 + 2\alpha_1 (S' x_2 - S x'_2)(C x'_2 - C' x_2) + \gamma_1 (S' x_2 - S x'_2)^2\,. \tag{72}$$

Sorting via x, x' and comparing the coefficients, we can finally relate the Twiss functions between the two locations in the ring:

$$\beta(s_2) = C^2 \beta(s_1) - 2SC\alpha(s_1) + S^2 \gamma(s_1) \tag{73}$$
$$\alpha(s_2) = -CC'\beta(s_1) + (SC' + S'C)\alpha(s_1) - SS'\gamma(s_1)\,, \tag{74}$$
$$\gamma(s_2) = C'^2 \beta(s_1) - 2S'C'\alpha(s_1) + S'^2 \gamma(s_1)\,. \tag{75}$$

Once more – for the sake of elegance in our notation – we prefer to combine these relations in matrix form and get

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_2} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & SC' + CS' & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \cdot \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_1}\,. \tag{76}$$

So if we know by calculation or measurement the optical functions at one position in the ring, we can determine them via the single-element matrices in the lattice at any other location.

Now it is high time to handle the things with care: There is something special for transfer lines and linear accelerators that we have to mention and unfortunately it is a quite uncomfortable item: unlike rings, be they synchrotrons or storage rings, the optical functions α, β, and γ are not defined in

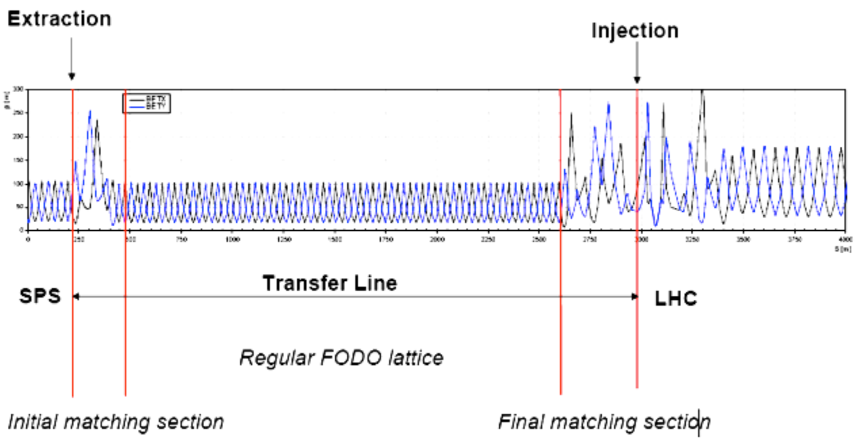


Fig. 31: Optics of the transfer line between the SPS and the HC

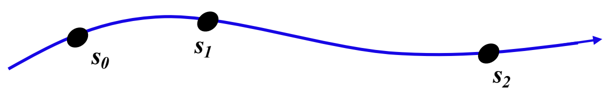


Fig. 32: The beam size must be measured at three locations in the transfer line

a non-periodic structure. Hill's equation requires a periodic system and so does its solution. Still, this description is very powerful and so our colleagues from the non-periodic world love to use our language. However, we should be aware of this issue. In Eq. (76) we have learnt how to transform the optical functions from one position in the lattice to another, knowing the focusing elements in between. This means that if we know, e.g., the Twiss functions at the beginning of a transfer line, we can calculate them through the complete linear structure. An example is given in Fig. 31.

In the case of the example in Fig. 31, the optical functions at the beginning of the structure are defined, i.e., uniquely determined by the periodicity of the SPS synchrotron. At the beginning and the end of the transfer line, special matching sections have been introduced to transform the periodic β-functions from the SPS lattice onto the transfer line structure and from here to the LHC cells. Usually, such a matching section leads for a moment to a somehow increased beam amplitude (due to the distorted β-functions) and we have to take care to limit the aperture needs to reasonable values.

More complicated, however, is the case, where a circular pre-accelerator does not exist. In such a case we have to 'guess' the initial values of α, β, and γ, or, better, we have to measure them at the initial position before we can apply Eq. (76). To do so, we must perform three measurements at three different locations in the line (see Fig. 32) and obtain three values for the beam size:

$$\sigma_1 = \sqrt{(\varepsilon\beta_1)}$$
$$\sigma_2 = \sqrt{(\varepsilon\beta_2)}$$
$$\sigma_3 = \sqrt{(\varepsilon\beta_3)}$$

And this is the way the trick goes [13]:

Assume that we measure the particle distribution, which might look like the example in Fig. 33. Fitting a reasonable ellipse to the distribution, we can deduce the standard deviation of the, hopefully,

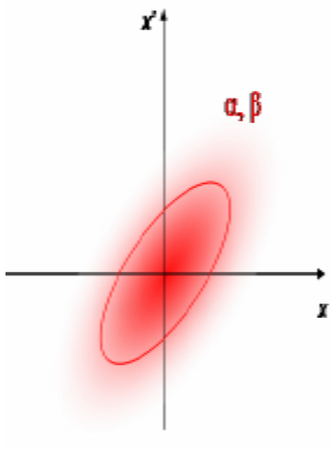


Fig. 33: Example of the transverse beam size measured in a transfer line

nicely Gaussian distributed particle density. (OK, for other than Gauss-distributions it will also work, more or less). So we obtain

$$\sigma = \sqrt{\varepsilon\beta} \tag{77}$$

and as the beam emittance is constant, we can write – referring to each of the three positions s_0, s_1, and s_2 –

$$\varepsilon = \frac{\sigma_0^2}{\beta_0} = \frac{\sigma_1^2}{\beta_1} = \frac{\sigma_2^2}{\beta_2} \,. \tag{78}$$

Now, from Eq. (76) we know how the β-functions transform through the lattice and so we can express β_1 and β_2 as a function of the initial value β_0:

$$\beta_1 = C_1^2\beta_0 - 2C_1S_1\alpha_0 + \frac{S_1^2}{\beta_0}\left(1+\alpha_0^2\right) \tag{79}$$

$$\beta_2 = C_2^2\beta_0 - 2C_2S_2\alpha_0 + \frac{S_2^2}{\beta_0}\left(1+\alpha_0^2\right) \,. \tag{80}$$

Using this information, we can determine both α_0 and β_0:

$$\alpha_0 = \frac{1}{2}\beta_0\Gamma \,, \tag{81}$$

$$\beta_0 = \frac{1}{\sqrt{\left(\frac{\sigma_1}{\sigma_0}\right)^2/S_1^2 - (C_1/S_1)^2 + (C_1/S_1)\Gamma - \Gamma^2/4}} \,, \tag{82}$$

where we introduced the parameter

$$\Gamma = \frac{(\sigma_2/\sigma_0)^2/S_2^2 - (\sigma_1/\sigma_0)^2/S_1^2 - (C_2/S_2)^2 + (C_1/S_1)^2}{C_1/S_1 - C_2/S_2} \,.$$

6 Dipole errors and quadrupole misalignment

The design orbit, and thus the geometry of the ring or transfer line, is defined by the strength and arrangement of the dipole magnets. Under the influence of imperfections in the dipole field and (transverse) misalignment of the quadrupole magnets, unwanted deflection fields ('kicks') are created that influence

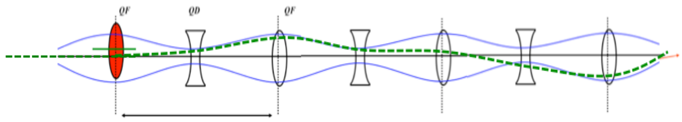


Fig. 34: Effect of a misaligned quadrupole in a transfer line

this orbit. If these distortions are small enough (and hopefully they are), we will still obtain an orbit that is not too far away from the design. It is this 'closed orbit' that acts as a reference for the single-particle trajectories and we have to take care that it does not differ too much from the design.

A special issue, however, arises from the fact that in a ring this reference orbit has, by definition, to be closed. While in a transfer line the effect of external distortions is somehow straightforward, in a periodic situation we have to be a bit more careful. A small field error δB will result in a sudden change of the particle's angle x' and so we describe the effect of a dipole error as

$$\Delta\theta = \Delta x' = \frac{\mathrm{d}l}{\rho} = \frac{\int \Delta B \mathrm{d}l}{B\rho}\,, \tag{83}$$

where we have again normalized the field error by the beam rigidity to obtain the deflection angle x'.

In a misaligned quadrupole, we get exactly the same problem. An offset Δx in the presence of a field gradient g leads to an effective dipole field that deflects the beam:

$$\Delta\theta = \frac{\int \Delta B \mathrm{d}l}{B\rho} = \frac{\int \Delta x g \mathrm{d}l}{B\rho}\,. \tag{84}$$

For a transfer line, the resulting effect is trivial. Assuming a short deflection, the amplitude x of the particle is not affected by the field, but the angle x' is and so we can write:

$$x_f = x_i = 0\,, \tag{85}$$

$$x'_f = x'_i + \Delta x' = x'_i + \frac{\int \Delta B \mathrm{d}l}{B\rho}\,. \tag{86}$$

From this moment on, the originally ideal trajectory will be transformed through the lattice elements in the usual way:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M} \begin{pmatrix} x \\ x' \end{pmatrix}_0\,, \tag{87}$$

which is shown qualitatively for a transfer line in Fig. 34.

In a circular machine, things get a bit more complicated. As we talk about a ring, the periodic boundary conditions after one turn have to be taken into account. Mathematically we express this fact by:

$$x(s+L) = x(s), \qquad\qquad x'(s+L) + \Delta x' = x'(s)\,. \tag{88}$$

While the orbit amplitude remains unchanged and the trajectory has to close upon itself, the angle x' after one turn has to take the distortion $\mathrm{d}x'$ into account (Fig. 35). Starting from the general solution of Hill's equation, and using the periodicity condition we write for the amplitude

$$x(s) = a\sqrt{\beta(s)}\cos(\psi(s) - \phi)\,, \tag{89}$$

$$x(s+L) = x(s)\,, \tag{90}$$

$$a\sqrt{\beta(s+L)}\cos(\psi(s) + 2\pi Q - \phi) = a\sqrt{\beta(s)}\cos(\psi(s) - \phi)\,. \tag{91}$$

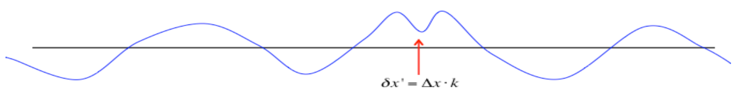


Fig. 35: Effect of a misaligned quadrupole in a ring: Following the periodicity condition, the orbit must close upon itself.

The amplitude factor a will be determined later by the periodicity conditions and, clearly enough, the phase advance per turn increases by

$$\psi(s+L) = 2\pi Q \; .$$

As the amplitude function β is periodic, by definition we obtain as first condition

$$\cos(2\pi Q - \phi) = \cos(-\phi) = \cos(\phi) \, , \tag{92}$$

$$\phi = \pi Q \, . \tag{93}$$

The boundary condition set by the amplitude fixes the initial condition for the phase ϕ.

Following the same arguments, but now for the angle x', we get

$$x(s) = a\sqrt{\beta(s)}\cos(\psi(s) - \phi) \, , \tag{94}$$

$$x'(s) = a\sqrt{\beta(s)}\,(-\sin(\psi(s) - \phi))\,\psi' + \frac{\beta'(s)}{2\sqrt{\beta}} a\cos(\psi(s) - \phi) \, , \tag{95}$$

and writing $\Delta x'$ for the local kick due to the field distortion

$$x'(s+L) + \Delta x' = x'(s) \tag{96}$$

we get

$$- a\frac{1}{\sqrt{\beta(\tilde{s}+L)}}\sin(2\pi Q - \phi) + \frac{\beta'(\tilde{s}+L)}{2\beta(\tilde{s}+L)}\sqrt{\beta(\tilde{s}+L)}a\cos(2\pi Q - \phi) + \delta x' = - a\frac{1}{\sqrt{\beta(\tilde{s})}}\sin(-\phi) + \frac{\beta'(\tilde{s})}{2\beta(\tilde{s})}\sqrt{\beta(\tilde{s})}a\cos(-\phi) \, . \tag{97}$$

Here we have explicitly written $\tilde{s}$ for the position of the dipole field error, to emphasize, that, e.g., the optical functions are to be taken at this position. Knowing that, from the periodicity condition we derived,

$$\beta(s+L) = \beta(s), \qquad\qquad \phi = \pi Q \, , \tag{98}$$

we can solve for the amplitude a and get

$$a = \delta x' \sqrt{\beta(\tilde{s})}\frac{1}{2\sin(\pi Q)} \; . \tag{99}$$

Inserting this equation into Eq. (89), we get the final result that the amplitude of the closed orbit under the influence of a dipole distortion (or quadrupole misalignment) is given by

$$x(s) = \delta x' \sqrt{\beta(\tilde{s})}\sqrt{\beta(s)}\frac{\cos(\psi(s) - \phi)}{2\sin(\pi Q)} \; . \tag{100}$$

We conclude that the distorted orbit depends on the kick strength, the local β-function at the location $\tilde{s}$ of the distortion, and the β-function at the observation point s. In addition, there is a resonance denominator,

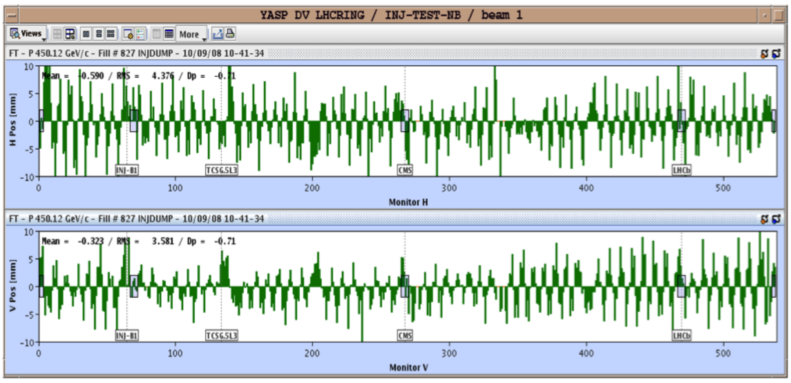


Fig. 36: Measured closed orbit in LHC during commissioning of the machine

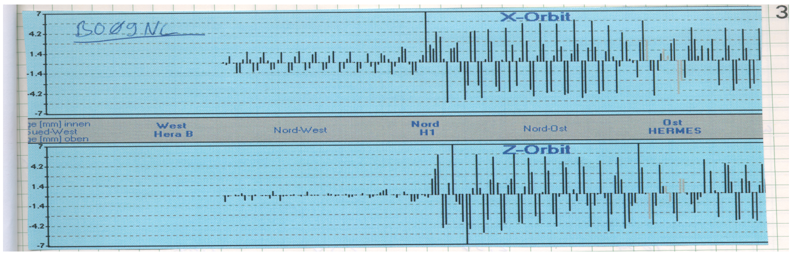


Fig. 37: Measured orbit in a beam transfer line, including the effect of a quadrupole lens that is misaligned in both planes. At the location of the misaligned magnet a sudden increase of the trajectory amplitude is observed .

which will amplify any external orbit distortion, if the tune in the corresponding plane is on, or close to an integer value. In such a case, the particle amplitude will grow ad infinitum and lead very quickly to particle losses; so better watch your tune!

For completeness: if we do not refer to a special starting point and express the orbit distortion as normalized dipole strength $\rho(\tilde{s})$, we get the general expression

$$x(s) = \frac{\sqrt{\beta(s)}}{2\sin(\pi Q)} \oint \sqrt{\beta(\tilde{s})} \frac{1}{\rho(\tilde{s})} \cos(|\psi(\tilde{s}) - \psi(s)| - \pi Q) \mathrm{d}\tilde{s} \,. \tag{101}$$

We would not like to close this section without showing a real example of orbits for both cases, a closed orbit in a storage ring and a real beam trajectory in transfer lines or linacs. We have seen an example of the first case already in Fig. 14, and we plot it here once more, for simplicity (Fig. 36). It shows the closed orbit of the LHC storage ring during the start-up phase of the machine, where considerable amplitudes in both planes were observed. The tune of the machine was set to $Q_x = 64.31$, so sufficiently away from resonance conditions. Still, the alignment tolerances of the magnets of $\Delta x \approx \Delta y \approx 150$ µm caused a considerable orbit distortion of up to 10 mm.

Figure 37 refers to the situation in a transfer line (observed at the HERA collider at DESY). While in the first part of the structure the oscillations are well corrected and small, suddenly a strong orbit fluctuation is created due to a misaligned quadrupole lens in the middle of the lattice. As the transfer line is, by definition, not closed upon itself, the observed orbit develops according to Eqs. (86) and (87).

6.1 Emittance in electron rings or linacs

There is a special issue about electron beams, that should not be forgotten: after all, they told us that this is an Accelerator School on electron machines. In Professor L. Rivkin's lecture [14], we heard that, whenever we bend an electron beam, synchrotron radiation is emitted that has a strong influence on the beam dynamics. Summarizing briefly what we learnt there, we can state:

- The power of the radiated synchrotron light depends on the energy of the particle and the bending radius of the trajectory under the influence of the field acting on it,

$$\Delta P = \frac{e^2 c}{6\pi\varepsilon_0} \frac{\gamma^4}{\rho^2} \,. \tag{102}$$

- The energy loss per turn in a circular machine is given by

$$\Delta E = \frac{e^2}{3\varepsilon_0} \frac{\gamma^4}{\rho} \,. \tag{103}$$

- The critical energy of the emitted radiation is given by

$$\omega_{\mathrm{c}} = \frac{3c}{2} \frac{\gamma^3}{\rho} \,. \tag{104}$$

The damping effect of the light emission and the quantum effect of the emitted photons lead to an equilibrium emittance of the beam that is given by

$$\varepsilon_{x_0} = \frac{C_q E^2}{J_0} \frac{\langle H \rangle_{\mathrm{mag}}}{\rho} \,, \tag{105}$$

where J_x is the so-called damping partition number, which is determined by the lattice and usually close to $J_x = 1$. The H-function describes the influence of the optical parameters α, β, and γ and the dispersion D.

$$\langle H \rangle = \gamma D^2 + 2\alpha D D' + \beta D'^2 \,; \tag{106}$$

C_q is a constant that we usually introduce to make our equations more compact. It is given by

$$C_q = \frac{55}{32\sqrt{3}} \frac{\hbar c}{(m_e c^2)^3} \,, \tag{107}$$

and for electrons it has the numerical value

$$C_q = 1.46810^{-6} \left[\frac{\mathrm{m}}{\mathrm{GeV}^2} \right] \,, \tag{108}$$

Now, while this all well known and clear and has a strong impact on the magnet strength and design of the lattice, it also affects the orbit correction scheme. Any (!) external field, including off-centre quadrupoles, and including the effect of orbit corrector dipoles, will influence the beam emittance. In the quest for smallest possible beam emittances, therefore, these effects must be taken into account.

Any deflecting field will change (i.e., create) additional beam emittance and we have to be careful when it comes to orbit corrections. Assume the two extreme cases, which are shown schematically in Figs. 38 and 39.

In the case of Fig. 38, the quadrupole magnets are slightly misaligned, leading to the orbit oscillations described above. The BPMs are assumed to be perfect and so they will show us how the real orbit looks like. Thus an orbit correction algorithm will calculate the most effective settings and tell us how to power our orbit correctors in a way to obtain a close-to-ideal orbit. Each misaligned quadrupole will lead to a orbit defection, which can and should be corrected by the corrector dipole next to it. The result will be a nearly perfect compensation of the quadrupole offsets and a nicely small beam emittance.

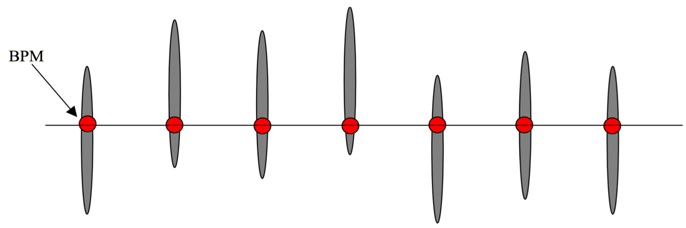


Fig. 38: Lattice with misaligned quadrupoles and perfect beam position monitors (BPMs)

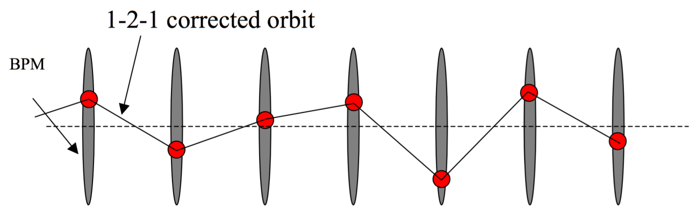


Fig. 39: Lattice with perfectly aligned quadrupoles but offsets in the reading of the beam position monitors (BPM). While the actual orbit is perfect, the beam position monitor readings simulate an orbit distortion.

Consider, however, the case of Fig. 39. Here the quadrupoles are exactly aligned, the orbit is perfect and it is the beam position monitor system that causes the trouble. Nobody is perfect and so even beam position monitors can have some reading errors that lead to artificial beam position offsets. A straightforward orbit correction approach, as explained, will reduce the beam position monitor readings, but in reality lead to a distorted orbit and so create additional dispersion in the machine and lead to increased emittances. In particular, in the vertical plane, this effect is most serious, as vertical bending fields will usually not be present in the machine and the emittance in this plane should be minimized. Special techniques are needed and have been developed to avoid such a problem. Dispersion-free steering methods are widely used [15, 16]; instead of naively correcting the orbit (i.e., the beam position monitor readings) we concentrate directly on the dispersion that can be measured in the machine and power our corrector magnets in such a way that the dispersion is minimized around the storage ring.

References

[1] B.J. Holzer, in Proc. of the CAS-CERN Accelerator School, Erice, Italy, 24 April - 4 May 2013, CERN-2014-005 (CERN, Geneva, 2014), pp. 41-56. https://doi.org/10.5170/CERN-2014-005.41

[2] F. Tecker, in Proc. of the CAS-CERN Accelerator School, Trondheim, Norway, 18 - 29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp. 9-29. https://doi.org/10.5170/CERN-2014-009.9

[3] C. Bernardini, *Phys. Persp.* **6** (2004) 156. https://doi.org/10.1007/s00016-003-0202-y

[4] LHC design report Vol. 1: the LHC Main Ring, CERN-2004-003-V-1 (CERN, Geneva, 2004), https:/doi.org/10.5170/CERN-2004-003-V-1.

[5] K. Wille, *The Physics of Particle Accelerators* (Oxford University Press, Oxford, 2000).

[6] J. Rossbach and P. Schmüser, in Proc. of the CAS-CERN Accelarator School: 5th General Acceler-

ator Physics Course, Jyväskylä, Finland, 7–18 September 1992, edited by S. Turner, CERN-1994-001, (CERN, Geneva, 1994), pp. 17-88. https://doi.org/10.5170/CERN-1994-001.17

[7] B.J. Holzer, in Proc. of the CAS-CERN Accelerator School, Trondheim, Norway, 18 - 29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp. 61-100. https://doi.org/10.5170/CERN-2014-009.61

[8] E. Jaeschke *et al.*, The Heidelberg test storage ring for heavy ions TSR, Proc. EPAC, Rome, 1988, Ed. S. Tazzari (World Scientific, Teaneck, NJ, 1988), p. 365.

[9] H. Goldstein, *Klassische Mechanik* (Akademische Verlag-Gesellschaft, Wiesbaden, 1981).

[10] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators* (Imperial College Press, London, 2014), p. 59. https://doi.org/10.1142/9781783262786-0002

[11] HERA Design Team, HERA: a proposal for a large electron proton colliding beam facility at DESY, DESY-HERA-81/10 (DESY, Hamburg, 1981).

[12] W. Shakespeare, *Henry V*, Act 3, Scene 1, Line 1.

[13] P.J. Bryant, in Proc. of the CAS-CERN Accelerator School, 5th General Accelerator Physics Course, Jyväskylä, Finland, 7–18 September 1992, CERN-1994-001 (CERN, Geneva, 1994), pp. 219-238. https://doi.org/10.5170/CERN-1994-001.17

[14] L. Rivkin, these proceedings

[15] T.O. Raubenheimer and R.D. Ruth, *Nucl. Instrum. Methods* **A302** (1991) 191. https://doi.org/10.1016/0168-9002(91)90403-D

[16] R. Assmann, *et al.*, Emittance optimization with dispersion free steering at LEP, CERN-SL-2000-078-OP, (CERN, Geneva, 2000).

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Space Charge Mitigation

M. Ferrario
Frascati National Laboratory, National Institute for Nuclear Physics, Rome, Italy

Abstract
In this paper we introduce, from basic principles, the main concepts of beam focusing and transport of space charge dominated beams in high brightness accelerators using the beam envelope equation as a convenient mathematical tool. Matching conditions suitable for preserving beam quality are derived from the model for significant beam dynamics regimes.

Keywords
Beam matching; rms emittance; laminar beam; space charge effects; rms envelope equations.

1 Introduction

Light sources based on high-gain free electron lasers or future high-energy linear colliders require the production, acceleration, and transport up to the interaction point of low divergence, high charge-density electron bunches [1]. Many effects contribute in general to degradation of the final beam quality, including chromatic effects, wake fields, emission of coherent radiation, and accelerator misalignments. Space charge effects and mismatch with focusing and accelerating devices typically contribute to emittance degradation of high charge-density beams [2]; hence, control of beam transport and acceleration is the leading edge for high-quality beam production.

Space charge effects represent a very critical issue and a fundamental challenge for high-quality beam production and its applications. Without proper matching, significant emittance growth may occur when the beam is propagating through different stages and components owing to the large differences of transverse focusing strength. This unwanted effect is even more serious in the presence of finite energy spread.

In this paper we introduce, from basic principles, the main concepts of beam focusing and transport in modern accelerators using the beam envelope equation as a convenient mathematical tool. Matching conditions suitable for preserving beam quality are derived from the model for significant beam dynamics regimes. A more detailed discussion of the previous topics can be found in the many classical textbooks on this subject, as listed in Refs. [3–6].

2 Laminar and non-laminar beams

An ideal high-charge particle beam has orbits that flow in layers that never intersect, as occurs in a laminar fluid. Such a beam is often called a laminar beam. More precisely, a laminar beam satisfies the following two conditions [6]:

i) all particles at a given position have identical transverse velocities. On the contrary, the orbits of two particles that start at the same position could separate and later cross each other;

ii) assuming that the beam propagates along the z axis, the magnitudes of the slopes of the trajectories in the transverse directions x and y, given by $x'(z) = \mathrm{d}x/\mathrm{d}z$ and $y'(z) = \mathrm{d}y/\mathrm{d}z$, are linearly proportional to the displacement from the z axis of beam propagation.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.177

Trajectories of interest in beam physics are always confined to the inside of small, near-axis regions, and the transverse momentum is much smaller than the longitudinal momentum, $p_{x,y} << p_z \approx p$. As a consequence, it is possible in most cases to use the small angle, or *paraxial*, approximation, which allows us to write the useful approximate expressions $x' = p_x/p_z \approx p_x/p$ and $y' = p_y/p_z \approx p_x/p$.

To help understand the features and advantages of a laminar beam propagation, the following figures compare the typical behaviour of a laminar and a non-laminar (or thermal) beam.

Figure 1 illustrates an example of orbit evolution of a laminar mono-energetic beam with half width x_0 along a simple beam line with an ideal focusing element (solenoid, magnetic quadrupoles, or electrostatic transverse fields are usually adopted to this end), represented by a thin lens located at the longitudinal coordinate $z = 0$. In an ideal lens, focusing (defocusing) forces are linearly proportional to the displacement from the symmetry axis z, so that the lens maintains the laminar flow of the beam.

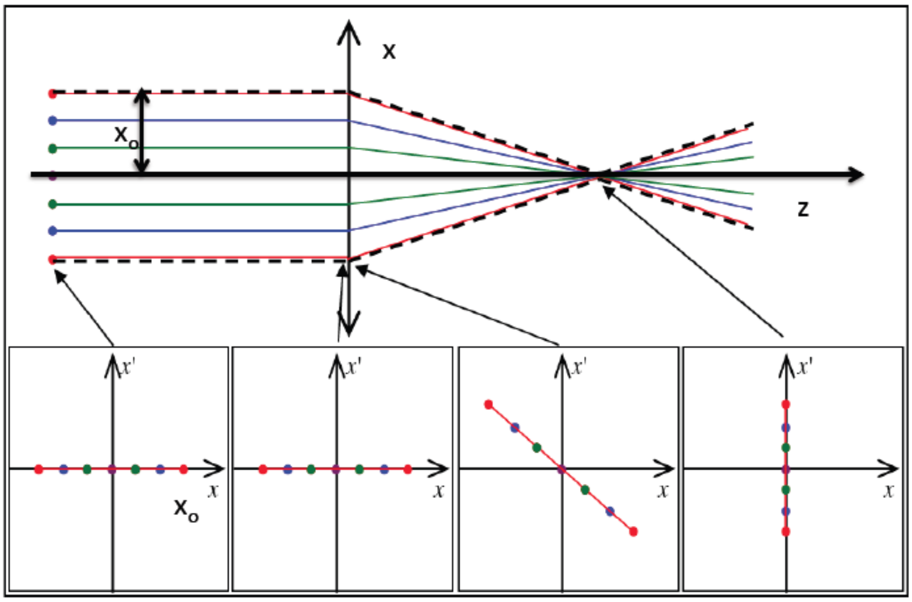


Fig. 1: Particle trajectories and phase space evolution of a laminar beam [7]

The beam shown in Fig. 1 starts propagating completely parallel to the symmetry axis z; in this particular case, the particles all have zero transverse velocity. There are no orbits that cross each other in such a beam. Ignoring collisions and inner forces, such as Coulomb forces, a parallel beam could propagate an infinite distance with no change in its transverse width. When the beam crosses the ideal lens, it is transformed into a converging laminar beam. Because the transverse velocities after the linear lens are proportional to the displacement off axis, particle orbits define similar triangles that converge to a single point. After passing through the singularity at the focal point, the particles follow diverging orbits. We can always transform a diverging (or converging) beam into a parallel beam by using a lens of the proper focal length, as can be seen by reversing the propagation axis of Fig. 1.

The small boxes in the lower part of the figure depict the particle distributions in the trace space (x, x'), equivalent to the canonical phase space (x, $p_x \approx x'p$) when p is constant, i.e., without beam acceleration. The phase space area occupied by an ideal laminar beam is a straight segment of zero thickness. As can be easily verified, the condition that the particle distribution has zero thickness proceeds from condition 1; the segment straightness is a consequence of condition 2. The distribution of a laminar beam propagating through a transport system with ideal linear focusing elements is thus a straight segment with variable slope.

Particles in a non-laminar beam have a random distribution of transverse velocities at the same location and a spread in directions, as shown in Fig. 2. Because of the disorder of a non-laminar beam, it is impossible to focus all particles from a location in the beam toward a common point. Lenses can influence only the average motion of particles. Focal spot limitations are a major concern for a wide variety of applications, from electron microscopy to free electron lasers and linear colliders. The phase space plot of a non-laminar beam is no longer a straight line: the beam, as shown in the lower boxes of Fig. 2, occupies a wider area of the phase space.

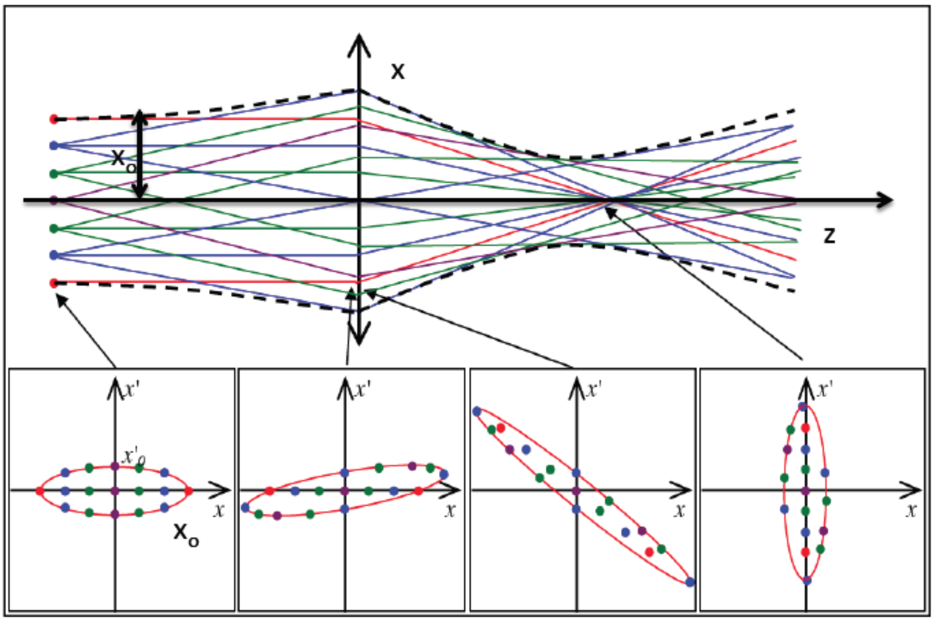


Fig. 2: Particle trajectories and phase space evolution of a non-laminar beam [7]

3 The emittance concept

The phase space surface A occupied by a beam is a convenient figure of merit for designating the quality of a beam. This quantity is the emittance ε_x and is usually represented by an ellipse that contains the whole particle distribution in the phase space (x, x'), such that $A = \pi\varepsilon_x$. An analogous definition holds for the (y, y') and (z, z') planes. The original choice of an elliptical shape comes from the fact that when linear focusing forces are applied to a beam, the trajectory of each particle in phase space lies on an ellipse, which may be called the trajectory ellipse. Being the area of the phase space, the emittance is measured in metres radians. More often is expressed in millimetres milliradians or, equivalently, in micrometres.

The ellipse equation is written as

$$\gamma_x x^2 + 2\alpha_x xx' + \beta_x x'^2 = \varepsilon_x \ , \tag{1}$$

where x and x' are the particle coordinates in the phase space and the coefficients $\alpha_x(z)$, $\beta_x(z)$, and $\gamma_x(z)$ are called Twiss parameters, which are related by the geometrical condition:

$$\beta_x \gamma_x - \alpha_x^2 = 1 \ . \tag{2}$$

As shown in Fig. 3, the beam envelope boundary $X_{\max}$, its derivative $(X_{\max})'$, and the maximum beam divergence $X'_{\max}$, i.e., the projection on the axes x and x' of the ellipse edges, can be expressed as a function of the ellipse parameters:

$$\begin{cases} X_{\max} = \sqrt{\beta_x \varepsilon_x} \\ \left(X_{\max}\right)' = -\alpha \sqrt{\dfrac{\varepsilon}{\beta}} \\ X'_{\max} = \sqrt{\gamma_x \varepsilon_x} \end{cases} \quad . \tag{3}$$

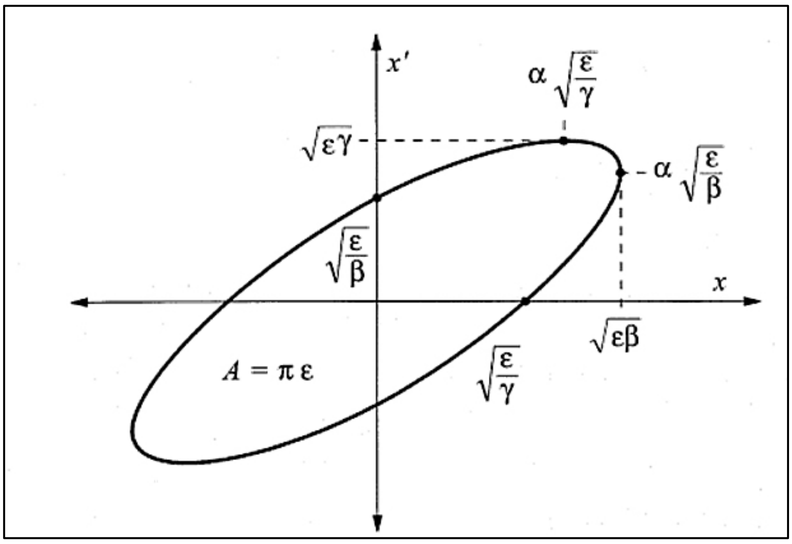


Fig. 3: Phase space distribution in a skewed elliptical boundary, showing the relationship of Twiss parameters to the ellipse geometry [6].

According to Liouville's theorem, the six-dimensional (x, p_x, y, p_y, z, p_z) phase space volume occupied by a beam is constant, provided that there are no dissipative forces, no particles lost or created, and no Coulomb scattering among particles. Moreover, if the forces in the three orthogonal directions are uncoupled, Liouville's theorem also holds for each reduced phase space surface, (x, p_x), (y, p_y), (z, p_z), and hence emittance also remains constant in each plane [3].

Although the net phase space surface occupied by a beam is constant, non-linear field components can stretch and distort the particle distribution in the phase space, and the beam will lose its laminar behaviour. A realistic phase space distribution is often very different from a regular ellipse, as shown in Fig. 4.

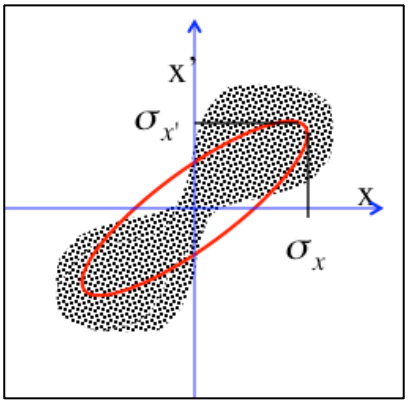


Fig. 4: Typical evolution of phase space distribution (black dots) under the effects of non-linear forces with the equivalent ellipse superimposed (red line).

We introduce, therefore, a definition of emittance that measures the beam quality rather than the phase space area. It is often more convenient to associate a statistical definition of emittance with a generic distribution function $f(x, x', z)$ in the phase space; this is the so-called *root mean square (rms) emittance*:

$$\gamma_x x^2 + 2\alpha_x x x' + \beta_x x'^2 = \varepsilon_{x,\mathrm{rms}} \ . \tag{4}$$

The rms emittance is defined such that the equivalent-ellipse projections on the x and x' axes are equal to the rms values of the distribution, implying the following conditions:

$$\begin{cases} \sigma_x = \sqrt{\beta_x \varepsilon_{x,\mathrm{rms}}} \\ \sigma_{x'} = \sqrt{\gamma_x \varepsilon_{x,\mathrm{rms}}} \end{cases} , \tag{5}$$

where

$$\begin{cases} \sigma_x^2(z) = \langle x^2 \rangle = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x^2 f(x, x', z)\, \mathrm{d}x \mathrm{d}x' \\ \sigma_{x'}^2(z) = \langle x'^2 \rangle = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x'^2 f(x, x', z)\, \mathrm{d}x \mathrm{d}x' \end{cases} \tag{6}$$

are the second moments of the distribution function $f(x, x', z)$. Another important quantity that accounts for the degree of (x, x') correlations is defined as

$$\sigma_{xx'}(z) = \langle x x' \rangle = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x x' f(x, x', z)\, \mathrm{d}x \mathrm{d}x' \ . \tag{7}$$

From Eq. (3) it also holds that

$$\sigma'_x = \frac{\sigma_{xx'}}{\sigma_x} = -\alpha_x \sqrt{\frac{\varepsilon_{x,\mathrm{rms}}}{\beta_x}}$$

see also Eq. (16), which allows us to link the correlation moment, Eq. (7), to the Twiss parameter as

$$\sigma_{xx'} = -\alpha_x \varepsilon_{x,\mathrm{rms}} \ . \tag{8}$$

One can easily see from Eqs. (3) and (5) that

$$\alpha_x = -\frac{1}{2} \frac{\mathrm{d}\beta_x}{\mathrm{d}z}$$

also holds.

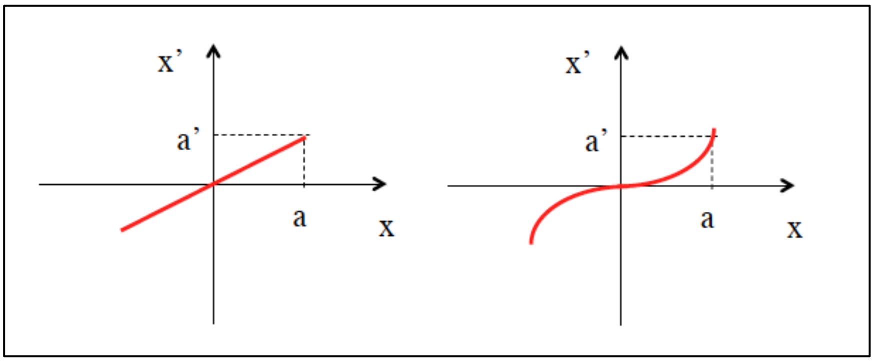
By substituting the Twiss parameter defined by Eqs. (5) and (8) into condition 2 we obtain [5]

$$\frac{\sigma_{x'}^2}{\varepsilon_{x,\mathrm{rms}}} \frac{\sigma_x^2}{\varepsilon_{x,\mathrm{rms}}} - \left(\frac{\sigma_{xx'}}{\varepsilon_{x,\mathrm{rms}}} \right) = 1 \ . \tag{9}$$

Reordering the terms in Eq. (8) we obtain the definition of *rms emittance* in terms of the second moments of the distribution:

$$\varepsilon_{\mathrm{rms}} = \sqrt{\sigma_x^2 \sigma_{x'}^2 - \sigma_{xx'}^2} = \sqrt{\left(\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2 \right)} \ , \tag{10}$$

where we omit, from now on, the subscript x in the emittance notation: $\varepsilon_{\mathrm{rms}} = \varepsilon_{x,\mathrm{rms}}$. The rms emittance tells us some important information about phase space distributions under the effect of linear or non-linear forces acting on the beam. Consider, for example, an idealized particle distribution in phase space that lies on some line that passes through the origin, as illustrated in Fig. 5.



(a) (b)

Fig. 5: Phase space distributions under the effect of (a) linear or (b) non-linear forces acting on the beam

Assuming a generic correlation of the type $x' = Cx^n$ and computing the rms emittance according to Eq. (10) we have

$$\varepsilon^2_{\mathrm{rms}} = C\sqrt{\langle x^2\rangle\langle x^{2n}\rangle - \langle x^{n+1}\rangle^2} \quad \begin{cases} n=1 \Rightarrow \varepsilon_{\mathrm{rms}} = 0 \\ n>1 \Rightarrow \varepsilon_{\mathrm{rms}} \neq 0 \end{cases} . \tag{11}$$

When $n = 1$, the line is straight and the rms emittance is $\varepsilon_{\mathrm{rms}} = 0$. When $n > 1$ the relationship is non-linear, the line in phase space is curved, and the rms emittance is, in general, not zero. Both distributions have zero area. Therefore, we conclude that even when the phase space area is zero, if the distribution is lying on a curved line, its rms emittance is not zero. The rms emittance depends not only on the area occupied by the beam in phase space, but also on distortions produced by non-linear forces.

If the beam is subject to acceleration, it is more convenient to use the rms normalized emittance, for which the transverse momentum $p_x = p_z x' = m_0 c\beta\gamma x'$ is used instead of the divergence:

$$\varepsilon_{n,\mathrm{rms}} = \frac{1}{m_0 c}\sqrt{\sigma_x^2\sigma_{p_x}^2 - \sigma_{xp_x}{}^2} = \frac{1}{m_0 c}\sqrt{\left(\langle x^2\rangle\langle p_x^2\rangle - \langle xp_x\rangle^2\right)} = \sqrt{\left(\langle x^2\rangle\langle(\beta\gamma x')^2\rangle - \langle x\beta\gamma x'\rangle^2\right)} . \tag{12}$$

The reason for introducing a normalized emittance is that the divergences of the particles $x' = p_x/p$ are reduced during acceleration as p increases. Thus, acceleration reduces the un-normalized emittance, but does not affect the normalized emittance.

It is interesting to estimate the fundamental limit of the beam emittance that is set by quantum mechanics on the knowledge of the two conjugate variables (x, p_x). The state of a particle is actually not exactly represented by a point, but by a small uncertainty volume of the order of $\hbar^3$ in the 6D phase space. According to the Heisenberg uncertainty relation $\sigma_x\sigma_{p_x} \geq \frac{}{2}$ one gets from Eq. (12) $\varepsilon^{QM}_{n,rms} \geq \frac{1}{2}\frac{}{m_o c} = \frac{_c}{2}$, where $_c$ is the reduced Compton wavelength. For electrons it gives: $\varepsilon^{QM}_{n,rms} \geq 1.9\times10^{-13} m$.

In the classical limit we see also from Eq. (12) that the single particle emittance is zero.

Assuming a small energy spread within the beam, the normalized and un-normalized emittances can be related by the approximated relation $\langle\beta\gamma\rangle\varepsilon_{rms}$. This approximation, which is often used in conventional accelerators, may be strongly misleading when adopted for describing beams with significant energy spread, like those currently produced by plasma accelerators. A more careful analysis is reported next [8].

When the correlations between the energy and transverse positions are negligible (as in a drift without collective effects), Eq. (12) can be written as

$$\varepsilon_{n,\mathrm{rms}}^2 = \langle\beta^2\gamma^2\rangle\langle x^2\rangle\langle x'^2\rangle - \langle\beta\gamma\rangle^2\langle xx'\rangle^2 \ . \tag{13}$$

Consider now the definition of relative energy spread

$$\sigma_\gamma^2 = \frac{\langle\beta^2\gamma^2\rangle - \langle\beta\gamma\rangle^2}{\langle\beta\gamma\rangle^2}$$

which can be inserted into Eq. (13) to give

$$\varepsilon_{n,\mathrm{rms}}^2 = \langle\beta^2\gamma^2\rangle\sigma_\gamma^2\langle x^2\rangle\langle x'^2\rangle + \langle\beta\gamma\rangle^2\left(\langle x^2\rangle\langle x'^2\rangle - \langle xx'\rangle^2\right) . \tag{14}$$

Assuming relativistic particles ($\beta = 1$), we get

$$\varepsilon_{n,\mathrm{rms}}^2 = \langle\gamma^2\rangle\left(\sigma_\gamma^2\sigma_x^2\sigma_{x'}^2 + \varepsilon_{\mathrm{rms}}^2\right) . \tag{15}$$

If the first term in the parentheses is negligible, we find the conventional approximation of the normalized emittance as $\langle\gamma\rangle\varepsilon_{\mathrm{rms}}$. For a conventional accelerator, this might generally be the case. Considering, for example, beam parameters for the SPARC_LAB photoinjector [9]: at 5 MeV the ratio between the first and the second term is ~10^{-3}; while at 150 MeV it is ~10^{-5}. Conversely, using typical beam parameters at the plasma–vacuum interface, the first term is of the same order of magnitude as for conventional accelerators at low energies; however, owing to the rapid increase of the bunch size outside the plasma ($\sigma_{x'}$ ~ mrad) and the large energy spread ($\sigma_\gamma > 1\%$), it becomes predominant compared with the second term after a drift of a few millimetres. *Therefore, the use of approximated formulas when measuring the normalized emittance of plasma accelerated particle beams is inappropriate* [10].

4 The root mean square envelope equation

We are now interested in following the evolution of the particle distribution during beam transport and acceleration. One can use the collective variable defined in Eq. (6), the second moment of the distribution termed the rms beam envelope, to derive a differential equation suitable for describing the rms beam envelope dynamics [11]. To this end, let us compute the first and second derivative of σ_x [4]:

$$\begin{aligned}
\frac{\mathrm{d}\sigma_x}{\mathrm{d}z} &= \frac{\mathrm{d}}{\mathrm{d}z}\sqrt{\langle x^2\rangle} = \frac{1}{2\sigma_x}\frac{\mathrm{d}}{\mathrm{d}z}\langle x^2\rangle = \frac{1}{2\sigma_x}2\langle xx'\rangle = \frac{\sigma_{xx'}}{\sigma_x} \\
\frac{\mathrm{d}^2\sigma_x}{\mathrm{d}z^2} &= \frac{\mathrm{d}}{\mathrm{d}z}\frac{\sigma_{xx'}}{\sigma_x} = \frac{1}{\sigma_x}\frac{\mathrm{d}\sigma_{xx'}}{\mathrm{d}z} - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{1}{\sigma_x}\left(\langle x'^2\rangle + \langle xx'\rangle\right) - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{\sigma_{x'}^2 + \langle xx''\rangle}{\sigma_x} - \frac{\sigma_{xx'}^2}{\sigma_x^3}
\end{aligned} . \tag{16}$$

Rearranging the second derivative in Eq. (16), we obtain a second-order non-linear differential equation for the beam envelope evolution,

$$\sigma_x'' = \frac{\sigma_x^2 \sigma_{x'}^2 - \sigma_{xx'}^2}{\sigma_x^3} + \frac{\langle xx'' \rangle}{\sigma_x} \;, \tag{17}$$

or, in a more convenient form, using the rms emittance definition Eq. (10),

$$\sigma_x'' - \frac{1}{\sigma_x}\langle xx'' \rangle = \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_x^3} \;. \tag{18}$$

In Eq. (18), the emittance term can be interpreted physically as an outward pressure on the beam envelope produced by the rms spread in trajectory angle, which is parameterized by the rms emittance.

Let us now consider, for example, the simple case with $\langle xx'' \rangle = 0$, describing a beam drifting in free space. The envelope equation reduces to

$$\sigma_x^3 \sigma_x'' = \varepsilon_{\mathrm{rms}}^2 \;. \tag{19}$$

With initial conditions σ_0, σ'_0 at z_0, depending on the upstream transport channel, Eq. (19) has a hyperbolic solution:

$$\sigma(z) = \sqrt{\left(\sigma_0 + \sigma_0'(z - z_0)\right)^2 + \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_0^2}(z - z_0)^2} \;. \tag{20}$$

Considering the case of a beam at waist ($\langle xx' \rangle = 0$) with $\sigma'_0 = 0$, using Eq. (5), the solution Eq. (20) is often written in terms of the β function as

$$\sigma(z) = \sigma_0 \sqrt{1 + \left(\frac{z - z_0}{\beta_{\mathrm{w}}}\right)^2} \;. \tag{21}$$

This relation indicates that without any external focusing element the beam envelope increases from the beam waist by a factor √2 with a characteristic length $\beta_w = \sigma_0^2 / \varepsilon_{\mathrm{rms}}$, as shown in Fig. 6.

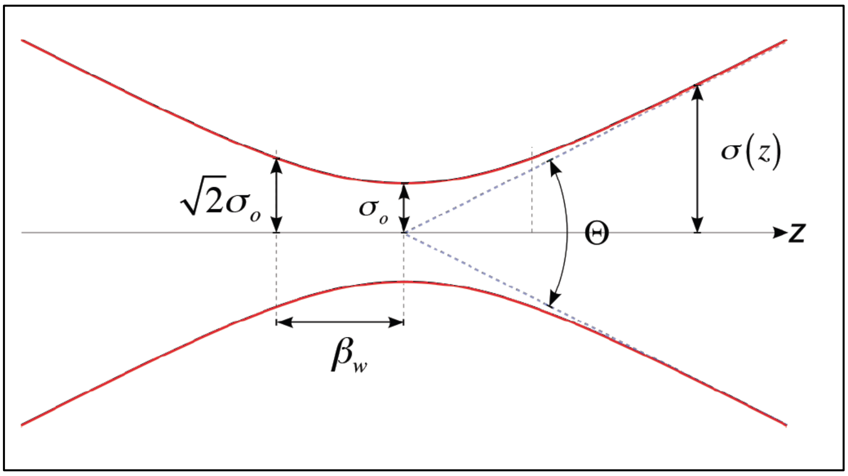


Fig. 6: Schematic representation of the beam envelope behaviour near the beam waist

At the waist, the relation $\varepsilon^2_{\mathrm{rms}} = \sigma^2_{0,x}\sigma^2_{0,x'}$ also holds, which can be inserted into Eq. (20) to give $\sigma^2_x(z) = \sigma^2_{0x'}(z - z_0)^2$. Under this condition, Eq. (15) can be written as

$$\varepsilon_{n,\mathrm{rms}}^2(z)=\langle\gamma^2\rangle\left(\sigma_\gamma^2\sigma_{x'}^4(z-z_0)^2+\varepsilon_{\mathrm{rms}}^2\right),$$

showing that beams with large energy spread and divergence undergo a significant normalized emittance growth even in a drift of length $(z-z_0)$ [8, 12].

Notice also that the solution Eq. (21) is exactly analogous to that of a Gaussian light beam for which the beam width $w = 2\sigma_{\mathrm{ph}}$ increases away from its minimum value at the waist w_0 with characteristic length $Z_{\mathrm{R}}=\pi w_0^2/\lambda$ (Rayleigh length) [4]. This analogy suggests that we can identify an effective emittance of a photon beam as $\varepsilon_{\mathrm{ph}}=\lambda/4\pi$.

For the effective transport of a beam with finite emittance, it is mandatory to make use of some external force providing beam confinement in the transport or accelerating line. The term $\langle xx''\rangle$ accounts for external forces when we know x'', given by the single particle equation of motion:

$$\frac{\mathrm{d}p_x}{\mathrm{d}t}=F_x \quad (22)$$

Under the paraxial approximation $p_x << p = \beta\gamma mc$, the transverse momentum p_x can be written as $p_x = px' = \beta\gamma m_0cx'$, so that

$$\frac{\mathrm{d}p_x}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}(px')=\beta c\frac{\mathrm{d}}{\mathrm{d}z}(px')=F_x\ , \quad (23)$$

and the transverse acceleration results in

$$x''=-\frac{p'}{p}x'+\frac{F_x}{\beta cp} \quad (24)$$

It follows that

$$\langle xx''\rangle=-\frac{p'}{p}\langle xx'\rangle+\frac{\langle xF_x\rangle}{\beta cp}=\frac{p'}{p}\sigma_{xx'}+\frac{\langle xF_x\rangle}{\beta cp} \quad (25)$$

Inserting Eq. (25) into Eq. (18) and recalling Eq. (16), $\sigma_x'=\sigma_{xx'}/\sigma_x$, the complete rms envelope equation is:

$$\sigma_x''+\frac{p'}{p}\sigma_x'-\frac{1}{\sigma_x}\frac{\langle xF_x\rangle}{\beta cp}=\frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3}\ , \quad (26)$$

where we have included the normalized emittance $\varepsilon_{n,\mathrm{rms}} = \gamma\varepsilon_{\mathrm{rms}}$. Notice that the effect of longitudinal accelerations appears in the rms envelope equation as an oscillation damping term, called ‘adiabatic damping’, proportional to p'/p. The term $\langle xF_x\rangle$ represents the moment of any external transverse force acting on the beam, such as that produced by a focusing magnetic channel.

5 External forces

Let’s now consider the case of an external linear force acting on the beam in the form $F_x=\mp kx$. It can be focusing or defocusing, according to the sign. The moment of the force is

$$\left\langle xF_x \right\rangle = \mp k \left\langle x^2 \right\rangle = \mp k \sigma_x^2 \tag{27}$$

and the envelope equation becomes

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' \mp k_{\text{ext}}^2 \sigma_x = \frac{\varepsilon_{n,\text{rms}}^2}{\gamma^2 \sigma_x^3} , \tag{28}$$

where we have explicitly used the momentum definition $p = \gamma mc$ for a relativistic particle with $\beta \approx 1$ and defined the wavenumber

$$k_{\text{ext}}^2 = \frac{k}{\gamma m_0 c^2} .$$

Typical focusing elements are quadrupoles and solenoids [3]. The magnetic quadrupole field is given in Cartesian coordinates by

$$\begin{cases} B_x = B_0 \dfrac{y}{d} = B_0' y \\ B_y = B_0 \dfrac{x}{d} = B_0' x \end{cases} , \tag{29}$$

where d is the pole distance and B_0' is the field gradient. The force acting on the beam is $\vec{F}_\perp = qv_z B_0' \left(y\hat{j} - x\hat{i} \right)$ and, when B_0 is positive, is focusing in the x direction and defocusing in the y direction. The focusing strength is

$$k_{\text{quad}} = \frac{qB_0'}{\gamma m_0 c} = k_{\text{ext}}^2 .$$

In a solenoid the focusing strength is given by

$$k_{\text{sol}} = \left(\frac{qB_0}{2\gamma m_0 c} \right)^2 = k_{\text{ext}}^2 .$$

Notice that the solenoid is always focusing in both directions, an important property when the cylindrical symmetry of the beam must be preserved. However, being a second-order quantity in γ, it is more effective at low energy.

It is interesting to consider the case of a uniform focusing channel without acceleration described by the rms envelope equation

$$\sigma_x'' + k_{\text{ext}}^2 \sigma_x = \frac{\varepsilon_{\text{rms}}^2}{\sigma_x^3} . \tag{30}$$

By substituting $\sigma_x = \sqrt{\beta_x \varepsilon_{\text{rms}}}$ into Eq. (30) one obtains an equation for the 'betatron function' $\beta_x(z)$ that is independent of the emittance term:

$$\beta_x'' + 2k_{\text{ext}}^2 \beta_x = \frac{2}{\beta_x} + \frac{\beta_x'^2}{2\beta_x} . \tag{31}$$

Equation (31) contains just the transport channel focusing strength and, being independent of the beam parameters, suggests that the meaning of the betatron function is to account for the transport line characteristic. The betatron function reflects exterior forces from focusing magnets and is highly dependent on the particular arrangement of the quadrupole magnets. The equilibrium, or matched, solution of Eq. (31) is given by $\beta_{\text{eq}} = \dfrac{1}{k_{\text{ext}}} = \dfrac{\lambda_\beta}{2\pi}$, as can be easily verified. This result shows that the matched β_x function is simply the inverse of the focusing wavenumber or, equivalently, is proportional to the 'betatron wavelength' λ_β. The corresponding envelope equilibrium condition, i.e., a stationary solution of Eq. (30), is given by: $\sigma_{\text{eq},x} = \sqrt{\dfrac{\varepsilon_{\text{rms}}}{k_{\text{ex}}}}$.

In analogy with the kinetic theory of gases we can define the beam temperature in a transverse direction at equilibrium and without correlations as

$$k_{\text{B}} T_{\text{beam},x} = \gamma m_0 \left\langle v_x^2 \right\rangle = \frac{\sigma_{p_x}^2}{\gamma m_0} = m_0 c^2 \frac{\varepsilon_{n,\text{rms}}^2}{\gamma \sigma_{\text{eq},x}^2} = \gamma m_0 \beta^2 c^2 \frac{\varepsilon_{\text{rms}}}{\beta_{\text{eq},x}},$$

where k_{B} is the Boltzmann constant and we have used Eq. (12), showing that the conditions for a cold beam are typically: low emittance, low energy, high betatron function.

By means of the beam temperature concept one can also define the beam emittance at the source called the thermal emittance. Assuming that electrons are in equilibrium with the cathode temperature $T_{\text{c}} = T_{\text{beam}}$ and $\gamma=1$, the thermal emittance is given by $\varepsilon_{\text{th,rms}}^{\text{cat}} = \sigma_x \sqrt{\dfrac{k_{\text{B}} T_{\text{c}}}{m_0 c^2}}$ which, per unit rms spot size at the cathode, is $\varepsilon_{\text{th,rms}} = 0.3$ μm/mm at $T_{\text{c}} = 2500$ K. For comparison, in a photocathode illuminated by a laser pulse with photon energy $\hbar\omega$ the expression for the variance of the transverse momentum of the emitted electrons is given by $\sigma_{p_x} = \sqrt{\dfrac{m_0}{3}(\ \omega - \phi_{\text{eff}})}$, where $\phi_{\text{eff}} = \phi_{\text{w}} - \phi_{\text{Schottky}}$, ϕ_{w} being the material work function and ϕ_{Schottky} the Schottky work function [19]. The corresponding thermal emittance is $\varepsilon_{\text{th,rms}}^{\text{ph}} = \sigma_x \sqrt{\dfrac{\omega - \phi_{\text{eff}}}{3 m_0 c^2}}$ that, with the typical parameters of a Copper photocathode illuminated by a UV laser, gives a thermal emittance per unit spot size of about 0.5 μm/mm.

6 Space charge forces

Another important force acting on the beam is the one produced by the beam itself due to the internal Coulomb forces. The net effect of the Coulomb interaction in a multiparticle system can be classified into two regimes [3]:

i) *collisional regime*, dominated by binary collisions caused by close particle encounters;

ii) *collective regime* or *space charge regime*, dominated by the self-field produced by the particles' distribution, which varies appreciably only over large distances compared with the average separation of the particles.

A measure for the relative importance of collisional versus collective effects in a beam with particle density n is the relativistic *Debye length*,

$$\lambda_{\text{D}} = \sqrt{\frac{\varepsilon_0 \gamma^2 k_{\text{B}} T_{\text{b}}}{e^2 n}} \ . \tag{32}$$

As long as the Debye length remains small compared with the particle bunch transverse size, the beam is in the space charge dominated regime and is not sensitive to binary collisions. Smooth functions for the charge and field distributions can be used in this case, and the space charge force can be treated as an external applied force. The space charge field can be separated into linear and non-linear terms as a function of displacement from the beam axis. The linear space charge term defocuses the beam and leads to an increase in beam size. The non-linear space charge terms also increase the rms emittance by distorting the phase space distribution. Under the paraxial approximation of particle motion, we can consider the linear component alone. We shall see next that the linear component of the space charge field can also induce emittance growth when correlations along the bunch are taken into account.

For a bunched beam of uniform charge distribution in a cylinder of radius R and length L, carrying a current $\hat{I}$ and moving with longitudinal velocity $v_z = \beta c$, the linear component of the longitudinal and transverse space charge field are given approximately by [13]

$$E_z(\zeta) = \frac{\hat{I}L}{2\pi\varepsilon_0 R^2 \beta c} h(\zeta) \quad , \tag{33}$$

$$E_{\mathrm{r}}(r,\zeta) = \frac{\hat{I}r}{2\pi\varepsilon_0 R^2 \beta c} g(\zeta) \quad . \tag{34}$$

The field form factor is described by the functions:

$$h(\zeta) = \sqrt{A^2 + (1-\zeta)^2} - \sqrt{A^2 + \zeta^2} - |1-\zeta| + |\zeta| \tag{35}$$

$$g(\zeta) = \frac{(1-\zeta)}{2\sqrt{A^2 + (1-\zeta)^2}} + \frac{\zeta}{2\sqrt{A^2 + \zeta^2}} \quad , \tag{36}$$

where $\zeta = z/L$ is the normalized longitudinal coordinate along the bunch, $\zeta = 0$ being the bunch tail, and $A = R/\gamma L$ is the beam aspect ratio. The field form factors account for the variation of the fields along the bunch and outside the bunch for $\zeta<0$ *and* $\zeta>L$. As γ increases, $g(\zeta) \to 1$ and $h(\zeta) \to 0$, thus showing that space charge fields mainly affect transverse beam dynamics. It shows also that an energy increase corresponds to a bunch lengthening in the moving frame $L' = \gamma L$, leading to a vanishing longitudinal field component, as in the case of a continuous beam in the laboratory frame.

To evaluate the force acting on the beam, one must also account for the azimuthal magnetic field associated with the beam current, which, in cylindrical symmetry, is given by

$$B_\vartheta = \frac{\beta}{c} E_r \ .$$

Thus, the Lorentz force acting on each single particle is given by

$$F_r = e(E_r - \beta c B_\vartheta) = e(1-\beta^2)E_r = \frac{eE_r}{\gamma^2} \quad . \tag{37}$$

The attractive magnetic force, which becomes significant at high velocities, tends to compensate for the repulsive electric force. Therefore, space charge defocusing is primarily a non-relativistic effect and decreases as γ^{-2}.

To include space charge forces in the envelope equation, let us start by writing the space charge forces produced by the previous fields in Cartesian coordinates:

$$F_x = \frac{e\hat{I}x}{8\pi\gamma^2\varepsilon_0\sigma_x^2\beta c}g\left(\zeta\right) \quad . \tag{38}$$

Then, computing the moment of the force, we need

$$x'' = \frac{F_x}{\beta cp} = \frac{eIx}{8\pi\varepsilon_0\gamma^3 m_0\beta^3 c^3\sigma_x^2} = \frac{k_{\mathrm{sc}}\left(\zeta\right)}{\left(\beta\gamma\right)^3\sigma_x^2} \tag{39}$$

where we have introduced the generalized beam perveance,

$$k_{\mathrm{sc}}\left(\zeta\right) = \frac{\hat{I}}{2I_{\mathrm{A}}}g\left(\zeta\right) \tag{40}$$

where $I_{\mathrm{A}} = 4\pi\varepsilon_0 m_0 c^3/e = 17$ kA is the Alfvén current for electrons. Notice that in this case the perveance in Eq. (40) explicitly depends on the slice coordinate ζ. We can now calculate the term that enters the envelope equation for a relativistic beam,

$$\left\langle xx''\right\rangle = \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x^2}\left\langle x^2\right\rangle = \frac{k_{\mathrm{sc}}}{\gamma^3} \ , \tag{41}$$

leading to the complete envelope equation

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' + k_{\mathrm{ext}}^2\sigma_x = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} + \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x} \quad . \tag{42}$$

From the envelope equation Eq. (42), we can identify two regimes of beam propagation: *space charge dominated* and *emittance dominated.* A beam is space charge dominated as long as the space charge collective forces are largely dominant over the emittance pressure. In this regime, the linear component of the space charge force produces a quasi-laminar propagation of the beam, as one can see by integrating one time Eq. (39) under the paraxial ray approximation x' 1.

A measure of the relative importance of space charge effects versus emittance pressure is given by the *laminarity parameter*, defined as the ratio between the space charge term and the emittance term:

$$\rho = \frac{\hat{I}}{2I_{\mathrm{A}}\gamma}\frac{\sigma^2}{\varepsilon_n^2} \quad . \tag{43}$$

When ρ greatly exceeds unity, the beam behaves as a laminar flow (all beam particles move on trajectories that do not cross), and transport and acceleration require a careful tuning of focusing and accelerating elements to keep laminarity. Correlated emittance growth is typical in this regime, which can be made reversible if proper beam matching conditions are fulfilled, as discussed next. When $\rho < 1$, the beam is emittance dominated (thermal regime) and space charge effects can be neglected. The transition to the thermal regime occurs when $\rho \approx 1$, corresponding to the transition energy

$$\gamma_{\mathrm{tr}} = \frac{\hat{I}}{2I_{\mathrm{A}}}\frac{\sigma^2}{\varepsilon_n^2} \quad . \tag{44}$$

For example, a beam with $\hat{I} = 100$ A, $\varepsilon_n = 1$ µm, and $\sigma = 300$ µm is leaving the space charge dominated regime and is entering the thermal regime at the transition energy of 131 MeV. From this example, one may conclude that the space charge dominated regime is typical of low-energy beams. Actually, for such applications as linac-driven free electron lasers, peak currents exceeding kA are required. Space charge effects may recur if bunch compressors are active at higher energies and a new energy threshold with higher $\hat{I}$ must be considered.

7 Correlated emittance oscillations

When longitudinal correlations within the bunch are important, like that induced by space charge effects, beam envelope evolution is generally dependent also on the coordinate along the bunch ζ. In this case, the bunch should be considered as an ensemble of n longitudinal slices of envelope $\sigma_s(z,\zeta)$, whose evolution can be computed from n slice envelope equations equivalent to Eq. (42), provided that the bunch parameters refer to each single slice: γ_s, γ'_s, $k_{sc,s} = k_{sc}g(\zeta)$. Correlations within the bunch may cause emittance oscillations that can be evaluated, once an analytical or numerical solution [13] of the slice envelope equation is known, by using the following correlated emittance definition:

$$\varepsilon_{\mathrm{rms,cor}} = \sqrt{\langle\sigma_s^2\rangle\langle\sigma_s'^2\rangle - \langle\sigma_s\sigma_s'\rangle^2} \ , \tag{45}$$

where the average is performed over the entire slice ensemble, assuming uniform charge distribution within each slice. In the simplest case of a two-slice model, the previous definition reduces to

$$\varepsilon_{\mathrm{rms,cor}} = |\sigma_1\sigma_2' - \sigma_2\sigma_1'| \ , \tag{46}$$

which represents a simple and useful formula for an estimation of the emittance scaling [14].

The total normalized rms emittance is given by the superposition of the correlated and uncorrelated terms as

$$\varepsilon_{\mathrm{rms,cor}} = \langle\gamma\rangle\sqrt{\varepsilon_{\mathrm{rms}}^2 + \varepsilon_{\mathrm{rms,cor}}^2} \ . \tag{47}$$

An interesting example to consider here, showing the consequences of non-perfect beam matching, is the propagation of a beam in the space charge dominated regime nearly matched to an external focusing channel, as illustrated in Fig. 7. To simplify our computations, we can neglect acceleration, as in the case of a simple beam transport line made by a long solenoid ($k^2_{\mathrm{ext}} = k_{\mathrm{sol}}$). The envelope equation for each slice, indicated as σ_s, reduces to

$$\sigma_s'' + k_{\mathrm{ext}}^2\sigma_s = \frac{k_{\mathrm{sc,s}}}{\gamma^3\sigma_s} \ . \tag{48}$$

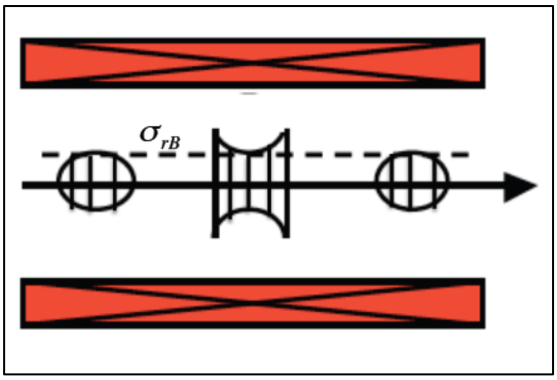


Fig. 7: Schematic representation of a nearly matched beam in a long solenoid. The dashed line represents the reference slice envelope matched to the Brillouin flow condition. The other slice envelopes are oscillating around the equilibrium solution.

A stationary solution corresponding to slice propagation with constant envelope, called *Brillouin flow*, is given by

$$\sigma_{r,B} = \frac{1}{k_{ext}} \sqrt{\frac{\hat{I}g(\zeta)}{2\gamma^3 I_A}} , \tag{49}$$

where the local dependence of the current $\hat{I}_s = \hat{I}g(\zeta)$ within the bunch has been explicitly indicated. This solution represents the matching conditions for which the external focusing completely balances the internal space charge force. Unfortunately, since k_{ext} has a slice-independent constant value, the Brillouin matching condition is different for each slice and usually cannot be achieved at the same time for all of the bunch slices. Assuming that there is a reference slice perfectly matched (49) with an envelope $\sigma_{r,B}$ and negligible beam energy spread, the matching condition for the other slices can be written as:

$$\sigma_{sB} = \sigma_{rB} + \frac{\sigma_{rB}}{2}\left(\frac{\delta I_s}{\hat{I}}\right) , \tag{50}$$

with respect to the reference slice. Considering a slice with a small perturbation δ_s with respect to its own equilibrium Eq. (50) in the form

$$\sigma_s = \sigma_{s,B} + \delta_s , \tag{51}$$

and substituting into Eq. (48), we can obtain a linearized equation for the slice offset

$$\delta_s'' + 2k_{ext}^2 \delta_s = 0 , \tag{52}$$

which has a solution given by

$$\delta_s = \delta_0 \cos\left(\sqrt{2} k_{ext} z\right) , \tag{53}$$

where $\delta_0 = \sigma_{so} - \sigma_{sB}$ is the amplitude of the initial slice mismatch, which we assume, for convenience, is the same for all slices. Inserting Eq. (53) into Eq. (51) we get the perturbed solution:

$$\sigma_s = \sigma_{s,B} + \delta_0 \cos\left(\sqrt{2} k_{ext} z\right) . \tag{54}$$

Equation (54) shows that slice envelopes oscillate together around the equilibrium solution with the same frequency for all slices ($\sqrt{2}k_{ext}$, often called the plasma frequency) dependent only on the external focusing forces. This solution represents a collective behaviour of the bunch, similar to that of the electrons subject to the restoring force of ions in a plasma. Using the two-slice model and Eq. (54), the emittance evolution Eq. (46) results in

$$\varepsilon_{rms,cor} = \frac{1}{4} k_{ext} \sigma_{rB} \left| \frac{\Delta I}{\hat{I}} \delta_0 \sin\left(k_{ext} z\right) \right| , \tag{55}$$

where $\Delta I = \hat{I}_1 - \hat{I}_2$. Notice that, in this simple case, envelope oscillations of the mismatched slices induce correlated emittance oscillations that periodically return to zero, showing the reversible nature of the correlated emittance growth. It is, in fact, the coupling between transverse and longitudinal motion induced by the space charge fields that allows reversibility. With proper tuning of the transport line length or of the focusing field, one can compensate for the transverse emittance growth.

At first, it may seem surprising that a beam with a single charge species can exhibit plasma oscillations, which are characteristic of plasmas composed of two-charge species. However, the effect of the external focusing force can play the role of the other charge species, providing the necessary restoring force that is the cause of such collective oscillations, as shown in Fig. 8. The beam can actually be considered as a single-component, relativistic, cold plasma.

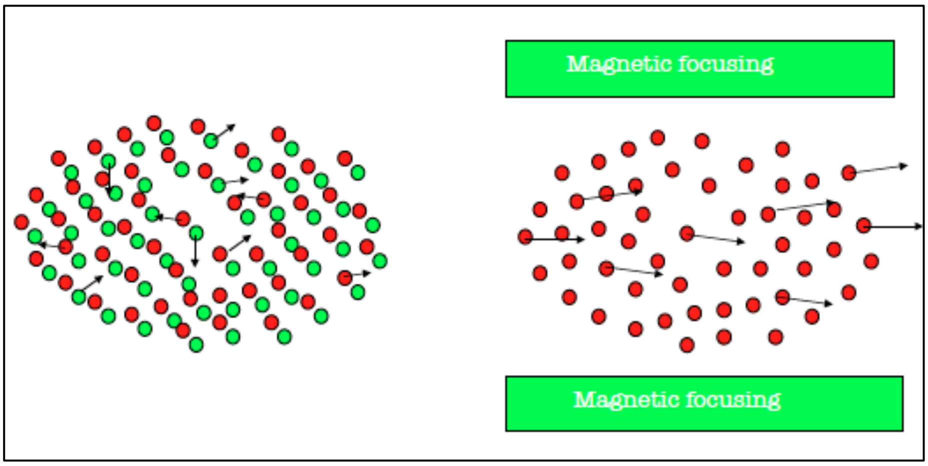


Fig. 8: The restoring force produced by the ions (green dots) in a plasma may cause electron (red dots) oscillations around the equilibrium distribution. In a similar way, the restoring force produced by a magnetic field may cause beam envelope oscillations around the matched envelope equilibrium.

It is important to bear in mind that beams in linacs are also different from plasmas in some important respects [5]. One is that beam transit time through a linac is too short for the beam to reach thermal equilibrium. Also, unlike a plasma, the Debye length of the beam may be larger than, or comparable to, the beam radius, so shielding effects may be incomplete.

8 Matching conditions in a radiofrequency linac

In order to prevent space charge induced emittance growth in a radiofrequency (rf) linac, as in the case of a high brightness photoinjector, and to drive a smooth transition from the space charge to the thermal regime, space charge induced emittance oscillations have to be damped along the linac in such a way that an emittance minimum is obtained at the transition energy Eq. (44). To this end the beam has to be properly matched to the accelerating sections with a Brillouin like flow in order to keep under control emittance oscillations that in this case are provided by the ponderomotive rf focusing force [2] acting in the rf structures. In some case rf focusing is too weak to provide sufficient beam containment. A long solenoid around the accelerating structure is a convenient replacement to provide the necessary focusing.

The matching conditions for a beam subject to acceleration (assuming $\gamma(z)=\gamma_o+\gamma' z$ and $\gamma''=0$) can be obtained following the previous example (Brillouin flow). This process can be described using the envelope equation (42) for a generic slice σ_s with external focusing provided by $k_{\text{ext}}^2=k_{\text{sol}}+k_{\text{rf}}^2$, where $k_{rf}^2=\frac{\eta}{8}\left(\frac{\gamma'}{\gamma}\right)^2$ and $\gamma'=\frac{eE_{acc}}{mc^2}$. The quantity η is a measure of the higher spatial harmonic amplitudes of the rf wave and it is generally quite close to unity in standing wave (SW) structures and close to 0 in travelling wave (TW) structures [15].

Being now γ(*z*) a time-dependent function, a stationary solution of Eq. (42) cannot be found by simply looking for a constant envelope solution. A possible way to find an 'equilibrium' solution is described hereafter. By substituting the reduced variable $\hat{\sigma}=\sqrt{\gamma}\sigma_s$ [16] in the envelope equation (42) we obtain

$$\hat{\sigma}''+\hat{k}_{ext}^2\hat{\sigma}=\frac{\hat{K}_{sc}}{\hat{\sigma}}+\frac{\varepsilon_n^2}{\hat{\sigma}^3} \tag{56}$$

with the scaled parameters $\hat{k}_{\text{ext}}^2 = k_{\text{ext}}^2 + \frac{1}{4}\left(\frac{\gamma'}{\gamma}\right)^2 = k_{\text{sol}} + \frac{1}{4}\left(\frac{\gamma'}{\gamma}\right)^2\left(1+\frac{\eta}{2}\right)$ and $\hat{K}_{sc} = k_{sc}/\gamma^2$. Eq. (56) is equivalent to Eq. (42) but the damping term has disappeared and the $\hat{k}_{ext}^2$ and $\hat{K}_{sc}$ parameters have the same γ^{-2} dependence. In the space charge regime the emittance term can be neglected in Eq. (56) and an equilibrium solutions in the reduced variables (called the 'invariant envelope' in the literature [2]) is given by $\hat{\sigma}_{sc} = \frac{\sqrt{\hat{K}_{sc}}}{\hat{k}_{ext}}$, corresponding to the matching conditions for the beam envelope:

$$\sigma_{sc} = \sqrt{\frac{2\hat{I}}{\gamma I_A\left(\Theta^2 + \gamma'^2\left(\frac{\eta}{2}+1\right)\right)}} \qquad \text{for } \rho > 1 \tag{57}$$

where $\Theta = \frac{eB}{mc}$.

The expression for the emittance oscillation in the space charge dominated regime, i.e. when $\gamma < \gamma_{tr}$, can be obtained from Eq. (55) using reduced variables and results:

$$\varepsilon_n = \frac{1}{\gamma}\sqrt{\frac{\hat{I}}{34I_A}}\left|\frac{\Delta I}{\hat{I}}\delta_o \sin\left(\frac{\left(\Theta^2 + \gamma'^2\left(\frac{\eta}{2}+1\right)\right)^{1/2}}{2\gamma}z\right)\right| \quad . \tag{58}$$

Before the transition energy is achieved the emittance performs damped oscillations with wavelength depending on the external fields and with amplitude depending on the current profile. A careful tuning of the external fields and bunch charge profile can minimize the value of the emittance at the injector extraction. A successful application of the emittance compensation technique can be seen in [17, 18].

When the beam enters in the thermal regime an equilibrium solution can be found directly from Eq. (42) neglecting the space charge term. The result is

$$\sigma_{\text{th}} = \sqrt{\frac{2\varepsilon_n}{\left(\Theta^2 + \frac{\eta}{2}\gamma'^2\right)^{1/2}}} \qquad \text{for } \rho < 1 \tag{59}$$

and no correlated emittance oscillations are expected. Note also that Eq. (57) scales like $\gamma^{-1/2}$ while Eq. (59) is independent of γ.

Acknowledgements

I wish to thank A. Cianchi, E. Chiadroni, J.B. Rosenzweig, A.R. Rossi, A. Bacci, and L. Serafini for the many helpful discussions and suggestions.

References

[1] T. Shintake, Review of the worldwide SASE FEL development, Proc. 22nd Particle Accelerator Conf., Albuquerque, NM, 2007 (IEEE, New York, 2007), p. 89. https://doi.org/10.1109/PAC.2007.4440331

[2] L. Serafini and J.B. Rosenzweig, *Phys. Rev. E* **55** (1997) 7565. https://doi.org/10.1103/PhysRevE.55.7565

[3] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York, 1994). https://doi.org/10.1002/9783527617623

[4] J.B. Rosenzweig, *Fundamentals of Beam Physics* (Oxford University Press, Oxford, 2003). https://doi.org/10.1093/acprof:oso/9780198525547.001.0001

[5] T. Wangler, *Principles of RF Linear Accelerators* (Wiley, New York, 1998). https://doi.org/10.1002/9783527618408

[6] S. Humphries, *Charged Particle Beams* (Wiley, New York, 2002).

[7] N. Pichoff, Beam dynamics basics in RF linacs, Proc. CERN Accelerator School: Small Accelerators, Zeegse, The Netherlands, 2005, CERN-2006-012. https://doi.org/10.5170/CERN-2006-012

[8] M. Migliorati *et al.*, *Phys. Rev. ST Accel. Beams* **16** (2013) 011302. https://doi.org/10.1103/PhysRevSTAB.16.011302

[9] M. Ferrario *et al.*, *Nucl. Instrum. Methods Phys. Res. B* **309** (2013) 183. https://doi.org/10.1016/j.nimb.2013.03.049

[10] A. Cianchi *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **720** (2013) 153. https://doi.org/10.1016/j.nima.2012.12.012

[11] F.J. Sacherer, *IEEE Trans. Nucl. Sci.* **NS-18** (1971) 1105. https://doi.org/10.1109/TNS.1971.4326293

[12] K. Floettmann, *Phys. Rev. ST Accel. Beams* **6** (2003) 034202. https://doi.org/10.1103/PhysRevSTAB.6.034202

[13] M. Ferrario *et al.*, *Int. J. Mod. Phys. A* **22** (2007) 4214. https://doi.org/10.1142/S0217751X07037779

[14] J. Buon, Beam phase space and emittance, Proc. CERN Accelerator School: 5th General Accelerator Physics Course, Jyvaskyla, Finland, 1992, CERN-94-01. https://doi.org/10.5170/CERN-1994-001

[15] J. B. Rosenzweig, L. Serafini, Phys. Rev. E 49, 1599 (1994). https://doi.org/10.1103/PhysRevE.49.1599

[16] C. Wang, Phys. Rev. E 74, 046502 (2006). https://doi.org/10.1103/PhysRevE.74.046502

[17] R. Akre et al., *Phys. Rev. ST Accel. Beams* 11, 030703 (2008). https://doi.org/10.1103/PhysRevSTAB.11.030703

[18] M. Ferrario et al., Phys. Rev. Lett. 99, 234801 (2007). https://doi.org/10.1103/PhysRevLett.99.234801

[19] D. H. Dowell et al., *Phys. Rev. ST Accel. Beams* **12**, 074201 (2009). https://doi.org/10.1103/PhysRevSTAB.12.074201

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Historical Survey of Free Electron Lasers

M. E. Couprie
Synchrotron SOLEIL, Saint-Aubin, France

Abstract

Free electron lasers are presently unique tuneable powerful lasers ranging from the infrared to the X-ray, serving for the exploration of light–matter interaction. They make use of a simple and elegant gain medium. The coherent radiation is generated using free electrons in a periodic permanent magnetic field generated by a so-called undulator as an amplification medium. The light–electron interaction in the undulator leads to a bunching process, setting in phase the electron emitters and establishing a longitudinal coherence. The amplification of the light wave takes place to the detriment of the kinetic energy of the electrons. First, the origins of the free electron laser, starting from the photon–matter interaction, the early times of synchrotron radiation and the undulator, the development of vacuum tubes, masers, and lasers are introduced. Then, motivated by the search of an exotic laser, the invention of the free electron conceptual idea laser is discussed, with a quantum approach followed by a classical one. The first low gain free electron laser experiments are then presented. Finally, some insight is given in the case of the high gain free electron lasers.

Keywords
Free electron laser; laser; undulator.

1 Introduction

In the early times, free electron lasers (FELs) appeared as exotic lasers. In 1990, C. Brau (ninth FEL Prize in 1996) introduced them as such: They “represent an altogether new and exciting class of coherent optical sources. Making use of a simple and elegant gain medium—an electron beam in a magnetic field—they have already demonstrated broad wavelength tuneability and excellent optical-beam quality. For the future they offer the possibility of generating the greatest focused power ever achieved by a laser. But even before this is achieved, the unique advantages of free electron lasers, especially their tuneability, will make them useful for a variety of important applications in science and medicine” [1]. Indeed, nowadays, they provide the most powerful lasers in the X-ray range, with ultra-short pulse duration. The present advent of tuneable X FELs with unpreceded intensities enables new investigation of matter with ultra-high intensities, ultra-short pulses, etc. It could be considered as a second revolution, following the invention of the laser, which led to the development of optical lasers which has changed our current life.

The FEL spontaneous emission corresponds to the undulator radiation emitted by the relativistic electrons. The electrons are not bound to nuclei in atoms and molecules, and vibrate at specific frequencies. In contrast, the FEL vibration frequency can be adjusted by changing the magnetic field or the energy of the electrons, resulting in a broad wavelength tuneability. Indeed, “The electrons in a free electron laser have the form of an electron beam in a vacuum, much like the beam in the picture tube of a television set except that the electrons have much higher energy and intensity. Electrons bound in atoms and molecules vibrate only at specific frequencies. Thus the laser light from conventional lasers, which make use of bound electrons, appears only at these specific frequencies. On the other hand, the electrons in a free electron laser are forced to vibrate by their passage through alternating magnetic field. Thus, the vibration frequency can be adjusted by altering the construction of the magnetic field or by

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.195

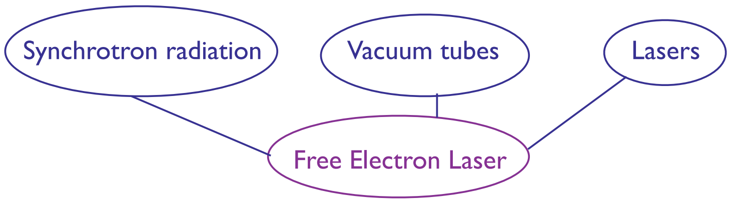


Fig. 1: Schematic presentation of the origins of the free electron laser

changing the speed of the electrons passing through the magnetic field. This changes the laser frequency or, equivalently, the wavelength". The gain process results from the energy exchange between the light and the electrons. The optical beam presents an excellent optical beam quality and can achieve high power.

This historical survey on the free electron laser aims at describing how the new ideas have emerged and how the field progressed. Some citations are taken from the original papers, to show how the understanding at a given time was. This report gathers also the main results which define major steps to my own understanding and personal experience. The work of the FEL Prize winners, listed in the Appendix, is mentioned. The progress of the FEL field is also closely related to technological advances, which cannot be discussed in detail but which are underlined when crucial. Drawing a history survey [2,3] on FEL implies some choices in the discussed items, which are necessarily subjective. I apologize in advance for all the important works which have not been cited.

First, the scientific context enabling the emergence of the FEL concept is explained: the development of vacuum tubes in the twentieth century, together with the invention of the laser in 1958 [4] and the first laser operation in 1960 [5]. Then, the discovery of the FEL [6] by J. M. J. Madey (1943–2016, first FEL Prize, 1988) which took place in such a context is reported. FELs combine specificities of synchrotron radiation, vacuum tubes, and lasers, as indicated in Fig. 1. The FEL was treated with a quantum approach, as stimulated Compton Back Scattering.

Then the first applied classical treatments to FELs are reported. A section describing the first experimental results follows, with some newly proposed concepts. The last section deals with high gain developments, including high gain FEL theory and the proposed concept of Self-Amplified Spontaneous Emission (SASE), the major experimental steps for SASE operation and some hints on further developments (seeding, echo).

2 The origins of the free electron laser

The scientific context favourable for the emergence of the free electron laser concept is hereafter described. The understanding of the interaction between light and matter was established at the beginning of the twentieth century. Then, the century knew an intense development of vacuum tubes accompanied by the discovery of the laser in the sixties. The emergence of the free electron laser concept benefited from the interplay between these two domains.

2.1 The photon–matter interaction processes

The laser concept relies on the prediction of energy enhancement by atom de-excitation by Albert Einstein (1879–1955, Nobel Prize in 1921) in 1917 [7] in the analysis of the black-body radiation, while absorption and spontaneous emission were the known light matter interactions at that time. The process was called later stimulated emission by J. Van Vleck in 1924 [8,9]. In the absorption case, a photon is absorbed and drives an atom to an excited state. The excited atom being unstable, it emits a spontaneous photon after a duration depending on the lifetime of the excited level. In the stimulated emission case, a photon is absorbed by an excited atom, which results in the emission of two photons with identical wavelength, direction, phase, polarization, while the atom returns to its fundamental state. Einstein was mainly interested by thermal radiation and exchanges of momentum in different process, but not specifically by the production of light by matter. Stimulated emission was seen as addition of photons to already existing photons, and not as the amplification of a monochromatic wave with conservation of its phase. The notion of light coherence, related to its undulatory properties, was not considered at that time.

2.2 The early times of synchrotron radiation

Synchrotron radiation, the electromagnetic radiation emitted by accelerated charged particles, is generally produced artificially in particle accelerators, but it can be also observed in astrophysics [10,11], such as in the Sun where hydrogen submitted to loops of magnetic fields emits visible light in the centre and on the edges in the X-ray domain.

Let's introduce some notations. Let's consider a laboratory region, without current, with a laboratory referential $(0, x, z, s)$ with the frame (x, z, s) [12], as shown in Fig. 2 where s is the longitudinal, x the horizontal, and z the vertical direction.

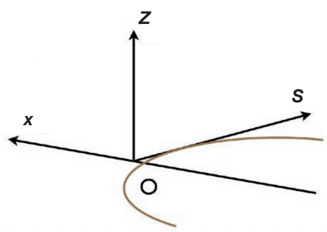


Fig. 2: Axis coordinates

Consider then a relativistic electron of energy E and velocity v with respect to the laboratory frame. Its relativistic factor γ is given by $\gamma = \frac{E}{m_o c^2}$ (with m_o the electron mass, e the particle charge, and c the speed of light). It can be expressed as :

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{1}$$

with $\vec{\beta}$ the normalized velocity of the electron, expressed as :

$$\vec{\beta} = \frac{\vec{v}}{c}. \tag{2}$$

Theoretical foundations of synchrotron radiation were established at the end of the 19th century by J. Larmor [13], who first proposed a specific prediction of time dilation: "... individual electrons describe corresponding parts of their orbits in times shorter for the [rest] system in the ratio, as given by $\sqrt{1-v^2/c^2}$".

Then, A. M. Liénard [14] provided the first correct calculation of the power emitted by an accelerated charged particle, as proportional to $(E/mc^2)^4/R^2$ with m the particle mass, R the orbit radius. The angular and spectral distribution and polarization properties were then described by G. A. Schott [15,16]. Radiation is emitted in a narrow cone of aperture $1/\gamma$.

In 1944, D. D. Ivanenko and I. Y. Pomeranchuk estimated a calculated limit on the energy obtainable in a betatron (of the order of 0.5 GeV) due to energy losses from radiating electrons [17]. Particles slow down and lose synchronism. Because of the spread in revolution frequency with energy, the frequency cannot simply be reduced to maintain synchronism [18] but the particle bunch should be injected in the radiofrequency (RF) at a proper phase (phase stability), as proposed by E. M. McMillan [19] and V. I. Veksler [20] in a 'synchrotron'-type accelerator, which was then built [21]. Julian Seymour

Schwinger (1918–1994, Nobel Prize in 1965) described the peaked spectrum and predicted that higher photon energies should be observed [22, 23].

After the construction of first accelerators, J.-P. Blewett measured the particle energy loss on the 100 MeV betatron and he found it to be in good agreement with the theoretical expectation, but failed to observe synchrotron radiation in the microwaves [24, 25]. The first synchrotron radiation was then observed in the visible tangent to the electron orbit one year later on the 70 MeV General Electric synchrotron, of 29.3 m radius and 0.8 T peak magnetic field [26]. The rapid increase of the intensity with the electron beam energy was measured (fourth power of the energy). The emitted light was found to be polarized with an electron vector parallel to the plane of the electron orbit.

2.3 The undulator

2.3.1 Invention of the undulator

After considering the synchrotron radiation emitted in bending magnets, it is of interest to analyse what happens with a succession of alternated dipoles, as mentioned by V. L. Ginzburg (1916–2009, Nobel Prize in 2003) who first pointed out the radiation emitted by relativistic electrons performing transverse oscillations [27].

H. Motz calculated the field created by a relativistic particle in a magnetic sinusoidal field (i.e. such as produced by undulators) [28, 29], as shown in Fig. 3. He also examined the influence of the bunching of the electrons on the coherence of the radiation. He then observed the polarized visible radiation from an undulator installed on the 100 MeV Stanford accelerator [30]. A buncher set-up after a 3.5 MeV accelerator achieved 1 W peak power at 1.9 mm thanks to the bunching of the electrons. In parallel, the emission of the radiation spectrum produced from an undulator installed on a 2.3 MeV accelerator was investigated by R. Combe and T. Frelot [31]. For the following, it is useful to recall the electron trajectory when it is submitted to the undulator magnetic field and the basics of undulator radiation.

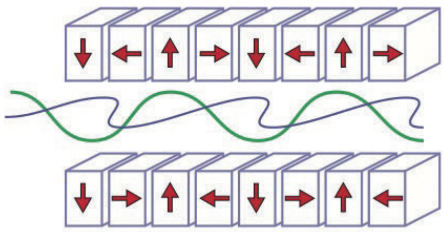


Fig. 3: Planar undulator scheme, creating a periodic magnetic field created by two arrays of permanent magnets arranged in the Halbach configuration [32]. Vertical magnetic field in green, electron trajectory in blue.

2.3.2 Electron movement in an undulator

Let's consider a relativistic electron travelling in an undulator of total length L_u and spatial period λ_u. Let's suppose that the relativistic electron of energy E and velocity v with respect to the laboratory frame is introduced along the s direction. For a relativistic factor $\gamma \gg 1$, $\frac{1}{\gamma} \ll 1$ and the reduced velocity $\beta^2 = 1 - \frac{1}{\gamma^2}$ can be approximated as $\beta \approx 1 - \frac{1}{2\gamma^2}$.

2.3.2.1 Electron movement in a planar undulator

In the case of a planar undulator, creating a field along the vertical direction expressed in the $[0, L_\mathrm{u}]$ interval as

$$\vec{B_{\mathrm{u}}} = B_{\mathrm{u}} \cos\left(\frac{2\pi}{\lambda_{\mathrm{u}}} s\right) \vec{z} = B_{\mathrm{u}} \cos(k_{\mathrm{u}} s) \vec{z} \tag{3}$$

with the undulator wavenumber k_{u} given by $k_{\mathrm{u}} = \frac{2\pi}{\lambda_{\mathrm{u}}}$. The undulator deflection parameter K_{u} expressed as $K_{\mathrm{u}} = \frac{eB_{\mathrm{u}}\lambda_{\mathrm{u}}}{2\pi m_{\mathrm{o}} c}$ is given in practical units as $K_{\mathrm{u}} = 0.934 B_{\mathrm{u}}(T) \lambda_{\mathrm{u}}$ (cm).

The integration of the fundamental equation of the dynamics applying the Lorentz force leads to

$$\vec{\beta} \begin{pmatrix} \frac{K_{\mathrm{u}}}{\gamma} \sin(k_{\mathrm{u}} s) \\ 0 \\ 1 - \frac{1}{2\gamma^2} - \frac{K_{\mathrm{u}}^2}{2\gamma^2} \sin^2(k_{\mathrm{u}} s) \end{pmatrix} \tag{4}$$

and on average over one undulator period:

$$\langle \vec{\beta} \rangle \begin{pmatrix} 0 \\ 0 \\ 1 - \frac{1}{2\gamma^2}(1 + \frac{K_{\mathrm{u}}^2}{2}) \end{pmatrix}. \tag{5}$$

Since the velocity in the vertical direction is zero, the movement takes place in the horizontal plane (x, s). Introducing $k_{\mathrm{u}} s = \omega_{\mathrm{u}} t$ with $\omega_{\mathrm{u}} = k_{\mathrm{u}} c$, a further integration without taking into account the integration constants (there are indeed termination magnets enabling the electron to enter at the origin position) leads to

$$\begin{cases} x = \frac{K_{\mathrm{u}} c}{\gamma} \int \sin(\omega_{\mathrm{u}} t) \mathrm{d}t = \frac{K_{\mathrm{u}} c}{\gamma \omega_{\mathrm{u}}} \cos(\omega_{\mathrm{u}} t), \\ z = 0, \\ s = c\left(1 - \frac{1}{2\gamma^2} - \frac{K_{\mathrm{u}}^2}{4\gamma^2}\right) t + \frac{K_{\mathrm{u}}^2 \lambda_{\mathrm{u}}}{16\,\pi\gamma^2} \sin(2\omega_{\mathrm{u}} t) = \langle v_{\mathrm{s}} \rangle ct + \frac{K_{\mathrm{u}}^2 \lambda_{\mathrm{u}}}{16\,\pi\gamma^2} \sin(2\omega_{\mathrm{u}} t). \end{cases} \tag{6}$$

The maximum amplitude of the transverse motion is $\frac{K_{\mathrm{u}}}{\gamma} \frac{\lambda_{\mathrm{u}}}{2\pi}$. In the longitudinal direction, oscillations occur at twice the frequency, with a maximum amplitude of $\frac{K_{\mathrm{u}}^2 \lambda_{\mathrm{u}}}{16\,\pi\gamma^2}$ generally of much smaller amplitude than in the transverse direction.

2.3.2.2 Electron movement in a helical undulator

In the case of a helical undulator creating a magnetic field in both horizontal and vertical directions given by

$$\begin{cases} \vec{B_{\mathrm{ux}}} = B_{\mathrm{ux}} \sin(\frac{2\pi}{\lambda_{\mathrm{u}}} s) \vec{x} = B_{\mathrm{u}} \sin(k_{\mathrm{u}} s) \vec{x}, \\ \vec{B_{\mathrm{uz}}} = B_{\mathrm{uz}} \cos\left(\frac{2\pi}{\lambda_{\mathrm{u}}} s\right) \vec{z} = B_{\mathrm{u}} \cos(k_{\mathrm{u}} s) \vec{z}, \\ \vec{B_{\mathrm{us}}} = 0, \end{cases} \tag{7}$$

a first integration of the Lorentz equation leads to

$$\vec{\beta} \begin{pmatrix} -\frac{K_{\mathrm{u}}}{\gamma} \sin(k_{\mathrm{u}} s) \\ -\frac{K_{\mathrm{u}}}{\gamma} \cos(k_{\mathrm{u}} s) \\ 1 - \frac{1}{2\gamma^2} - \frac{K_{\mathrm{u}}^2}{2\gamma^2} \end{pmatrix}. \tag{8}$$

Taking integration constants equal to zero, a further integration leads to

$$\begin{cases} x = -\frac{K_u c}{\gamma} \int \sin(\omega_u t) dt = \frac{K_u c}{\gamma \omega_u} \cos(\omega_u t), \\ z = \frac{K_u c}{\gamma} \int \cos(\omega_u t) dt = -\frac{K_u c}{\gamma \omega_u} \sin(\omega_u t), \\ \quad s = c\left(1 - \frac{1}{2\gamma^2} - \frac{K_u^2}{4\gamma^2}\right) t = \langle v_s \rangle ct. \end{cases} \tag{9}$$

The electron trajectory on axis is helical. There is no oscillatory movement in the longitudinal direction at twice the frequency.

2.3.3 Undulator radiation

Let's now consider the specific features of the undulator radiation.

2.3.3.1 Resonance

Electrons wiggling inside the undulator emit synchrotron radiation, as in a succession of bending magnets. They emit synchrotron radiation due to their acceleration in the transverse plane. For each period, the radiation is emitted in a narrow cone of aperture $1/\gamma$ in the forward direction.

The radiation emitted along the undulator interferes constructively depending on the phase lag between the electron and the front of the emitted wave train. One can then introduce the resonance condition: when the electron progresses by λ_u, the wave has travelled by $(\lambda_u + \lambda)$ or more generally by $(\lambda_u + n\lambda)$ with n an integer, the radiation of one electron from the different periods interfere and can add constructively for these wavelengths λ_n, as shown in Fig. 4.

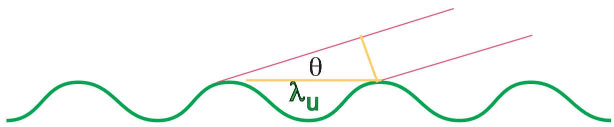


Fig. 4: Undulator resonance condition: when the electron progresses by λ_u, the wave has travelled by $\lambda_u + \lambda$, t_o being the time origin, v_s being longitudinal velocity of the electrons.

In introducing the path difference between the two rays: $n\lambda_n$, one has $c\lambda_u/v_s - \lambda_u \cos\theta/c = n\lambda_n$ leading to

$$n\lambda_n = \lambda_u(1 - \beta_s \cos\theta)/\beta_s. \tag{10}$$

Synchrotron radiation being emitted ahead for small angles, one can approximate $\cos\theta$ by $(1 - \theta^2/2)$, and using $\beta_s = \langle\beta_s\rangle = 1 - 1/2\gamma^2 - K_u^2/4\gamma^2$ for a planar undulator, the resoanant wavelength becomes

$$\lambda_n = \frac{\lambda_u}{2n\gamma^2}\left(1 + \frac{K_u^2}{2} + \gamma^2\theta^2\right). \tag{11}$$

In the case of a helical undulator (with $\beta_s = \langle\beta_s\rangle = 1 - 1/2\gamma^2 - K_u^2/2\gamma^2$), the resonant wavelength is given by

$$\lambda_n = \frac{\lambda_u}{2n\gamma^2}(1 + K_u^2 + \gamma^2\theta^2). \tag{12}$$

This is the so-called 'undulator resonance' wavelength, setting the undulator radiation as a series of harmonics, of order n. The wavelength λ_n of the emitted radiation can be varied by changing the electron beam energy or by a modification of the undulator magnetic field (by changing the gap for permanent magnet insertion devices or the power supply current for electromagnetic insertion devices). The infrared spectral range can be reached with reasonable beam energies. The X-ray regime requires the use of high electron beam energies. Larger wavelengths are obtained for off-axis radiation.

2.3.3.2 Radiated spectrum

The wave packet emitted by each electron contains only a finite number N_u of oscillations as shown in Fig. 5. Thus, in the time domain, the observer receives a train of N_u magnetic periods. The frequency is imperfectly defined. The radiation spectrum corresponds to the Fourier transform of the wave packet emitted by the electron.

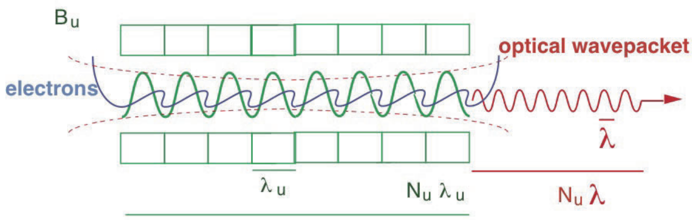


Fig. 5: Radiation train emitted from the undulator of period N_u periods λ_u. The electron moves along the undulator length with a speed v ($vt = N_\mathrm{u}\lambda_\mathrm{u}$) and emits a wave packet whose length is $(c-v)t = N_\mathrm{u}\lambda$. The wave packet contains the same number of periods as the undulator, i.e. N_u.

The electron radiates uniformly from the undulator entrance to its end, so the wave packet has a square envelope. For the optical wave central wavelength $\lambda_l = \frac{2\pi c}{\omega_\mathrm{l}}$, the intensity is given by

$$I(\omega) \propto \left| \int_0^{N_\mathrm{u}\lambda_\mathrm{u}/c} \exp\left[-\mathrm{i}(\omega-\omega_\mathrm{l})t\right]\mathrm{d}t \right|^2 \propto \left\{ \mathrm{sinc}\frac{2\pi N_\mathrm{u}(\omega-\omega_\mathrm{l})}{2\omega_\mathrm{l}} \right\}^2 \tag{13}$$

with n the harmonic number. The on axis radiation spectrum, square of the Fourier transform of this train, is then composed of a series of square sinus cardinals, centred on odd harmonics. The radiation spectrum in the forward direction is thus nearly monochromatic, i.e. it is composed of narrow spectral lines at a well-defined wavelength λ_n. The linewidth of the radiation, called per analogy to conventional lasers the 'homogeneous linewidth' is of the order of

$$\left(\frac{\Delta\lambda}{\lambda_\mathrm{n}}\right)_\mathrm{hom} \approx \frac{1}{nN_\mathrm{u}}. \tag{14}$$

This so-called 'homogeneous linewidth' refers to the case of a single electron. The emission presents then a narrowband in the frequency domain. In other words, the emitted field interferes between different points of the trajectory, leading to sharp spectral peak emission. The higher the number of undulator periods, the smaller the radiated bandwidth. In the case of a single electron, the undulator intensity then scales as N_u^2.

An example of spectrum is given in Fig. 6 with an ideal and real electron beam. For the one electron emission (or ideal electron beam, i.e. filament mono-energetic electron beam) shown in Fig. 6 (a), the radiated line width is ruled by the homogeneous width. However, a real electron beam is not mono-energetic (it has intrinsic energy spread) and presents a transverse size and divergence (emittance contribution).

When adding the emittance term in Fig. 6(b), a satellite peak appears on the red side of the line, and the even harmonics are growing. The linewidth broadening can be interpreted with the contributions

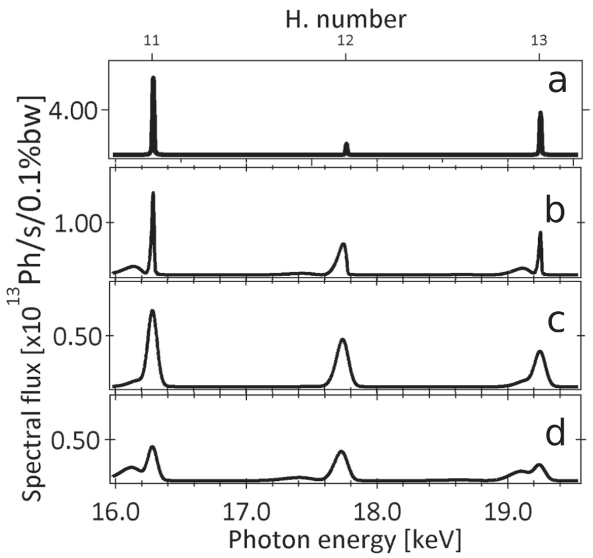


Fig. 6: Spectral flux in the case of a U20 undulator (0.97 T peak field) with 2.75 GeV electrons on the 11th and 13th harmonics: (a) case of a filament monoenergetic electron beam, (b) emittance of 3.9 nm mrad in horizontal and 39 pm mrad in vertical direction, (c) energy spread of 0.1% energy spread, (d) with the contribution of the emittance of (b) and energy spread of (c).

given by the electron beam emittance (with the beam size σ and divergence θ) through the electron beam divergence and size as

$$\left(\frac{\Delta\lambda}{\lambda_{\mathrm{n}}}\right)_{\mathrm{div}} \approx \frac{\gamma^2\theta^2}{1+\frac{K_{\mathrm{u}}^2}{2}}, \tag{15}$$

$$\left(\frac{\Delta\lambda}{\lambda_{\mathrm{n}}}\right)_{\sigma} \approx \frac{2\pi^2 K_{\mathrm{u}}^2}{(1+K_{\mathrm{u}}^2/2)}\frac{\sigma_z^2}{\lambda_{\mathrm{u}}^2}. \tag{16}$$

When adding the energy spread Fig. 6 (c), the line widens symmetrically. This energy spread ($\frac{\sigma_\gamma}{\gamma}$) induced spectral broadening contribution is modelled as

$$\left(\frac{\Delta\lambda}{\lambda_{\mathrm{n}}}\right)_{\sigma_\gamma} \approx \frac{2\sigma_\gamma}{\gamma} \tag{17}$$

The two contributions are added in (Fig. 6d): the lines are damped and widened. The modification of the undulator line can be interpreted as 'inhomogeneous' broadening, which results from different contributions: the electron beam energy spread, size, and divergence. Analytically, one can make the quadratic sum of the different contributions as a first approximation.

When inhomogeneous bandwidth becomes dominant, the intensity is proportional to N_{u}, i.e. $I \propto N_{\mathrm{u}}$. Such a linewidth provides a certain longitudinal coherence length, but is far from the Fourier limit.

2.3.3.3 Undulator emission angle

Because of the wiggling trajectory in the transverse plane (i.e. horizontal for a vertical field), the opening angle is given by the excursion of the reduced velocity, i.e. K_{u}/γ. For the planar undulator (vertical field case), the vertical opening angle is given by $1/\gamma$ as for the usual synchrotron radiation case. For a helical undulator, the angles are given by the velocity excursions in both planes. The radiation is well collimated and presents some transverse coherence, depending on the considered wavelength and on the electron beam contribution.

2.3.4 Temporally coherent emission

Let's discuss further the temporal properties of the radiation. In general, the radiation from the different electrons adds incoherently, and the radiated intensity is proportional to the number of electrons N_e. Longitudinal coherence occurs if the different electrons emit in phase, leading to a radiated intensity scaling as N_e^2. Electrons being in phase can occur either if the electron bunch length is small with respect to the considered wavelength of emission, or if a modulation is imprinted on the electron bunch [33], such as in the free electron laser process. Figure 7 illustrates the three cases of incoherent beam in (a), a bunch short than the wavelength in (b), and of a bunched beam in (c). The bunched case was already considered by Motz [28]. Intermediate case can occur also with abrupt changes in the electronic density, such as with edge radiation from bending magnets.

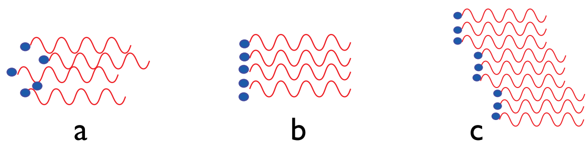


Fig. 7: Distribution of electrons: (a) random distribution leading to incoherent emission, (b) short electron bunch with respect to the radiated wavelength, (c) case of the bunched electron beam.

The normalized longitudinal distribution $n(s)$ can be expressed as $n(s) = N_\mathrm{e}S(s)$ with $S(s)$ being the form factor. The corresponding electric field is then expressed by $E(\omega) = E_\mathrm{o}(\omega)N_\mathrm{e}f(\omega)$ with $E_\mathrm{o}(\omega)$ the electric field of one electron and $f(\omega)$ the form factor in the frequency domain, given by

$$f(\omega) = \int_{-\infty}^{\infty} S(s)\exp\left(\mathrm{i}\frac{\omega_\mathrm{s}}{c}\right)\mathrm{d}s. \tag{18}$$

The electric field results from the sum of the fields emitted by each electron, according to

$$I(\omega) = I_\mathrm{o}(\omega)[N_\mathrm{e}(N_\mathrm{e}-1)f(\omega)^2 + N_\mathrm{e}]. \tag{19}$$

The intensity expression comports two terms, one scaling as the square of the number of electrons (for the coherent emission) and another scaling (for the incoherent synchrotron radiation). The coherent term involves the form factor. In case of short Gaussian electron bunches, one gains typically several orders of magnitude on the radiated intensity.

In the case of a bunched beam, one has

$$S(s) = \sum_{m=1}^{M} S(s - m\lambda_\mathrm{n}). \tag{20}$$

When a form factor is introduced, the undulator emission can become longitudinally coherent emission for λ_n. The form factor resulting from the bunching efficiency is similar to the form factor in Bragg diffraction. The first observation was achieved in 1989 [34].

2.4 The development of vacuum tubes

Let's consider now the domain of vacuum tubes, when bunched electrons are currently generated. The electron beam in vacuum tubes witnessed a rapid and spectacular development at the beginning of the twentieth century for the current amplifier applications such as radiodiffusion and radar detection for icebergs or military use, where high frequency oscillations are needed. Electron beams in vacuum tubes rely on the interaction of a free electron of relativistic factor γ given by $\gamma = \frac{E}{m_\mathrm{o}c^2}$ (with E its energy,

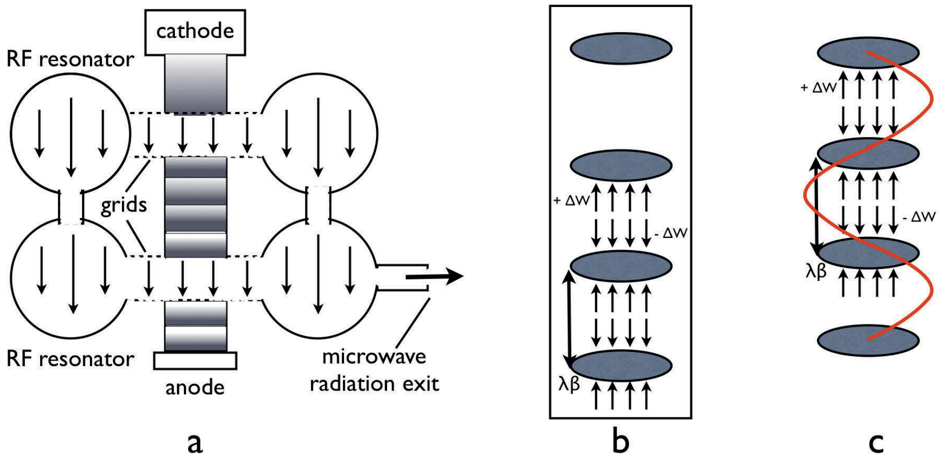


Fig. 8: Klystron principle: (a) klystron scheme, (b) electron bunching by energy modulation in the klystron drift space, electrons accumulate in bunches, (c) phased electron in the second klystron cavity, electric field in red.

m_o the electron mass, e the particle charge, and c the speed of light), $\vec{\beta}$ the normalised electron velocity and an electromagnetic wave of electric field $\vec{E}$ with $\vec{E} = \vec{E}\sin(ks - \omega.t)$ with k the wavenumber and ω the pulsation according to :

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{e\vec{\beta}.\vec{E}}{mc}. \tag{21}$$

A first example of the high-power electron vacuum tube is given by the magnetron, now used for microwave ovens. Electron bunches passing through open cavities excite RF wave oscillations by interaction with the magnetic field, the frequency being determined by the geometry of the cavity. The magnetron can act only as an oscillator for the generation of the microwave signal from the direct current supplied to the tube.

The second example is given by the klystron, invented by the Russel and Sigurd Varian brothers [35]. It consists of two cavities (metal boxes along the tube), as shown in Fig. 8(a) [36].

In the first cavity, an electric field oscillates on a length Δs at a frequency $\nu = 2\pi f$ ranging between 1 and 10 GHz (i.e. with corresponding wavelengths of 30–33 cm). The electrons, generated at the cathode, enter in the first cavity where the input RF signal is applied. They can gain energy according to

$$\Delta W_1 = \int_0^{\Delta s} ec\vec{\beta}.\vec{E}\mathrm{d}t \simeq ecE.\beta\cos(\omega t).\frac{\Delta s}{\beta} = ecE.\Delta s.\cos(\omega t). \tag{22}$$

The sign of ΔW_1 depends on the moment t when the electron arrives inside the cavity. ΔW_1 is modulated in time at a temporal period $T = \frac{2\pi}{\omega}$ or spatial period $\lambda\beta$. On average for the electrons, $\Delta W_1 = 0$ since the electrons have different phases.

Then, the electrons enter into the drift space (see Fig. 8(b)), and they accumulate in bunches. The drift space length is adjusted to enable an optimal electron bunching.

In the second cavity, the electrons have the same phase with respect to the electromagnetic wave, since they have been bunched (see in Fig. 8c). The second energy exchange is given by

$$\Delta W_2 = \sum_{\text{electrons}} \int_0^{l_2} ec\vec{\beta}.\vec{E_{\mathrm{RF}}}\mathrm{d}t = N_\mathrm{e}ecEL_2\cos(\omega t) \tag{23}$$

with L_2 the interaction region in the second cavity, E_{RF} the electric field. The phase of the electrons in the second cavity is ruled by the electrons themselves. The gain in electric field can be very high (practically, 10^3–10^6).

Schematically, the klystron can be understood as a block for the bunching, with another one for the phased interaction with the field. In such a case, a high intensity electron beam excites the RF wave in the second cavity (see Fig. 9). The klystron can be operated in the oscillator mode with a feedback loop on the radiation, where the cavity and the waveguides should be of the order of the wavelength. While looking for larger values of the frequency, i.e. for short wavelengths, the manufacturing of the cavities and waveguides thus limits the operation of the klystron to the microwave region. Another system should be realized for the micrometre and submicrometre spectral ranges.

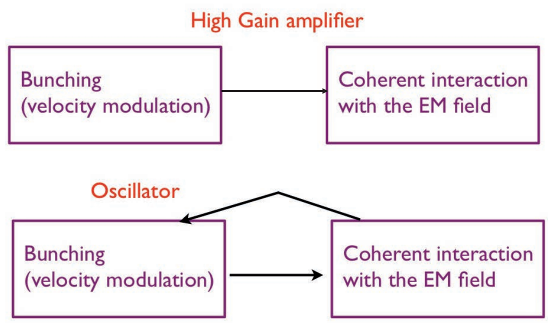


Fig. 9: Bloc diagram of the klystron: amplifier and oscillator cases

More generally, an electron bunch can be accelerated or decelerated by a wave for which the period is longer than the electron bunch one. The linear accelerator relies on such a principle, as seen in Fig. 10. The electrons are produced in an electron gun: a thermo-ionic gun or a photoinjector where the electrons are then generated in trains. With the conventional thermo-ionic gun, the electrons travel into the so-called buncher (a sub-harmonic or harmonic cavity) on the edge of the RF wave, for acquiring energy spread and being bunched by the velocity modulation, as in the klystron case. Then, the electron beam is accelerated by an intense RF wave produced by a klystron and sent in the cavities of the accelerating sections, which can be considered as a series of coupled cavities or as a waveguide where irises slow down the phase of the RF wave to become equal to that of the electrons. In the accelerating sections, the electrons should have the same phase with respect to the RF wave. To be so, they are arranged in small bunches (‘bunching’). For example, for a RF frequency of 1.3 GHz, the period is of 0.77 ns, $1°$ phase corresponds to 2.1 ps.

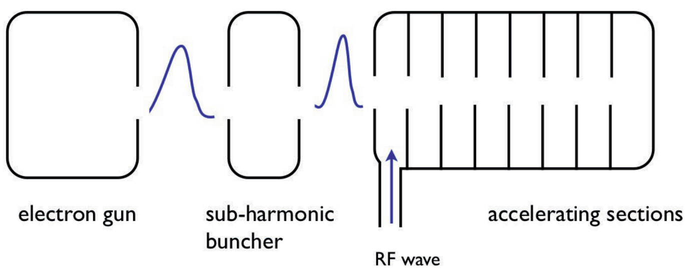


Fig. 10: Scheme of the linear accelerator: electron gun, buncher, accelerating sections

Vacuum tubes such as klystrons, magnetrons, and more generally electronics, discovered at the end of the thirties, underwent a wide development during the second World War with applications such as radiodiffusion and radar detection, where oscillators with high frequencies are needed. The sources generally use electron beams submitted to electric or magnetic fields, where the ‘bunching’ is the key concept for the wave amplification. The use of resonant cavities at the frequency of the emitted wavelength can efficiently insure the retroaction needed for the production of a coherent wavelength.

This field of electronics enabled us to understand that in setting a loop on a wide band amplifier (in connecting one part of its output to its entry), one can transform it into a very monochromatic oscillator. This concept will be used later for the maser and laser inventions.

2.5 The ubitron: undulating beam interaction

FEL precursor works considered whether wave amplification [37] was possible. Then, the ubitron, for ‘undulating beam interaction’ was invented by R. M. Phillips (FEL Prize in 1992 at General Electric Microwave Laboratory), who reports on its discovery [38] in the following terms: “The ubitron (acronym for undulating beam interaction) is a FEL which was setting records for RF power generation 15 years before the term ‘free electron laser’ was coined. As is often the case, the invention of the ubitron was accidental. The year was 1957 and I was searching, at the GE Microwave Lab, for an interaction which would explain why an X-band periodically focussed coupled cavity TWT oscillated when a solenoid focused version did not. The most apparent difference between the two was the behaviour of the electron beam; one wiggled while the other simply spiralled. Out of a paper study of ways of coupling an RF wave to an undulating axially symmetric electron beam came the idea of coupling to the TE_{01} mode by allowing the wave to slip through the beam such that the electric field would reverse direction at the same instant the electron velocity is reversed.”

The ubitron is a high-power travelling wave tube which makes use of the interaction between a magnetically undulated periodic electron beam and the TE_{01} mode in an unloaded waveguide [39]. The scheme is illustrated in Fig. 11. The basic idea is to couple the TE_{01} mode by allowing the wave to slip through the beam such that the electric field reverses direction at the same instant the electron velocity is reversed. Several beam guide ubitron configurations (planar, coaxial, circular) can be considered, and they can provide 100 times the interaction area of a TWT (Travelling Wave Tube). The electron–wave interaction exhibits the same type of first-order axial beam bunching characteristic of the conventional slow wave travelling wave tube. In consequence, the ubitron can be used in extended interaction klystrons and electron accelerators, as well as travelling wave tubes.

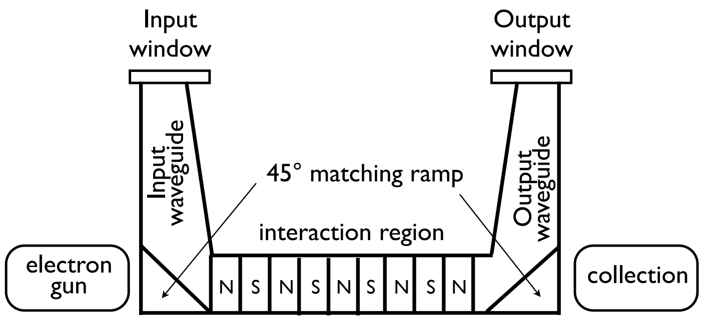


Fig. 11: Scheme of the ubitron: an electron beam from an electron gun, wiggling with axial symmetry, can couple to an RF wave. Alternated magnetic poles provide the axial symmetry.

First experiments used an undulated pencil beam in a rectangular waveguide [40]. They presented unique features such as a very broad interaction bandwidth which results from the absence of a dispersive slow wave circuit, a variable interaction phase velocity, and hence, variable saturation power level.

Among the physical embodiments of the ubitron are a number of higher-order mode waveguides and beam configurations. They opened at that time interesting prospects for high-power millimetre wave amplification. As reports C. Brau [1], “the ubitron uses the same configuration of electron beam and magnetic field as proposed by Motz, but at a high enough electron density that space-charge waves are excited by the electron beam”. High power (>1 MW) and high efficiency ($>10\%$) were obtained at wavelengths from 10 cm to 5 mm. However, other devices developed at about the same time, such as the travelling-wave tube, offered higher gain and other advantages, and the ubitron was not actively pursued.

Studies were extended to the interaction of relativistic particles and waves in the presence of a static alternated magnetic field [41]. The possibility of achieving stimulated emission [42] was also considered.

2.6 The maser discovery

After the second World War, RF sources and detectors developed for radar and transmission were also used for fundamental research, in particular Hertzian spectroscopy of atoms and molecules, radio-astronomy, and magnetic nuclear resonance. In the early fifties (1953), Charles Townes (1905–2015, Nobel Prize in 1964) [43] in the USA (Columbia University, New York), Nicolay Gennadiyevich Basov (1922–2001, Nobel Prize in 1964) in 1952), and Aleksandr Mikhailovich Prokhorov (1916–2002, Nobel Prize in 1964) in the Soviet Union (Lebedev Institute, Moscow) independently aimed at creating new microwave sources in replacing the amplification by an electron beam amplification with the help of the stimulated emission process in molecules. In order to make a ‘quantum’ microwave oscillator, they introduced excited molecules into a microwave cavity which was resonant to the frequency of the molecule transition. They met for the first time in 1959 in the USA at the first Quantum Electronics Conference. Some physicists were sceptical, including N. Bohr (1885–1962, Nobel Prize in 1922), who was not very familiar with recent advances in electronics and could hardly admit that the phase coherence of the oscillator could last longer than the excited state lifetime. To perform the population inversion required for the stimulated emission, Townes, Basov, and Prokhorov had the idea to use the spatial separation of excited molecules (Stern–Gerlach type), which is efficient but not very practical. The population inversion can also be performed in an easier manner by a proper excitation of the radiation of the atoms and molecules. In 1949, Alfred Kastler (1902–1984, Nobel Prize in 1966) and Jean Brossel proposed and developed ‘optical pumping’, based on the use of circularly polarized light for selectively filling some Zeeman sub-levels of atoms. In 1951, E. Purcell and R. Pound, working on nuclear magnetic resonance, showed that RF radiation enables us to create samples of ‘negative temperature’, i.e. a population inversion. Inspired by the resonators of vacuum tubes, the light feedback is ensured by a cavity resonant on its fundamental mode.

In 1954, the first maser (microwave amplification by stimulated emission of radiation) was operated in the microwave region by Charles Townes [43] at Columbia University with the NH_3 molecule. In 1954, N. Bloembergen (Nobel Prize 1981 on laser spectroscopy and non-linear optics), Basov, and Prokhorov proposed the 3-level maser concept: with a proper illumination of a solid such as a ruby crystal, population inversion takes place. It is easier to operate than the equivalent gas-based maser. It has been used in particular as a very low noise amplifier.

Masers also exist naturally in stars.

The new domain of ‘quantum electronics’ has emerged from the interplay between the scientific fields of electronic vacuum tubes and quantum properties of matter and it has seen an extraordinary spread and has raised a lot of interest. The question was then of the extension of the maser to the optical wavelengths.

2.7 The laser discovery

In order to achieve an optical maser, the maser cavity resonant on its fundamental mode must become extremely small (of the order of 1 micrometre) and this was not possible at that time. Nowadays, these cavities are manufactured using nanotechnologies (for VCSEL (vertical cavity surface emitting laser) semi-conductor lasers). Charles Townes and Arthur L. Schawlow (1921–1999, Nobel Prize in 1981 on laser spectroscopy and non-linear optics) at Bell Labs, G. Gould (1920–2005) at Columbia [44], and A. Prokhorov at the Lebedev Institute proposed feedback with an open resonant cavity (Fabry–Perot-type used in spectroscopy). These 'optical lasers' were named lasers for light amplification by stimulated emission of radiation.

In a Fabry–Perot cavity of length L_c, the light makes round trips between the two mirrors on which it is reflected. C. Townes and A. Schawlow said, in their reference paper [4]:

"The extension of maser techniques to the infrared and optical region is considered. It is shown that by using a resonant cavity of centimetre dimensions, having many resonant modes, maser oscillation at these wavelengths can be achieved by pumping with reasonable amounts of incoherent light. For wavelengths much shorter than those of the ultraviolet region, maser-type amplification appears to be quite impractical. Although using of a multimode cavity is suggested, a single mode may be selected by making only the end walls highly reflecting, and defining a suitably small angular aperture. Then extremely monochromatic and coherent light is produced. The design principles are illustrated by reference to a system using potassium vapor".

A scheme of such an optical cavity is shown in Fig. 12.

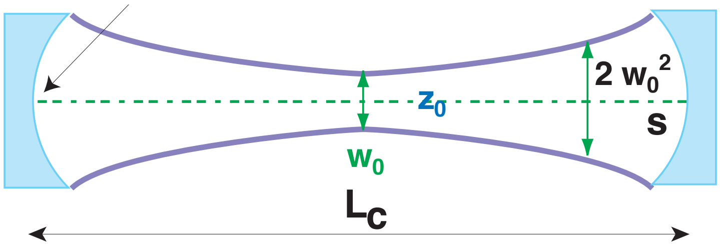


Fig. 12: Scheme of an optical resonator

For the light to interact at each pass with the amplifier medium, and to get larger, it should be in phase with the one from the previous pass. In other words, the optical path for one round trip should be equal to an integer number p of wavelengths λ, i.e. $2L_c = p\lambda$. For a fixed cavity length L_c, only the wavelengths verifying $\lambda = \frac{2L_c}{p}$ can be present in the 'optical maser' light, defining the longitudinal modes of the cavity associated with different values of p. The shift in frequency between two modes is given by $\nu = \frac{c}{\lambda} = \frac{c}{2L_c}$.

In practice, in order to focus the light transversally and to avoid diffraction losses, one of the mirrors should be concave. The light circulating in the optical resonator is not a plane wave, and the radius of the light changes along its propagation direction [45,46]. In case of a cavity with two concentric mirrors, the light radius is minimum at the waist w_0 and diverges according to $w(s) = w_0\sqrt{1 + \frac{s^2}{Z_R^2}}$ with Z_R the Rayleigh length, i.e. the distance from the waist for which the radius of the light beam is increased by a factor $\sqrt{2}$, given by $Z_R = \frac{\pi w_0^2}{\lambda}$. This corresponds to the diffraction of light by an aperture of diameter $2w_0$. The radiation at the entry and at the exit of the cavity have the same characteristics. The divergence of the light beam θ_r can be expressed as $\theta_r = \frac{\lambda}{\pi w_0}$. The higher the focus, the smaller the waist and the higher the beam divergence. The beam is very directional, it can be adapted (focused or expanded) to the users need with the help of mirrors and lenses. In the case of a He–Ne laser at 633 nm with a waist of 600 μm, the Rayleigh length is of the order of 2 m. Over 2 m propagation length, the light beam diameter remains practically constant.

Following the publication of the theoretical paper by Arthur L. Schawlow and C. Townes on 'Infra-red and Optical masers' in 1958 [4], different laboratories entered the race to demonstrate experimentally the 'optical maser'. It was won by an outsider in 1960, Theodore Maiman (1927–2007), who had the idea to realize a pulsed and not a CW (continuous) source, for which the oscillation conditions take place transiently. On May 6 1960, Maiman achieved the first working laser by generating pulses of coherent light from a fingertip-sized lump of ruby (chromium in corundum) illuminated by a flash lamp [5,47] in Malibu (USA). The device was extremely simple. Several 'optical masers' followed [48]. The calcium fluoride laser was achieved by Mirek Stevenson and Peter Sorokin at General Electric in 1960. The He–Ne laser was operated by Ali Javan (1926–2016), Bill Bennett, and Don Herriott in 1961 [49], with the population inversion achieved with a discharge on the Ne atoms bringing a fraction of the He atoms to metastable states, the He atoms being relaxed by collision with Ne atoms in transmitting to them their energy excess. Then, in 1962 followed the semi-conductor AsGa laser (diode laser) where a p–n junction of the gallium arsenide semiconductor through which a current was passed and it emitted near-infrared light from recombination processes with very high efficiency, first operated by R. Hall (1919–) [50] and others [51].

3 The invention of the free electron laser concept

3.1 The FEL concept emergence: motivations for an exotic laser

Early work on stimulated bremsstrahlung was conducted at the beginning of the twentieth century [52,53] and later [54].

Following the discovery of the laser, much less interest was devoted to the electron tube based systems. The Gaussian eigenmodes of free space provided an alternative to the coupled slow wave structures of the prior electron devices. In addition, the laser radiation is independent of phase.

In the original paper from A. Schawlow and C. Townes [4], it was written that "As one attempts to extend maser operation towards very short wavelengths, a number of new aspects and problems arise, which require a quantitative reorientation of theoretical discussions and considerable modification of the experimental techniques used" and "These figures show that maser systems can be expected to operate successfully in the infrared, optical, and perhaps in the ultraviolet regions, but that, unless some radically new approach is found, they cannot be pushed to wavelengths much shorter than those in the ultraviolet region".

J. M. J. Madey (1943–2016, first FEL Prize in 1988) [55], from Stanford University, thought that "A. Schawlow and C. Townes descriptions of masers and lasers coupled with the new understanding of the Gaussian eigenmodes of free space offered a new approach to high frequency operation that was not constrained by the established limits to the capabilities of electron tubes" [56] and he wondered whether there was "a Free Electron Radiation Mechanism that Could Fulfill these Conditions" and considered the different possible radiation processes. He first examined Compton scattering, as shown in Fig. 13, which appeared as the most promising candidate. Indeed, stimulated Compton backscattering has been analysed by Dreicer in the cosmic blackbody radiation [57].

First analysis of the stimulated Compton backscattering was carried out by Pantell (eighth FEL Prize in 1995, shared with G. Befeki (1925–1995)) [42]. Precursor works include the study of stimulated emission of bremsstrahlung [58], and of the possibility of frequency multiplication, and wave amplification by means of some relativistic effects [37], radiation transfer and investigations into whether negative absorption (i.e. amplification) could be possible in radio astronomy [59].

The Compton backscattering (CBS) process between a laser pulse and a bunch of relativistic particles (electrons, positrons, etc.) leads to the production of high-energy radiation coming from the head-on collision between the photons and the particles. In order to reduce the divergence of the scattered radiation, it is better to use a relativistic electron beam, which radiation cone is reduced to $1/\gamma$. For relativistic particles (i.e. $\gamma \gg 1$), the energy of the produced photons E_{CBS} is given by

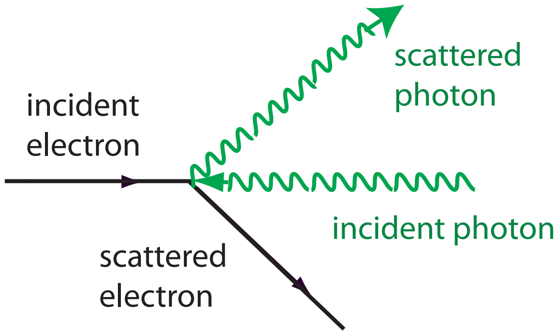


Fig. 13: Compton backscattering scheme

$$E_{\text{CBS}} = \frac{4\gamma^2 E_{\text{ph}}}{1 + (\gamma\theta)^2} \tag{24}$$

with E_{ph} the energy of the initial photon beam, θ the angle between the CBS photons and the electron beam trajectory. The energy of the relativistic electrons can easily be changed, so the CBS radiation could be tuneable.

J. M. J. Madey had the idea to make the phenomenon more efficient by using the magnetic field of an undulator [6]. He was aware of the theoretical [28] and experimental [30] work of Motz, where radiation from bunched beams has been observed. He considered that “Relativistic electrons can also not tell the difference between real and virtual incident photons, permitting the substitution of a strong, periodic transverse magnetic field for the usual counter-propagating real photon beam” [56]. The proposed FEL scheme is shown in Fig. 14.

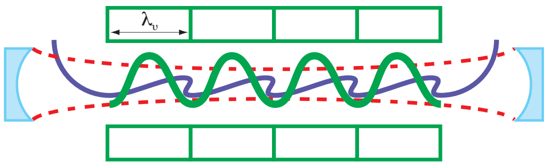


Fig. 14: Scheme of the FEL oscillator with the gain medium consisting of relativistic electrons in the undulator

3.2 The FEL quantum approach

J. M. J. Madey then calculated in the frame of quantum mechanics the gain due to the induced emission of radiation into a single electromagnetic mode parallel to the motion of a relativistic electron through a periodic transverse dc magnetic field [6]. He found that finite gain is available from the far-infrared through the visible region, raising the possibility of continuously tuneable amplifiers and oscillators in such a spectral range, and he envisioned further the possibility of partially coherent radiation sources in the ultraviolet and X-ray regions to beyond 10 keV. He introduced the notion of the ‘free electron laser’ [6].

According to the Weisächer–Williams approximation, the undulator field of period λ_u can be assumed to be a planar wave of virtual photons. It enables an easier way to relate the transition rates to more easily calculable Compton scattering rates. By Lorentz transformation, the wavelength λ' of a planar wave in the moving frame of the electrons in the undulator is given by

$$\lambda' = \frac{\lambda_{\mathrm{u}}}{\gamma_{\mathrm{s}}} \tag{25}$$

with γ_{s} the Lorentz factor of the scattered electron. Photon emission and absorption are forbidden by conservation of energy and momentum. For free electrons, one can then consider a two photon process, such as Compton scattering, as shown in Fig. 15.

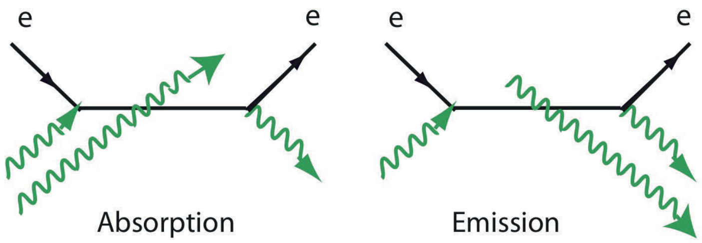


Fig. 15: Feynman diagrams of Compton scattering

The emission is then again given by the Doppler effect, according to

$$\lambda = \frac{\lambda'}{(1+\beta_{\mathrm{s}})\gamma_{\mathrm{s}}} = \frac{\lambda_{\mathrm{u}}^2}{2\gamma_{\mathrm{s}}^2} \approx \frac{\lambda_{\mathrm{u}}^2}{2\gamma^2}. \tag{26}$$

For stimulated Compton scattering, the diffusion transition rate τ_{d} should be larger than the absorption one τ_{a}. One can define the gain g as

$$g = \tau_{\mathrm{d}} - \tau_{\mathrm{a}}. \tag{27}$$

The original calculation, performed in [6, 60], is not reproduced here. It was also found that the gain expression does not depend on Planck's constant h. Further developments followed [61].

The ubitron can also be considered as another precursor of the FEL [39].

3.3 The FEL regimes

Different regimes can be considered [62]. The FEL can thus be described as a stimulated Compton scattering device, as shown in Fig. 16. If the electronic density is large enough, a plasma wave can develop.

3.3.1 The Compton FEL regime

In the Compton regime, the scattered wavenumber k'_{s} equals the incident wavenumber k'_{i}:

$$k'_{\mathrm{s}} = k'_{\mathrm{i}}. \tag{28}$$

3.3.2 The Raman FEL regime

In the Raman regime, the scattered wavenumber k'_{s} is the sum/difference of the incident wavenumber k'_{i} and of the plasma wavenumber k'_{p}, leading to the Stokes and anti-Stokes lines.

$$k'_{\mathrm{s}} = k'_{\mathrm{i}} \pm k'_{\mathrm{p}}. \tag{29}$$

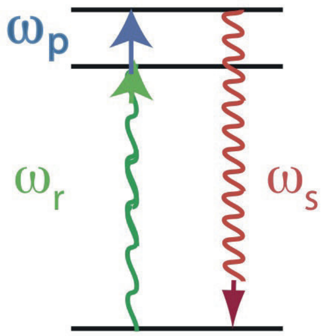


Fig. 16: Stimulated Compton scattering scheme in the electron frame, with ω_{p} the plasma pulsation, ω_r the resonant pulsation, and ω_{s} the scattered pulsation.

In the laboratory frame, it comes to $\omega_{\mathrm{s}} = \omega_{\mathrm{i}} \pm \omega_{\mathrm{p}}$ where the plasma pulsation ω_{p} is given by $\omega_{\mathrm{p}} = \sqrt{\frac{n_{\mathrm{e}}e^2}{\epsilon_{\mathrm{o}}m_{\mathrm{o}}\gamma^3}} = \sqrt{\frac{J_{\mathrm{e}}e}{\epsilon_{\mathrm{o}}m_{\mathrm{o}}c\gamma^3}}$ with n_{e} the electronic density, ϵ_{o} the vacuum permeability, and J_{e} the current density $J_{\mathrm{e}} = n_{\mathrm{e}}ec$.

Practically, one considers that the FEL is in the plasma regime if the number of plasma oscillations N_{p} done by the electron while it travels into the undulator is at least one:

$$N_{\mathrm{p}} = \frac{N_{\mathrm{u}}\lambda_{\mathrm{u}}}{\lambda_p} = \frac{N_{\mathrm{u}}\lambda_{\mathrm{u}}\omega_{\mathrm{p}}}{2\pi c} = \frac{N_{\mathrm{u}}\lambda_{\mathrm{u}}}{2\pi c}\sqrt{\frac{J_{\mathrm{e}}e}{\epsilon_{\mathrm{o}}m_{\mathrm{o}}c\gamma^3}}.$$

Thus $N_{\mathrm{p}} > 1$ if $J_{\mathrm{e}} < \frac{4\pi^2\epsilon_{\mathrm{o}}m_{\mathrm{o}}c^3\gamma^3}{eN_{\mathrm{u}}^2\lambda_{\mathrm{u}}^2}$.

The regimes of FEL are recapitulated in Fig. 17.

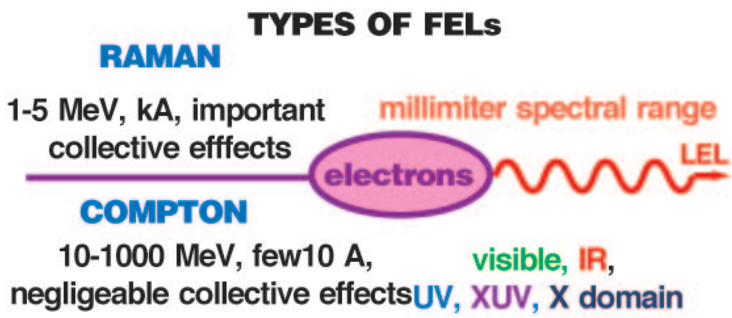


Fig. 17: Comparison between Compton and Raman FELs

4 The FEL classical approach (low gain regime)

4.1 Stimulated scattering in a plasma fluid type approach

P. Sprangle (fourth FEL Prize, in 1991) continued the exploration of stimulated Compton scattering of an electromagnetic wave from relativistic electrons [42, 61] using a plasma approach. Sprangle and Granatstein examined the stimulated cyclotron resonance scattering and production of powerful submillimetre radiation [63] and stimulated collective scattering from a magnetized relativistic electron beam [64] where the pump satisfies the dispersion relation associated with the beam in the magnetic field and the scattered waves consist of collective plasma oscillations as well as right- and left-polarized electromagnetic waves, travelling parallel and antiparallel to the beam. The frequency of the forward-scattered electromagnetic wave is Doppler shifted. Conditions for enhanced stimulated scattering and growth rates were found [64]. An original following work on noise excitation analysis [65] can be mentioned. Variants of FEL were considered, such as the gyrotron with a uniform magnetic field [66]. Saturation and phase (wave refractive index) were analysed [67]. The Raman-type theoretical developments are not detailed here [68, 69].

4.2 The FEL classical approach using the Weizsäcker–Williams approximation and electronic distribution function

F. A. Hopf *et al.* [70] continued the investigation of stimulated emission of radiation in a transverse magnetic field. He pointed out that the theories explored so far [6, 42, 57, 60, 61] were all "quantum mechanical in nature. They give impression that they have to be so, since it is argued that it is the electron recoil $\Delta p = h/\lambda_c$ where λ_c is the Compton wavelength, which is the source of the finite gain. Furthermore, quantum approaches, while agreeing on the structure of the gain formula, differ from one another by orders of magnitude in numerical coefficients" [70]. He then shows that "this problem is completely classical, and that the gain is produced by a bunching of the electron density in the presence of a field".

He works directly in the laboratory frame. Considering the Weizsäcker–Williams approximation and in the case of the extreme relativistic limit, the static undulator field of period λ_u can be replaced by a pure electromagnetic field of wavelength $\lambda_i = (1 + \beta_s)\lambda_u \approx 2\lambda_u$. The electron motion is treated via the collisionless relativistic Boltzmann equation, according to

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \dot{x}_i \frac{\partial f}{\partial x_i} + \dot{P}_i \frac{\partial f}{\partial P_i} = 0, \tag{30}$$

with **P** the canonical momentum, **x** the position and dot the total derivative with respect to time.

The total number of electrons $N(t)$ is given by $N(t) = \int \mathrm{d}^3x \int f(\mathbf{x}, \mathbf{P}, t)\mathrm{d}^3P$.

The Boltzmann equation is coupled to the transverse current $\mathbf{J_t}$ via the Maxwell equation, and is given by

$$\mathbf{J_t} = e \int \mathbf{v_T} f(\mathbf{x}, \mathbf{P}, t)\mathrm{d}^3P, \tag{31}$$

with $\mathbf{v_T}$ the transverse velocity. The scheme is sketched in Fig. 18.

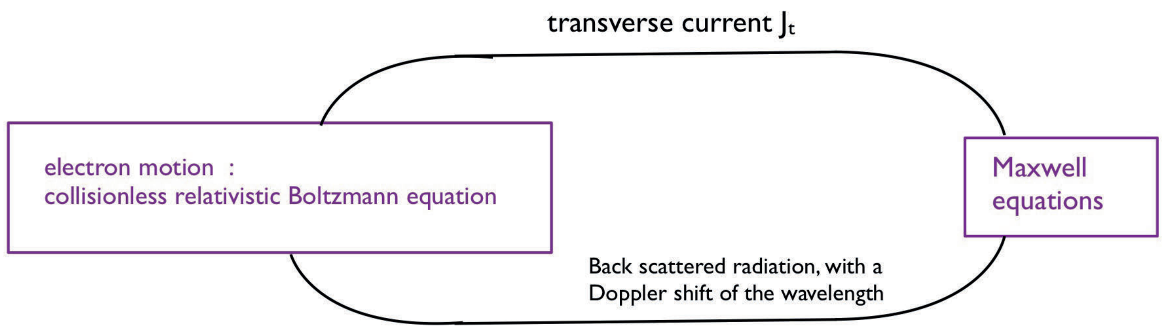


Fig. 18: Diagram of the low gain FEL classical theory using the Boltzmann equation for the movement of the electrons in the undulator and the scattering process.

By assuming that:

- the electromagnetic field is transverse and depends on s and t;
- the transverse velocity spread of the electronic distribution is neglected since the electrons propagate with a relativistic velocity along the direction s;
- the mass shift for electric fields smaller than 10^{12} V/m,

the problem is reduced to a one-dimensional one. The interaction term of the reduced Boltzmann equation is similar to the one in the Klein–Gordon Hamiltonian, and the source term of the reduced Maxwell equation is "proportional to a density times an electric field. This is exactly the same as in usual scattering problems, where the d'Alembertian of the electric field is proportional to the second derivative of the polarization, which is in turn proportional to a density times the electric field. Hence, we see at this point that the problem at hand is nothing else than a usual classical scattering problem, complicated only by the fact that we deal here with relativistic particles" [70]. Then, the reduced distribution is developed in perturbation series, the first order giving the small-signal theory. In this case, only two modes of the field are kept, the incident one (i.e. the static periodic magnetic field in the Weizsäcker–Williams approximation), and the scattered one. The relevant scattered mode in the up-conversion scheme is the backscattered radiation, which propagates in the direction of the electron beam. Its wavelength is Doppler shifted, as seen previously. New assumptions are made:

- the amplitude and phase are slowly varying;
- the depletion of the incident field is neglected since it is assumed to be very intense;

enabling us to find the first-order term of the reduced electron distribution function $h^{(1)}(s, P_\mathrm{s}, t)$, expressed as

$$h^{(1)}(s, P_\mathrm{s}, t) = -\frac{e^2(k_\mathrm{s}+k_\mathrm{i})n_\mathrm{e}A_\mathrm{i}^*A_\mathrm{s}}{P_\mathrm{s}}\frac{\mathrm{d}F}{\mathrm{d}P_\mathrm{s}}\frac{\exp(-\mathrm{i}\mu s)-1}{\mu}\exp(-\mathrm{i}\Delta\omega(t-s/v_\mathrm{s}))+c.c. \tag{32}$$

where k_i (respectively k_s) is the wavenumber of the incident (scattered) field, A_i (respectively A_s) is the vector potential of the incident (scattered) field, $F(P_\mathrm{s})$ the initial electron momentum distribution, $\Delta\omega = \omega_\mathrm{s} - \omega_\mathrm{i}$ with ω_i (respectively ω_s) the pulsation of the incident (scattered) field, and $\mu = \Delta\omega - (k_\mathrm{s}+k_\mathrm{i})$. "$h^{(1)}(s, P_\mathrm{s}, t)$ describes electron density fluctuations [70] (bunching) which are responsible for the scattering". Here, Hopf is pointing out that the bunching is a key process for the FEL interaction. Introducing the reduced electron distribution into the Maxwell equations, one finds the small-signal low gain expression. In the case of the small cavity limit (where the initial electron momentum distribution function $F(P_\mathrm{s})$ can be taken as a δ function, i.e. to the limit of a homogeneously broadened medium), the total small-signal gain is given by

$$g \cong 64\pi^2 r_\mathrm{o}^2 F_f \frac{n_\mathrm{e}}{mc^2}\frac{k_\mathrm{i}^{1/2}}{k_\mathrm{s}^{3/2}}L_\mathrm{c}^2 I_\mathrm{i}\frac{\mathrm{d}(\sin\eta_\mathrm{c}/\eta_\mathrm{c})^2}{\mathrm{d}\eta_\mathrm{c}} \tag{33}$$

with $\eta_\mathrm{c} = \mu L_\mathrm{c}/2$, r_o the classical radius of the electron, F_f the filling factor term representing the transverse overlap (ratio of the electron beam transverse to the section of the optical beam in the cavity), I_i the incident flux. Different cases occur.

- If $\eta_\mathrm{c} = 0$ there is exact conservation of momentum, and no net gain.
- If $\eta_\mathrm{c} < 0$ (i.e. for electrons with a velocity $v > v_\mathrm{o}$), the net result is a gain. It is the equivalent of the Stokes line in Raman scattering.
- If $\eta_\mathrm{c} > 0$ (i.e. for electrons with a velocity $v < v_\mathrm{o}$), the net result is an absorption. It is the equivalent of the anti-Stokes line in Raman scattering.

The maximum gain is found for $\eta_c = -\pi/2$ in agreement with the Madey result derived in the quantum mechanics frame within a factor 0.8. It is worth citing part of the conclusion "that the free-electron laser is a completely classical device. The stimulated scattering producing amplification is due to electron bunching, rather than to the Compton recoil, as argued previously. This result not only is important from an academic viewpoint, but also greatly simplifies the analysis of the strong-signal regime and of the saturation of this new laser" [70]. The major step here is the understanding that the gain results from a bunching of the electronic density in the presence of a field.

Two months later, Hopf continued with the strong-signal case "In order to assess the potential of any practical laser device, it is necessary to *complement* the small-signal theory by an analysis of the mechanism of saturation" [71]. In the strong signal case (i.e. for 'long' undulator, 'high' current), the change of electric field in one pass cannot be neglected anymore. The undulator is still treated in the frame of the Weizsäcker–Williams approximation. The coupled Maxwell–Boltzmann equations are reduced differently from the small-signal case where the longitudinal part of the Boltzmann distribution function h was expanded in powers of $|\mathrm{A_i A_s}|$ limited to the first order. Here, h is then expressed as a harmonic expansion (previous expansion to higher orders would diverge), leading to a set of generalized Bloch equations in keeping the first term of the expansion. These generalized Bloch equations have "a striking resemblance to the optical Bloch equations" involving an equivalent "population inversion" and "polarization" term. "However, it differs from them in two respects". "This difference in structure lies in the fact that in a free electron laser, the gain is not proportional to the electron distribution function. It is its derivative with respect to p_s, (rather than the gain itself) which plays the role of an inversion." By supposing that the saturation mechanism corresponds to "deceleration of the electrons through the gain line to the point of zero gain", one can express a saturation flux and find the "maximum field extractable from a free-electron laser (i.e. the output field when the laser is in the saturation regime)" [71] as

$$A_{\mathrm{sat,max}} \cong \left[\frac{\lambda_s}{L_c}\right]^2 \frac{M^2 c 2 \gamma^4}{e^2} A_i. \tag{34}$$

He then computed the efficiency in the case of the Stanford experiment, and found it to be of the order of 5×10^{-3}. He deduces that "This implies that free electron lasers have the potential to work at high power, but they must be operated in a pulsed mode, with small per shot efficiency." He just pointed out that the previously described analysis is very simplified: "In reality, a more detailed analysis shows that a major contribution to the saturation is a strong alteration of the electron distribution such that the laser eventually reaches the large- cavity limit". Hopf *et al.* had shown here that FELs "have the potential to work at high power".

The expansion of the Boltzmann distribution function was then not limited to the first term, and the problem reduces to the Klein–Gordon equation [72]. A theory including Raman scattering has also been developed [62, 73, 74].

In Madey's and Hopf's approaches, the FEL has been explained in terms of collective phenomena. Single particle theory can also be applied, as described below.

4.3 The classical approach considering the relativistic motion of the electrons in the undulator

W. B. Colson (Second FEL prize, 1989) aimed at a broader theoretical framework. He analysed the radiation from electrons travelling through a static transverse periodic magnetic field with classical, semi-classical, and quantum field theories. He considered the radiation characteristics of the electrons in the undulator and developed stimulated emission rates and laser evolution equations describing exponential gain and saturation [75]. His paper was published the same month as the second publication by F. A. Hopf [71]. He insisted on the importance of including the filling factor term, representing the overlap between the electron beam and the optical wave. W. B. Colson received the Second FEL Prize after

J. M. J. Madey, in 1989 "Bill laid the foundation for the classical theory of Free Electron Laser, enabling a wide audience to understand the operating principles of FEL". He gave "an outstanding contribution to the understanding of free electron laser mechanisms".

4.3.1 Resolution of the one-body classical Lorentz equation in the presence of a periodic magnetic field and a plane electromagnetic wave

W. B. Colson looked for a more appropriate magnetic field description and by "solving the one-body classical Lorentz force equation in presence of periodic magnetic field and a plane electromagnetic wave" [76, 77]. He found that "the non-linear electron dynamics to the phase space paths of a simple pendulum in the limit of small gain. The position and energy of a single electron are simply expressed as a function of time". He was able to determine the gain and to link it to the derivative of the sinc-like function describing the spontaneous emission [75], the electron modulation, and the saturation for strong laser fields corresponding to closed phase space paths, where the electron beams becomes resonant and the gain drops. He insisted on the "importance of the electron beam produced by the FEL device. The combined magnet and radiation fields conspire in a controlled way to yield a *coherently modulated relativistic electron beam*". It is found that the evolution of the system follows the pendulum equation which has been widely adopted and will be presented below. The FEL Prize recipient receives a clock, symbolizing the FEL pendulum equations! Phase space paths are illustrated in Fig. 19. The electrons that are initially resonant and on a phase equal to an integer p times π, corresponding to points in phase space located at $(p\pi, 0)$, do not contribute to work. Electrons near these points evolve very slowly in time, with motion at 'even-p' points being stable and 'odd-p' unstable. In analogy, a simple pendulum would be at the bottom (p-even) or top (p-odd) of its arc. If an electron is not at a critical point, the radiation field alters its energy and position. A shift in relative position, proportional to the square of the pendulum frequency, occurs and redistributes electrons along the beam axis. For an initially uniform, mono-energetic beam, half the electrons within a given λ_r are positioned such that work is done on them; they gain energy and move ahead of the average flow. The rest of the electrons loose energy to the radiation field and move back. This causes the 'bunching' of the beam. Electrons can undergo closed and open orbits.

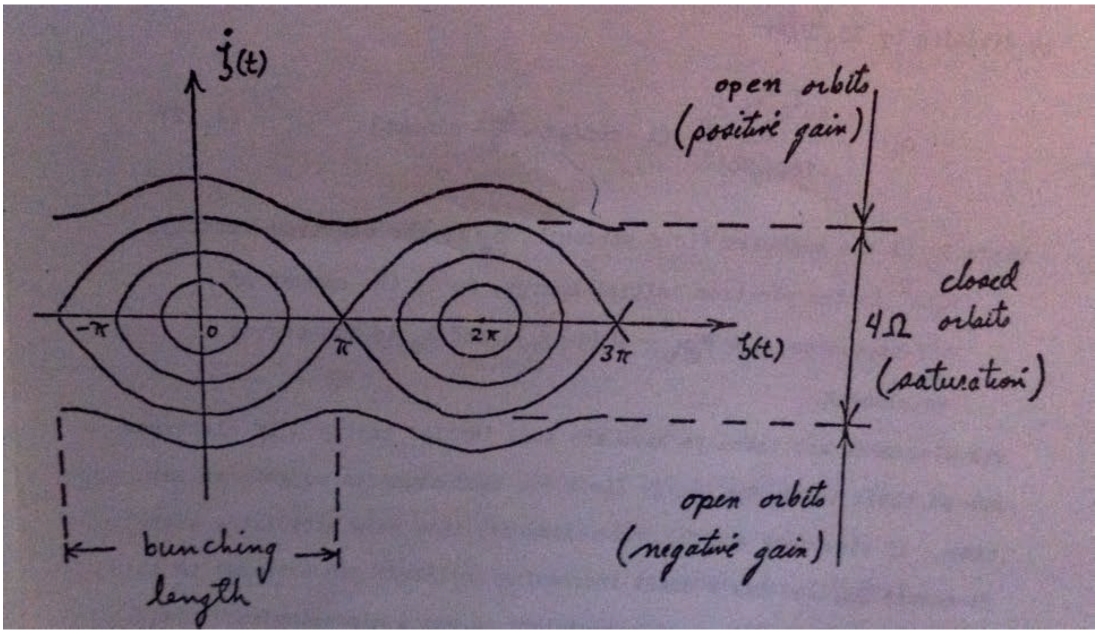


Fig. 19: Electron phase space paths in the case of the pendulum for a helical undulator (low gain regime), from [78]

Then, the electron motion is depicted with a Hamiltonian approach, while the phase space representation of the pendulum is deepened and the influence of the detuning (delay between the electrons in the undulator and optical pulses in the optical cavity) is studied [79]. The importance of the 'bunched' beam is emphasized and the use of an external laser with a static periodic field to create the modulation at optical wavelengths is considered.

The model is then described self-consistently, using single particle dynamics and Maxwell's equations. The optical wave evolution is governed by Maxwell's equation in the presence of an electron current. Assuming that the amplitude and phase are slowly varying, two differential equations describing the amplitude and the phase of the wave are found. The dynamics of the electrons is ruled by the Lorentz equations in the presence of the combined static and radiated fields. The total current results from the sum of all individual particle currents. The two sets of equations are then combined. It is found that the microscopic electron bunching drives the amplitude and phase of the optical wave [80] as shown in Fig. 20. The saturation is well described within the frame of the self-consistent pendulum equation: "When the radiation field becomes large, the electron becomes trapped in closed orbits of the pendulum phase space. In the beam frame, the bunching electrons will have moved on the order of an optical wavelength: at this point, the gains tops and the laser saturates" [80].

The model is then applied to study the operation on higher harmonics [81–83], as developed later.

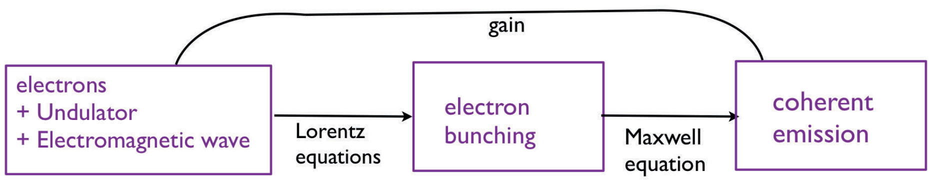


Fig. 20: Diagram of the low gain FEL classical theory using the movement of the electrons in the undulator and the interaction with the electromagnetic wave.

This approach using the description of the electron motion in the undulator, energy exchange, properly described with the pendulum equation enables us to explain the laser gain, saturation, and coherent electron beam modulation. It is described, complete with informative phase space plots, in the textbook [84]. A tutorial is also available [85].

4.3.2 Ponderomotive phase

Let's consider a plane wave travelling in the same direction as the electron, with its electric field in the trajectory plane. This wave can be the stored spontaneous emission in the optical cavity. The electrons in a vertical magnetic field of a planar undulator are submitted to the electric field $\overrightarrow{E_l}$ given by $\overrightarrow{E_l} = E_l \cos(ks - \omega t + \phi)\overrightarrow{e_s}$ propagating along the direction s, with ϕ the phase of the monochromatic plane wave with respect to each single electron. The work ΔW between the times t_1 and t_2 is given by

$$\Delta W = e\int_{t_1}^{t_2} c\overrightarrow{\beta}.\overrightarrow{E_l}\mathrm{d}t. \tag{35}$$

The energy exchange only takes place via the transverse component of the velocity, so the vertical magnetic field efficiently couples the electrons and the radiation. Using $\beta_x = \frac{K_u}{\gamma}\sin(k_u s)$:

$$\Delta W = \int_{t_1}^{t_2} \frac{eK_u E_l}{\gamma}\sin(k_u s)\cos(ks - \omega t + \phi)\mathrm{d}t. \tag{36}$$

Using $s = c\beta_s t$, $k_u s = \omega_u \beta_s t$, $ks = \omega\beta_s t$, and defining $\Delta\Omega_1 = (\omega_u\beta_s + \omega(1-\beta_s))$ and $\Delta\Omega_2 = (\omega_u\beta_s - \omega(1-\beta_s))$, after some trigonometry we have: $\Delta W = \int_{t_1}^{t_2} \frac{eK_u E_l}{2\gamma}(\sin(\Delta\Omega_1 t + \phi) + \sin(\Delta\Omega_2 t - \phi))\mathrm{d}t$. A wave beating with two frequencies $\omega_u\beta_s \pm \omega(1-\beta_s)$ takes place.

The so-called ponderomotive phase ψ, i.e. the phase of the nth harmonic of the electron wiggles with respect to the wave, is introduced as

$$\psi = (nk_\mathrm{u} + k)s - \omega t. \tag{37}$$

The energy exchange due to the ponderomotive phase ψ is developed at first order, resulting in $\Delta W = -\int_{t_1}^{t_2} \frac{eK_\mathrm{u}E_\mathrm{l}}{2\gamma}(\sin(\psi + \phi - (n-1)k_\mathrm{u}s) - \sin(\psi + \phi - (n+1)k_\mathrm{u}s)\mathrm{d}t$.

Provided that the energy changes slowly compared with the period of an undulator, the longitudinal motion can be expressed as the sum of the fast term along the s direction at twice the pulsation and the slow evolution $\tilde{s}$ caused by the FEL interaction, according to $s = \tilde{s} + s_w$ where the mean motion satisfies $\langle\beta_\mathrm{s}\rangle = 1 - \frac{1}{2\gamma^2}(1 + \frac{K_\mathrm{u}^2}{2})$ in which the energy γ varies along the length of the undulator. To first order in $\frac{1}{\gamma^2}$, the longitudinal oscillation can be written as $s_w = \frac{K_\mathrm{u}^2\lambda_\mathrm{u}}{16\pi\gamma^2}\sin(2\omega_\mathrm{u}t) = \frac{K_\mathrm{u}^2}{8\gamma^2 k_\mathrm{u}}\sin(2k_\mathrm{u}\tilde{s})$. The phase can be expressed as

$$\psi = \zeta + \psi_w, \tag{38}$$

with $\psi_w = \frac{K_\mathrm{u}^2 k}{8\gamma^2 k_\mathrm{u}}\sin(2k_\mathrm{u}\tilde{s})$ and $\zeta = (nk_\mathrm{u} + k)\tilde{s} - \omega t$. The ponderomotive phase evolves as

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = (nk_\mathrm{u} + k)\frac{\mathrm{d}\tilde{s}}{\mathrm{d}t} - \omega = (nk_\mathrm{u} + k)\left(1 - \frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}\right)c - kc = nk_\mathrm{u} - nk_\mathrm{u}\left(\frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}\right)c - k\frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}c.$$

Since $nk_\mathrm{u} \ll k$, it becomes

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = nk_\mathrm{u}c - k\frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}c.$$

Then ψ given by $\psi = \zeta + \psi_w$ is inserted into the energy exchange expression, leading to

$$\Delta W = -\int_{t_1}^{t_2} \frac{eK_\mathrm{u}E_\mathrm{l}}{2\gamma}(\sin(\zeta + \psi_w + \phi - (n-1)k_\mathrm{u}\tilde{s}) - \sin(\zeta + \psi_w + \phi - (n+1)k_\mathrm{u}\tilde{s})\mathrm{d}t. \tag{39}$$

4.3.3 FEL resonance

The plane wave travelling in the same direction as the electron is shown in Fig. 21. The electron is resonant with the light wave of wavelength λ_r if, when the electron progresses by λ_u, the wave has travelled $\lambda_\mathrm{u} + \lambda_\mathrm{r}$ or more generally, with n being an integer, by $\lambda_\mathrm{u} + n\lambda_\mathrm{r}$.

The travel times of the electron and the photon can be written as $\frac{\lambda_\mathrm{u}}{v_\mathrm{s}} = \frac{\lambda_\mathrm{u} + n\lambda_\mathrm{r}}{c}$, so

$$\lambda_\mathrm{r} = \frac{\lambda_\mathrm{u}}{n}\left(\frac{1}{\beta_\mathrm{s}} - 1\right) = \frac{\lambda_\mathrm{u}}{n}\Big(\frac{1-\beta_\mathrm{s}}{\beta_\mathrm{s}}\Big) = \lambda_\mathrm{u}\frac{1-\beta_\mathrm{s}^2}{\beta_\mathrm{s}(1+\beta_\mathrm{s})}.$$

In the planar undulator case, with $\beta_\mathrm{s} \approx 1$ and $1 - \beta_\mathrm{s}^2 = \frac{1}{\gamma^2} + \frac{K_\mathrm{u}^2}{\gamma^2}$, the resonant wavelength becomes

$$\lambda_\mathrm{r} = \frac{\lambda_\mathrm{u}}{2n\gamma^2}(1 + \frac{K_\mathrm{u}^2}{2}) \tag{40}$$

and in the helical undulator case

$$\lambda_\mathrm{r} = \frac{\lambda_\mathrm{u}}{2n\gamma^2}(1 + K_\mathrm{u}^2). \tag{41}$$

The infrared spectral range can be reached with reasonable beam energies. The resonance can also be scanned either by changing the electron beam energy or by modifying the magnetic field of the undulator.

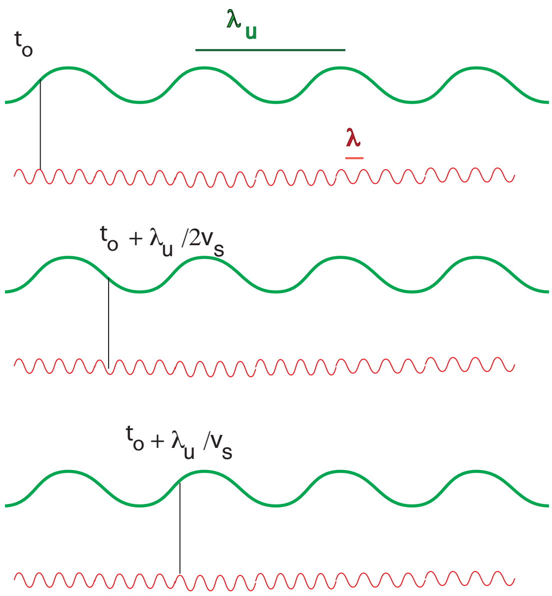


Fig. 21: Undulator resonance condition: when the electron progresses by λ_u, the wave has travelled by $\lambda_\mathrm{u} + \lambda$, t_o is the time origin, v_s is the longitudinal velocity of the electrons.

The resonance is generalized for the electron phase to be stationary $\frac{\mathrm{d}\zeta}{\mathrm{d}t} = 0$, it leads to the expression of the resonant energy given by the undulator γ_r,

$$\gamma_\mathrm{r}^2 = \frac{1}{2n}\left(1 + \frac{K_\mathrm{u}^2}{2}\right)\frac{\lambda_\mathrm{u}}{\lambda} = \frac{1}{2n}\left(1 + \frac{K_\mathrm{u}^2}{2}\right)\frac{k}{k_\mathrm{u}}. \tag{42}$$

One considers electrons with a relative energy difference with respect to the resonance given by

$$\eta = \frac{\gamma - \gamma_\mathrm{r}}{\gamma}. \tag{43}$$

4.3.4 Pendulum equations

Let us consider now the phase given by

$$\zeta = (nk_\mathrm{u} + k)\tilde{s} - \omega t. \tag{44}$$

It evolves as

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = (nk_\mathrm{u} + k)\frac{\mathrm{d}\tilde{s}}{\mathrm{d}t} - \omega = (nk_\mathrm{u} + k)\left(1 - \frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}\right)c - kc = nk_\mathrm{u} - nk_\mathrm{u}\left(\frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}\right)c - k\frac{1 + \frac{K_\mathrm{u}^2}{2}}{2\gamma^2}c.$$

Since $nk_\mathrm{u} \ll k$, it becomes $\frac{d\zeta}{\mathrm{d}t} = nk_\mathrm{u}c - k\frac{1+\frac{K_\mathrm{u}^2}{2}}{2\gamma^2}c$. Then, using $1 + \frac{K_\mathrm{u}^2}{2} = \gamma_\mathrm{r}^2 2n\frac{k_\mathrm{u}}{k_r}$, one finds

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = nk_\mathrm{u}c - nk_\mathrm{u}c\frac{\gamma_\mathrm{r}^2}{\gamma^2} = nk_\mathrm{u}c\left(1 - \frac{\gamma_\mathrm{r}^2}{\gamma^2}\right) = nk_\mathrm{u}c\frac{(\gamma - \gamma_\mathrm{r})(\gamma + \gamma_\mathrm{r})}{\gamma^2} = 2nk_\mathrm{u}c\frac{(\gamma - \gamma_\mathrm{r})}{\gamma}.$$

It then becomes

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = 2nk_\mathrm{u}c\eta. \tag{45}$$

With the Lorentz equation $\frac{d\gamma}{dt} = \frac{e\vec{E}.\vec{v}}{m_o c^2}$, one gets

$$\frac{d\gamma}{dt} = -\frac{eE_l K_u}{2\gamma m_o c}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]\sin(\zeta+\phi).$$

So

$$\frac{d\eta}{dt} = -\frac{eE_l K_u}{2\gamma^2 m_o c}[JJ]\sin(\zeta+\phi). \tag{46}$$

Combining the two equations:

$$\frac{d^2\zeta}{dt^2} = 2nk_u c\frac{d\eta}{dt} = -\frac{2nk_u ceE_l K_u}{2\gamma^2 m_o c}[JJ]\sin(\zeta+\phi) = \frac{ne^2 E_l B_u}{\gamma^2 m_o^2 c}[JJ]\sin(\zeta+\phi).$$

Noting that

$$\Omega = \frac{e}{\gamma m_o}\sqrt{\frac{nE_l B_u\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]}{c}} \tag{47}$$

we find that

$$\frac{d^2\zeta}{dt^2} = -\Omega^2\sin(\zeta+\phi). \tag{48}$$

There is a close analogy with the pendulum equation $\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin\theta = 0$, where g is acceleration due to gravity, ℓ is the length of the pendulum, and θ is the displacement angle. The analogy of θ is ψ, the phase of an electron with respect to the superposition of the optical and undulator fields. The pendulum equation is a non-linear differential equation, with an analytic solution using time-dependent elliptical Jacobi functions.

The electrons are submitted to the free electron laser sinusoidal ponderomotive potential given by $-\Omega\cos\psi$. It has the form of the potential of a pendulum, in which ψ is the angle of the pendulum at its equilibrium position. One usually represents the energy evolution in the energy phase space, as illustrated in Fig. 22.

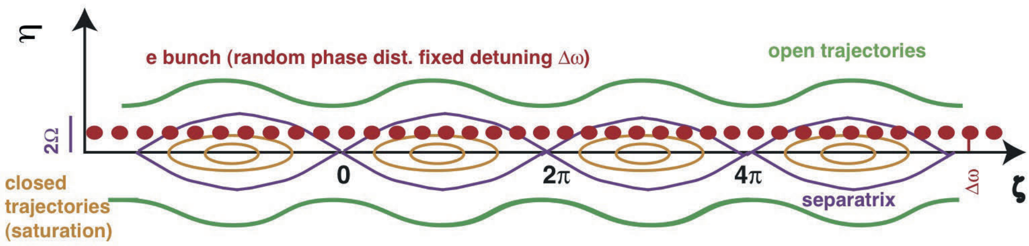


Fig. 22: Electron trajectories in energy phase space representation. The vertical axis represents the deviation with respect to resonance, the horizontal axis the electron phase with respect to the ponderomotive potential. Green: open trajectories with energy oscillations. Orange: closed trajectories of particles by the ponderomotive potential. Maximum kinetic energy is given to/taken from the optical wave for half a rotation, i.e. for highest and lowest positions.

The initial phase of the electron is simply given by its position along the electron bunch, this determines its energy variation and thus its bunching. The electrons enter the undulator with a specific phase. On resonance $\gamma = \gamma_r$, i.e. $\eta = 0$, there is no energy transfer. Near resonance, the optical wave and the electrons exchange energy, the electrons gather around positions for which the energy variation $\delta\gamma m_o c^2$ keeps a constant sign. The modulation depends on the electric field of the wave. Above resonance ($\gamma > \gamma_r$), there is a net positive energy transfer from the electron beam to the optical wave. Positive energy exchange (gain) or negative one (absorption) occurs $\eta \neq 0$ depending on the sign of η.

At small amplitudes, with the approximation $\sin\psi = \psi$, one gets the harmonic oscillator case. The equation can be analytically solved only assuming Ω small, i.e. for low amplification since $\Omega \propto \sqrt{E_1}$. With increasing angular momentum, the motion becomes unharmonic. The trajectories are closed, inside the 'bucket'. The closed motions correspond to oscillations of the pendulum. When trapped in the ponderomotive field $-\Omega^2\cos\phi$, the particles bounce back and forth on the borders of the potential, and rotate in phase space. Trapped particles undergo oscillations in the buckets of the potential. The closed trajectories in phase space correspond to those of an oscillating pendulum around its equilibrium position.

Above the peaks of the potential, at very large angular momentum, the motion becomes unbounded, the trajectories are open (green), and the movement corresponds to a complete rotation of the pendulum around its pivot. The particles can follow open trajectories from one potential well to another: they present oscillations in energy and an evolving phase. They can also be trapped in the ponderomotive field $-\Omega^2\cos\phi$, and rotate in phase space.

4.3.5 First-order energy exchange and bunching

4.3.5.1 First-order energy exchange

In the case of the optical wave resonant to the undulator wavelength, i.e. if the pulsation of the incident wave is equal to the resonant wavelength, which means for $\omega = \frac{\beta_s\omega_u}{1-\beta_s}$, $\Delta\Omega_2 = 2\omega_u\beta_s$ and $\Delta\Omega_1 = 0$. The work integrated over one undulator period, i.e. between $t_1 = 0$ and $t_2 = \frac{\lambda_u}{\beta_s c}$ is

$$\Delta W = -\int_{t_1=0}^{t_2=\frac{\lambda_u}{\beta_s c}} \frac{eK_u E_o}{2\gamma}(\sin(\phi) + \sin(2\beta_s\omega_u t - \phi))\mathrm{d}t = 0.$$

The work due to the force applied by the electric field averaged over one undulator period is zero. For $\lambda = \lambda_r$, there is no average energy exchange at first order: half of the electrons gain energy, half of them loose energy.

In the case of an optical wave slightly detuned with respect to the undulator wavelength, using $\frac{k}{k_r} = 4n\gamma_r^2\frac{1}{2+K_u^2}$, the oscillatory term of the phase becomes

$$\psi_w = \frac{K_u^2}{8\gamma^2}4n\gamma_r^2\frac{1}{2+K_u^2}\sin(2k_u\tilde{s}) = n\left(\frac{\gamma_r}{\gamma}\right)^2\frac{K_u^2}{4+2K_u^2}\sin(2k_u\tilde{s}) = n\xi\sin(2k_u\tilde{s}),$$

where ξ is defined by

$$\xi = \frac{K_u^2}{2(2+K_u^2)}\left(\frac{\gamma_r}{\gamma}\right)^2. \tag{49}$$

The phase then becomes

$$\psi = \zeta + n\xi\sin(2k_u\tilde{s}). \tag{50}$$

The electron phase ζ contains only the slowly varying part of the s motion $\tilde{s}$, the second term corresponds to the rapidly oscillatory term. In replacing this new expression of the phase in the energy exchange expression, the corresponding energy exchange term can be written as

$$\Delta W = -\int_{s_1}^{s_2}\frac{eK_u E_l}{2\gamma}\sin(\zeta + \phi - (n-1)k_u\tilde{s} + n\xi\sin(2k_u\tilde{s})) \\ - \sin(\zeta + \phi - (n+1)k_u\tilde{s} + n\xi\sin(2k_u\tilde{s}))\mathrm{d}t.$$

On expanding the sines, it becomes

$$\begin{cases} \sin\left(\zeta+\phi-(n-1)k_{\mathrm{u}}s+n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)\right) \\ =\sin\left(\zeta+\phi\right)\cos\left(n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)-(n-1)k_{\mathrm{u}}\tilde{s}\right)+\cos\left(\zeta+\phi\right)\sin\left(n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)-(n-1)k_{\mathrm{u}}\tilde{s}\right) \\ \quad -\sin\left(\zeta+\phi-(n+1)k_{\mathrm{u}}s+n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)\right) \\ =-\sin\left(\zeta+\phi\right)\cos\left(n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)-(n+1)k_{\mathrm{u}}\tilde{s}\right)-\cos\left(\zeta+\phi\right)\sin\left(n\xi\sin\left(2k_{\mathrm{u}}\tilde{s}\right)-(n+1)k_{\mathrm{u}}\tilde{s}\right). \end{cases}$$

Assuming that the energy γ and the light wave electric field E_{l} change slowly, one can average the oscillating terms over one undulator period. The second and fourth terms then vanish by symmetry. The average of the first and third terms is performed using the integral representation of the Bessel functions of order m and of variable z [86] given by

$$J_k(z)=\frac{1}{2\pi}\int_0^{2\pi}\cos\left(z\sin\theta-k\theta\right)\mathrm{d}\theta. \tag{51}$$

For k equal to half an integer, the integral vanishes by symmetry. Using $\theta=2k_{\mathrm{u}}\tilde{s}$, $z=n\xi$, and $m=\frac{n-1}{2}$ or $m=\frac{n+1}{2}$, it becomes

$$\Delta\gamma=-\frac{eE_{\mathrm{l}}K_{\mathrm{u}}N_{\mathrm{u}}\lambda_{\mathrm{u}}}{2\gamma m_{\mathrm{o}}c^2}\left[J_{\frac{n-1}{2}}(\xi)-J_{\frac{n+1}{2}}(\xi)\right]\sin\left(\zeta+\phi\right). \tag{52}$$

This expression gives zero for even m values. This recalls the vanishing of the even harmonics of the spontaneous emission on the axis, while considering that the electron beam average over one period is parallel to the undulator axis. In the slow varying phase ϕ and electric field E_{l} approximation, these functions can be estimated in using their values for $\tilde{s}$.

Besides, the homogeneous width of the spontaneous emission is given by $\frac{1}{nN_{\mathrm{u}}}$. This spontaneous emission width provides also the non-linear interaction region in the frequency space. By differentiating the resonance equation, one gets

$$\frac{\delta\gamma}{\gamma}=\frac{1}{2}\frac{\delta\lambda}{\lambda}=0\left(\frac{1}{2nN_{\mathrm{u}}}\right). \tag{53}$$

Only energies within a relative difference $\frac{1}{2nN_{\mathrm{u}}}$ can play a role. At a low-order derivation, the energy exchange can be written as

$$\Delta\gamma=-\frac{eE_{\mathrm{l}}K_{\mathrm{u}}N_{\mathrm{u}}\lambda_{\mathrm{u}}}{2\gamma_{\mathrm{r}}m_{\mathrm{o}}c^2}\left[J_{\frac{n-1}{2}}(\xi)-J_{\frac{n+1}{2}}(\xi)\right]\mathrm{sinc}(\pi N_{\mathrm{u}}\eta)\sin\left(\zeta+\phi\right), \tag{54}$$

with η the relative energy difference and ξ given by $\xi=\frac{K_{\mathrm{u}}^2}{2(2+K_{\mathrm{u}}^2)}$.

The sign of the $\Delta\gamma$ depends on the phase $\zeta+\phi$ between the electron and the optical wave. If one electron is accelerated, i.e. for $\Delta\gamma>0$, another electron located longitudinally one-half wavelength ahead or behind is decelerated by the same amount $\Delta\gamma<0$. The longitudinal distribution of the electrons being much wider than the wavelength, the phase ϕ is uniformly distributed between 0 and 2π. In consequence, the first-order net energy exchange $\langle\Delta\gamma_1\rangle_{\mathrm{electrons}}$ between the electron bunch and the optical wave is zero over the electron bunch:

$$\langle\Delta\gamma_1\rangle_{\mathrm{electrons}}=0. \tag{55}$$

For the interaction to occur, λ should be slightly different from λ_{r}: for $\lambda>\lambda_{\mathrm{r}}$ amplification occurs (gain and beam deceleration) whereas for $\lambda<\lambda_{\mathrm{r}}$ the optical wave is absorbed (the beam is accelerated).

4.3.5.2 *Root-mean-square energy variation*

The root-mean-square (RMS) energy variation, averaged over the electrons, can be expressed as

$$\langle \Delta\gamma^2 \rangle = \frac{1}{2}\left(\frac{eK_\mathrm{u}N_\mathrm{u}\lambda_\mathrm{u}}{2\gamma_\mathrm{r}m_\mathrm{o}c^2}\right)^2 \langle E_\mathrm{l}\rangle^2 \left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]^2 \mathrm{sinc}^2(\pi N_\mathrm{u}\eta) \tag{56}$$

where $\langle E_\mathrm{l}\rangle^2$ is the average of the square of the electric field over the electron beam. The electrons, after propagation, are then accelerated or decelerated by energy enhancement or loss. This leads to a longitudinal spatial modulation, known as 'electron bunching' or 'electron microbunching'. Electrons are bunched around a phase $\psi + \phi$, multiple of 2π. The electronic density is then modulated with a period equal to the resonant wavelength. The electrons are put in phase, the elementary oscillators are set in coherence. This bunching is similar to the one taking place in the klystron, as introduced earlier. On a planar undulator, the bunching also occurs at the odd harmonics of the resonant wavelength. This is the basic concept for 'coherent harmonic generation' [81–83].

4.3.5.3 *Ponderomotive field*

Considering the energy exchange given by $\frac{\mathrm{d}\gamma}{\mathrm{d}t} = -\frac{eE_\mathrm{l}K_\mathrm{u}}{2\gamma m_\mathrm{o}c}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]\sin(\zeta+\phi)$ and in considering the analogy with the interaction of an electron with an axial electric field according to $\frac{\mathrm{d}\gamma}{\mathrm{d}t} = -\frac{e}{m_\mathrm{o}c}\beta_\mathrm{s}E_\mathrm{p}$ with E_p the so-called ponderomotive field, one gets

$$\langle E_\mathrm{p}\rangle = -\frac{E_\mathrm{l}K_\mathrm{u}[JJ]}{2\gamma}\sin(\zeta+\phi) \quad \text{with} \quad [JJ] = \left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]. \tag{57}$$

The corresponding electron potential is $V_\mathrm{p} = e\int_0^t \langle E_\mathrm{p}\rangle \mathrm{d}s'$. The electrons behave as though they were particles in a sinusoidal potential given by $-\Omega\cos\psi$, or the so-called ponderomotive potential of the free electron laser. When in the potential the particles bounce back and forth on the borders of the potential. Particles undergo trapped oscillations in the buckets of the potential, as shown in Fig. 23.

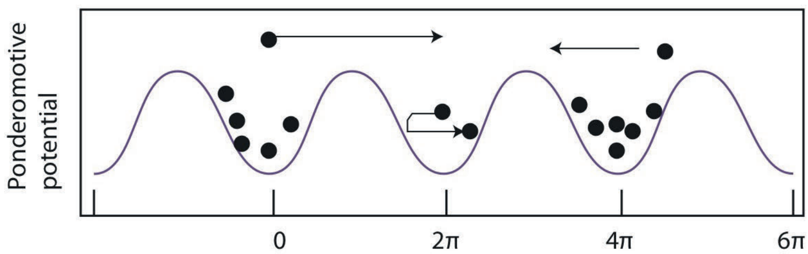


Fig. 23: Ponderomotive potential

4.3.5.4 *Bunching process*

Some electrons gain energy, others loose energy. The average longitudinal velocity is changing along the propagation in the undulator. From $\tilde{\beta_\mathrm{s}}^2 = 1 - \frac{1+\frac{K_\mathrm{u}^2}{2}}{\gamma^2}$, one gets $\Delta\beta_\mathrm{s} \approx (1+\frac{K_\mathrm{u}^2}{2})\frac{\Delta\gamma}{\gamma^3}$.

The energy variation averaged over the phases is zero at first order in E_l. An individual electron with a phase ϕ gains or loses energy, so its position relative to the unperturbed $s = \tilde{v}t$ position is advanced or retarded. Because the amplitude of the interaction only depends on the longitudinal position of the electron in the electron bunch with periodicity λ_r, the electrons tend to bunch along given positions, separated by λ_r. This bunching (λ_r separation) takes place by velocity modulation (electrons set in phase). As for the klystron, the electrons tend to gather around preferred positions separated by the resonant wavelength λ_r.

One first replaces the energy exchange in the longitudinal velocity variation for $n = 1$, and we get

$$\Delta\beta_s = -D(\cos(\Delta\Omega_2 s/\tilde{v}_s + \phi) - \cos\phi) \quad \text{with} \quad D = \frac{(1+\frac{K_u^2}{2})}{\gamma^4}\frac{eE_l K_u[JJ]}{2\, m_o c\Delta\Omega_2} \tag{58}$$

with $\Delta\Omega_2 = \omega_u\beta_s - \omega(1-\beta_s)$. Then one evaluates the longitudinal position as

$$s(t) = \int_0^t v_s(t')\mathrm{d}t' = \int_0^t (\tilde{v}_s + c\Delta\beta_s)\mathrm{d}t' = \tilde{v}_s t - cD\int_0^t (\cos(\Delta\Omega_2 t' + \phi) - \cos\phi)\mathrm{d}t'$$

$$s(t) = \int_0^t v_s(t')\mathrm{d}t' = \tilde{v}_s t - cD\left[\frac{(\sin(\Delta\Omega_2 t + \phi) - \sin\phi)}{\Delta\Omega_2} - t\cos\phi\right]. \tag{59}$$

One finds here the longitudinal bunching, as seen in the klystron case. It is illustrated in Fig. 24.

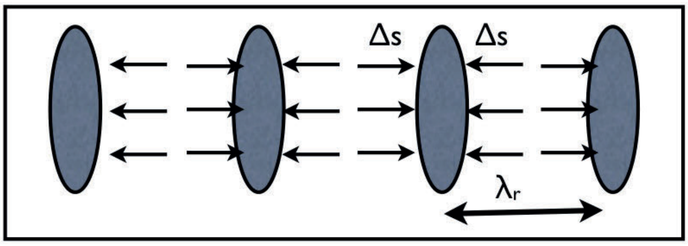


Fig. 24: Electron bunching due to the electron/optical wave interaction

One uses now the expression of the longitudinal position of the electron in the electric field expression. It becomes

$$E_l(t) = E_l\cos(ks - \omega t + \phi) = E_l\cos\left(k\tilde{v}_s t - \frac{kcD}{\Delta\Omega_2}[\sin(\Delta\Omega_2 t + \phi) - \sin\phi - \Delta\Omega_2 t\cos\phi] - \omega t + \phi\right). \tag{60}$$

One defines the phase slippage $\Delta\phi$ as

$$\Delta\phi = -\frac{\omega D}{\Delta\Omega_2}[\sin(\Delta\Omega_2 t + \phi) - \sin\phi - \Delta\Omega_2 t\cos\phi] \tag{61}$$

$\Delta\Omega_2$ is given by $\Delta\Omega_2 = \omega_u\beta_s - \omega(1-\beta_s)$. Using $1 - \langle\beta_s\rangle = \frac{1}{2\gamma^2}(1+\frac{K_u^2}{2}) = \frac{\lambda_r}{\lambda_u} = \frac{\omega_u}{\omega_r}$, i.e. $\omega_u = \omega_r(1-\beta_s)$, it becomes $\Delta\Omega_2 = \omega_u\beta_s - \omega\frac{\omega_u}{\omega_r}$. With $\beta_s \approx 1$

$$\Delta\Omega_2 = \omega_u\left(1 - \frac{\omega}{\omega_r}\right). \tag{62}$$

Considering this electron bunching will enable us to evaluate the second-order energy exchange.

4.3.6 Second-order energy exchange

The second-order energy exchange is calculated using in the energy exchange expression the electric field expression taking into account the density modulation of the electron beam.

For a low electric field E_l and low gain, $\Delta\phi$ is close to 0, and one develops $\sin(\Delta\Omega_2 t - \phi - \Delta\phi) = \sin(\Delta\Omega_2 t - \phi) - \Delta\phi\cos(\Delta\Omega_2 t - \phi)$. One then averages over all phases ϕ and it remains as

$$\langle\frac{\mathrm{d}\gamma}{\mathrm{d}t}\rangle_\phi = \frac{b}{2}(-\sin(\Delta\Omega_2 t) + \Delta\Omega_2 t\cos(\Delta\Omega_2 t)) \quad \text{with} \quad b = -\frac{eE_\mathrm{l}K_\mathrm{u}[JJ]\omega D}{2\gamma m_\mathrm{o} c\Delta\Omega_2}. \tag{63}$$

In integrating over the electron transit time through the undulator $\tau = L_\mathrm{u}/\tilde{v}_\mathrm{s}$, one obtains the second-order energy change per electron:

$$\langle\Delta\gamma_2\rangle_\phi = \int_0^{\tau=L_\mathrm{u}/\tilde{v}_\mathrm{s}} \langle\frac{\mathrm{d}\gamma}{\mathrm{d}t}\rangle_\phi \mathrm{d}t = \frac{b}{2\Delta\Omega_2}(2 - 2\cos\Delta\Omega_2\tau - \Delta\Omega_2 t\sin(\Delta\Omega_2\tau)).$$

By multiplying by τ^3, replacing $(1 + K_\mathrm{u}^2/2) = 2\gamma^2\lambda/\lambda_\mathrm{u}$, b, and D, one gets

$$\langle\Delta\gamma_2\rangle_\phi = \frac{e^2\pi}{2m_\mathrm{o}^2c^4}\frac{K_\mathrm{u}^2}{\lambda_\mathrm{u}}[JJ]^2E_\mathrm{l}^2\frac{L_\mathrm{u}^3}{\gamma^3}\frac{(2 - 2\cos\Delta\Omega_2\tau - \Delta\Omega_2\tau\sin(\Delta\Omega_2\tau))}{(\Delta\Omega_2\tau)^3}. \tag{64}$$

Let us define the function $g(x)$ by $g(x) = \frac{2-2\cos x - x\sin x}{x^3}$. The function $g(x)$ is antisymmetric in x and has a maximum of 0.135 at $x = 2.6$.

Multiplying by τ^3, replacing $(1 + K_\mathrm{u}^2/2) = 2\gamma^2\lambda/\lambda_\mathrm{u}$, b, and D, one gets

$$\langle\Delta\gamma_2\rangle_\phi = \frac{e^2\pi}{2m_\mathrm{o}^2c^4}\frac{K_\mathrm{u}^2}{\lambda_\mathrm{u}}[JJ]^2E_\mathrm{l}^2\frac{L_\mathrm{u}^3}{\gamma^3}\frac{(2 - 2\cos\Delta\Omega_2\tau - \Delta\Omega_2\tau\sin(\Delta\Omega_2\tau))}{(\Delta\Omega_2\tau)^3}. \tag{65}$$

4.3.7 Gain

4.3.7.1 Gain expression in the low gain regime

The optical wave is the FEL spontaneous emission given by the synchrotron radiation emitted by the electrons passing through N_u periods of the undulator and stored in an optical cavity, as shown in Fig. 25, where the electron bunching is indicated. The mirrors of the optical resonator perform the optical feedback, such that the light pulse performs multiple passes in the cavity. The gain is evaluated for small variations of the optical field.

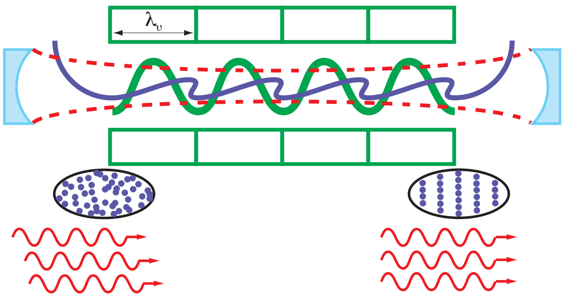


Fig. 25: FEL configuration with an optical cavity: the energy exchange between the optical light (initially the spontaneous emission stored in the optical cavity) and the electrons is then transformed into density modulation while the electron progress is due to velocity bunching. This results in a microbunching of the electrons which can then emit coherently in phase with the optical wave that gets amplified.

The first-order energy exchange averaged over the electrons is zero. The second-order energy exchange $\langle\Delta\gamma_2\rangle$ averaged over phases with the bunched electron distribution has been calculated. The change in electromagnetic power ΔP is given by

$$\Delta P = -\frac{I}{e} m_{\mathrm{o}} c^2 \langle \Delta\gamma_2 \rangle. \tag{66}$$

For a small variation of the optical field, the gain G per pass can be expressed as the second-order energy exchange divided by the incident field energy, according to

$$G = \frac{m_{\mathrm{o}} c I \langle \Delta\gamma_2 \rangle}{e\epsilon_{\mathrm{o}} \int E_1^2 \mathrm{d}S} = \frac{m_{\mathrm{o}} c^2 \rho_e \langle \Delta\gamma_2 \rangle}{\frac{1}{2}\epsilon_{\mathrm{o}} E_1^2} \tag{67}$$

where ϵ_{o} is the vacuum permeability, ρ_e the electronic density in the volume, the radiation field being integrated over the longitudinal coordinates. The small signal gain is given by

$$G = \frac{2\pi e^2}{\epsilon_{\mathrm{o}} m_{\mathrm{o}} c^2} \rho_e \frac{K_{\mathrm{u}}^2}{\lambda_{\mathrm{u}}} \left(\frac{L_{\mathrm{u}}}{\gamma}\right)^3 \left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]^2 \frac{\partial}{\partial\gamma} \left(\frac{\sin(\pi N_{\mathrm{u}} \eta)}{(\pi N_{\mathrm{u}} \eta)}\right)^2 . \tag{68}$$

For the interaction to occur, λ should be slightly different from λ_{r}. For $\lambda > \lambda_{\mathrm{r}}$ amplification occurs (gain and beam deceleration) whereas for $\lambda < \lambda_{\mathrm{r}}$ the optical wave is absorbed and the beam is accelerated. Depending on the sign of $(\lambda - \lambda_{\mathrm{r}})$, the optical wave is either absorbed to the benefit of a gain of kinetic energy of the electrons, or is amplified to the detriment of the kinetic energy of the electrons. The electrons are bunched and are in phase with the incident electric field. The emission from the bunched beam then adds coherently to the incident wave that gets amplified.

The small signal gain varies as $1/\gamma^3$. The higher the energy, the lower the gain. Since short wavelength operation requires the use of high electron beam energies (because of the resonance condition), for the same undulator length, the gain is smaller at short wavelengths than at longer ones. The gain is proportional to the electronic density (thus to the beam current I). The more electrons interact, the larger the gain. For short wavelength FELs where the gain is naturally small, one should employ beams with high electronic densities.

The gain is also proportional to the third power of the undulator length. The longer the undulator, the higher the gain up to certain limits that are given by the gain bandwidth ($1/nN_{\mathrm{u}}$, because of the interference nature of the interaction), and by the slippage (temporally, the light pulse should remain in the longitudinal bunch distribution for the interaction to occur). So the number of undulator periods cannot be excessively large. Similarly, both the optical light and electron bunch should overlap properly all long the undulator propagation.

4.3.7.2 Madey's theorems

Remarkably, the gain is the derivative of the spontaneous emission, as understood thanks to the Madey theorems [60, 87]. They are given by

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\Omega}(\theta = 0) = \frac{2\alpha m_{\mathrm{o}}^2 c^4 I \langle \Delta\gamma^2 \rangle}{e^2 \lambda^2 \langle E_1^2 \rangle}, \tag{69}$$

$$\langle \Delta\gamma_2 \rangle = \frac{1}{2} \frac{\partial \langle \Delta\gamma^2 \rangle}{\partial\gamma}, \tag{70}$$

with α the fine structure constant, I the beam current, $\frac{\mathrm{d}\Phi}{\mathrm{d}\Omega}(\theta = 0)$ the angular spectral flux on axis of the undulator spontaneous emission. The first theorem relates the energy spread $\langle E_1^2 \rangle$ introduced by the optical wave to the spontaneous emission of the undulator. According to the second theorem, the second-order energy exchange $\langle \Delta\gamma_2 \rangle$ is proportional to the derivative of the spontaneous emission of the undulator. Due to the resonance relationship linking the particles energy to the emission wavelength, the spectral 'gain' distribution is close to the wavelength derivative of the spontaneous emission spectrum versus λ. The Madey theorem is valid for a gain smaller than 0.2. The gain can be expressed as the derivative of the undulator spontaneous emission.

4.3.7.3 Gain correction terms

Gain corrections terms should be introduced.

The transverse filling factor F_f accounts for a non-perfect transverse overlap between the laser transverse modes and the transverse dimensions of the electron beam σ_x and σ_z. For a laser of TEM_{00} mode of waist w_o, it is given by $F_\mathrm{f} = \frac{1}{\sqrt{1+(\frac{w_\mathrm{o}}{2\sigma_x})^2}\sqrt{1+(\frac{w_\mathrm{o}}{2\sigma_\mathrm{z}})^2}}$. The filling factor has been calculated using Gaussian spherical wavefronts of the optical wave, leading to a deviation from the Madey's theorems, and a new optimization of the energy extraction [88], as shown in Fig. 26.

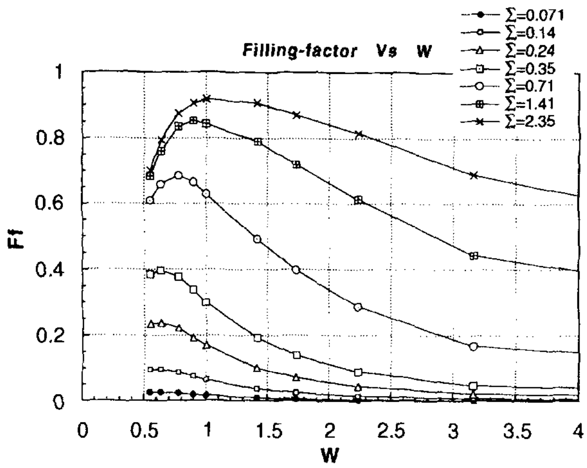


Fig. 26: Filling factor in the case for Gaussian optical beams for different values of the normalized waist $W_i = w_i\sqrt{\frac{\pi}{\lambda L_\mathrm{u}}}$ and electron beam transverse sizes $\Sigma_i = \sigma_\mathrm{i}\sqrt{\frac{\pi}{\lambda L_\mathrm{u}}}$. from [89].

Besides, according to the Madey's theorem, spontaneous emission inhomogeneous broadening (presented in 2.3.3.2) due to energy spread and emittance affect directly the gain. The inhomogeneous reduction factor F_inh is

$$F_\mathrm{inh} = \left[1 + \frac{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_{\sigma_\gamma}}{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_\mathrm{hom}}\right]^{-1} . \left[1 + \frac{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_\mathrm{div}}{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_\mathrm{hom}}\right]^{-1} . \left[1 + \frac{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_\sigma}{(\frac{\Delta\lambda}{\lambda_\mathrm{n}})^2_\mathrm{hom}}\right]^{-1} .$$

The longitudinal overlap between the electron bunch of RMS length σ_l and the optical wave should be maintained. The light wave is in advance by $N_\mathrm{u}\lambda$ with respect to the electrons, and for short electron bunch distributions, it could escape. The corresponding correction factor F_g is $F_g = \left[1 + \frac{N_\mathrm{u}\lambda}{\sigma_\mathrm{l}}\right]^{-1}$.

The small signal gain can be expressed as

$$G = \frac{2\pi e^2}{\epsilon_\mathrm{o} m_\mathrm{o} c^2}\rho_e F_\mathrm{f} F_\mathrm{inh} F_g \frac{K_\mathrm{u}^2}{\lambda_\mathrm{u}}(\frac{L_\mathrm{u}}{\gamma})^3 \left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]^2 \frac{\partial}{\partial\gamma}\mathrm{sinc}^2(\pi N_\mathrm{u}\eta), \tag{71}$$

$$G = n\frac{\pi^2 r_\mathrm{o}\lambda_\mathrm{u}^2 N_\mathrm{u}^3 K_\mathrm{u}^2}{\gamma^3} F_\mathrm{f} F_\mathrm{inh} F_g \rho_e \left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]^2 \frac{\partial}{\partial\gamma}\mathrm{sinc}^2(\pi N_\mathrm{u}\eta), \tag{72}$$

with $r_\mathrm{o} = \frac{1}{4\pi\epsilon_\mathrm{o}}\frac{e^2}{m_\mathrm{o} c^2}$ the classical radius of the electron (2.82×10^{-15}).

4.4 The classical approach in the moving frame

In such an approach, developed in Italy [90, 91] in particular by Alberto Renieri (seventh FEL Prize in 1994) and Guiseppe Datttoli (seventh FEL Prize in 1994), the Weizsäcker–Williams approximation, still valid for ultra-relativistic electrons, is used. The FEL corresponds to a stimulated scattering process from the so-called 'pseudo-radiation field' into a true radiation field travelling in the same direction as the electron beam. The FEL modelled by a stimulated Thomson scattering process is described using the Hamiltonian formalism [90].

The selected frame is a moving one [91], "chosen in such a way that the periodic structure transforms into a (pseudo) radiation field whose frequency coincides with the frequency of the stimulating field". This frame choice presents different advantages.

- "The physical processes of scattering from one field to the other and vice versa become apparent. Indeed, in that frame, the two fields are treated on the same step, although they are quite different in the laboratory frame."
- "Relativistic calculations can be avoided. In fact, in that frame, the electrons have non relativistic velocity, and the momentum exchanged with the fields is not sufficient to give the electrons a relativistic velocity."
- "In the limit in which the laser operation can be described in terms of ensemble averages over independent single particles, it becomes possible to follow the history for each electron in the field with simple equations of motion." [91]

This FEL description in the moving frame reduces to the pendulum equations, which are not Lorentz invariant and are valid in that frame only. A quantum mechanical density matrix description of the system is able to conciliate this approach with the one proposed by Hopf. A Hamiltonian completes the overall description of the process.

Under the approximation that the electrons do not contribute much to the laser intensity and follow adiabatically the field, the FEL evolution can be described by the pendulum equations. Figure 27 shows the phase space plots, with W scale momentum, Ω interaction frequency, ψ interaction phase, "essentially the position coordinate canonical to W" [91].

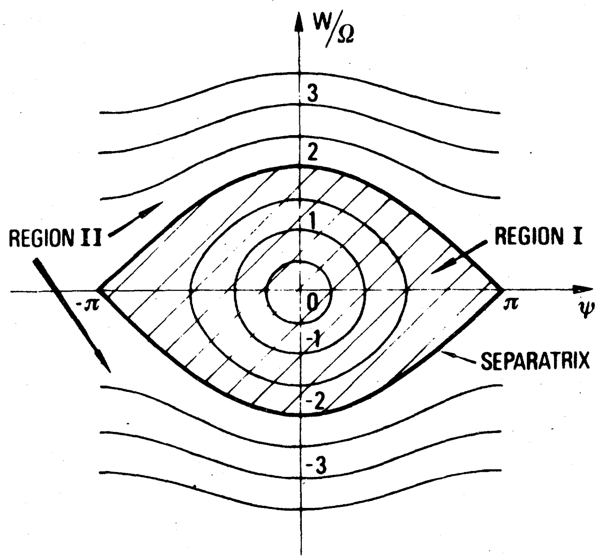


Fig. 27: Phase space plot of the FEL modelled using the pendulum equation, from [91]: W scale momentum, Ω interaction frequency, ψ interaction phase, "essentially the position coordinate canonical to W".

It shows two zones.

- Zone I with closed paths for W and ψ.
- Zone II with W with periodic trajectories while ψ increases steadily with time.
- Separatrix with both aperiodic motions of W and ψ, depending on the value of laser intensity through Ω and its evolution.

The electron motion can then be solved using the Jacobi elliptical functions.

The dimensionless laser intensity is introduced "it measures the ratio of the small-signal oscillation frequency (Rabi flipping) against the Doppler shift determined by the initial electron momentum" [91] and enables us to follow the evolution of the gain towards saturation. After this single particle classical theory, where amplification is due to the single electron stimulated Thomson scattering, the Hamiltonian description has been further examined [92] in considering the multiple electron effects [93, 94], gain-spread expressions [95], and in accounting for self-consistency. FEL pulse propagation and synchronization of the pulses in the optical resonator and the electron bunches in the resonator are examined: it is found that the 'lethargy', i.e. "the slowing down of the laser oscillation in the cavity owing to the interaction with the electrons" leads to the presence of "supermodes" [96, 97]. An equivalent refractive index can be defined. The FEL evolution has been described with the logistic function [98].

Different lectures are gathered in textbooks [99, 100].

4.5 Early FEL developments in the former Soviet Union

In Russia, which was quite isolated at that time, development took also place on the FEL. FEL progress was made independently in Russia and outside.

A meeting was held in December 1980 in the frame of the Academy of Science to discuss the development of free electron lasers [101]. Different work progress was reported: M. V. Fedorov (Lebedev Institute) on the different types of FELs, M. I. Petelin, A. A. Kolomenskii, A. A. Ruxadz (Institute of Applied Physics, Nijni-Novgorod) on the possibility of a mm range FEL on Sinus-4, A. N. Didenko (Institute of Nuclear Physics in Tomsk) on an undulator experiment on an induction linac and on the use of 'Sirius' synchrotron (500–900 MeV), N. A. Vinokurov (Institute of Nuclear Physics in Novosibirsk) on the use of the optical klystron on the VEPP3 storage ring, A. A. Varfolomeev and D. F. Zaretskii, S. P. Kapitsa (Institute of problems of Physics) on a proposition of FEL on a microtron, N. V. Karkov (Lebedev Institute in Karkhov) on the use of FEL for isotope separation at 16 μm. Prospects for short wavelength operation and high efficiency FEL were given.

In Novossibirsk, at the Institute of Nuclear Physics (presently called the Budker Institute), the team was working both on the theory and was also thinking of a test experiment for a storage ring FEL. In order to enhance the gain, N. A. Vinokurov (fourth FEL Prize in 1991) and A. N. Skrinsky proposed the optical klystron, a device to artificially enhance the gain. They investigated the maximum power that could be extracted.

4.5.1 The optical klystron

The optical klystron proposed by N. A. Vinokurov [102–105], represented in Fig. 28, is made of a first undulator creating the electron energy modulation, a dispersive section of length L_d and peak field B_d creating a wide wiggler of magnetic field enabling the energy modulation to be transformed into density modulation, and a second undulator where bunched electrons radiate. Assuming that the undulator segments and the dispersive section are well compensated, the electrons do not suffer velocity and position shifts during their travel in the device. The dispersive section acts as a magnetic chicane: the electrons are more or less deviated in the strong magnetic field according to their energy and become bunched thanks to a velocity modulation process. The concept of the optical klystron was then explored further around the world [106–108].

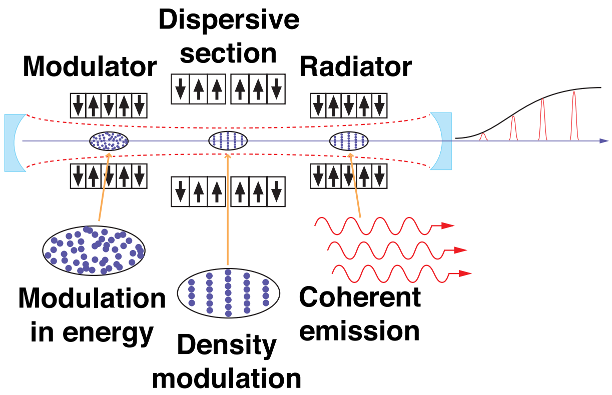


Fig. 28: Scheme of the optical klystron

The radiation emitted in the two undulator segments interfere, as in the Young slit experiment. The spectrum is contained in an envelope which corresponds to the spectral line of one single undulator emission, with an internal fine structure resulting from the interference [107, 109]. It can be expressed as

$$\left(\frac{\mathrm{d}^2 I}{\mathrm{d}\Omega \mathrm{d}\omega}\right)_{\text{optical klystron}} \approx \left(\frac{\mathrm{d}^2 I}{\mathrm{d}\Omega \mathrm{d}\omega}\right)_{\text{one undulator}} (1 + f \cos \alpha_{\text{optical klystron}}) \tag{73}$$

with

$$\alpha_{\text{optical klystron}} = 2\pi (N_{\mathrm{u}} + N_{\mathrm{d}}) \frac{\lambda_{\mathrm{r}}}{\lambda} \frac{\gamma_{\mathrm{r}}^2}{\gamma^2}. \tag{74}$$

Here

$$N_{\mathrm{d}} = \frac{\omega L_{\mathrm{d}}}{4\gamma_{\mathrm{r}}^2 c} \left[1 + \frac{e^2}{L_{\mathrm{d}} m^2 c^2} \int_0 L_{\mathrm{d}} \left[\int_0^u B_{\mathrm{d}}(s) \mathrm{d}s\right]^2 \mathrm{d}u\right] \tag{75}$$

is the equivalent number of periods of the dispersive section, and scales its strength. $(N_{\mathrm{u}} + N_{\mathrm{d}})$ represents the number of optical wavelengths which pass the electron during its travel in the dispersive section. The fringe contrast, called the modulation rate f, results from different contributions (magnetic field inhomogeneity, width of energy distribution of the electrons, transverse position of the electron beam), the main one coming from the electron beam energy spread, as

$$f_\gamma = \exp\left(-8\pi^2 (N_{\mathrm{u}} + N_{\mathrm{d}})^2 (\sigma_\gamma/\gamma)^2\right). \tag{76}$$

An example of an optical klystron spectrum is shown in Fig. 29.

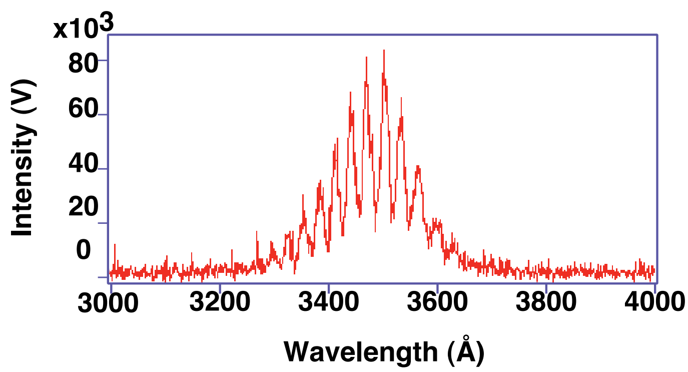


Fig. 29: Measured spectrum in the case of the Super-ACO optical klystron (Orsay, France)

The optical klystron provides a very easy means to measure the energy spread on an electron beam. Besides, the variation of the intensity in the spectra being much faster than in the single undulator case, the derivative of the spontaneous emission (proportional to the gain according to Madey's theorem) reaches much larger values than for a single undulator of total length L_{d}.

The gain enhancement for an optical klystron of length L_{ok} is

$$G_{\text{optical klystron}} = \frac{f L_{\mathrm{ok}}^2 (N_{\mathrm{u}} + N_{\mathrm{d}})}{N_{\mathrm{u}}^3 \lambda_{\mathrm{u}}} G_{\text{one undulator}}. \tag{77}$$

The gain enhancement takes place to the detriment of the total saturated power [104–106].

The concept of the multiple optical klystron was further developed [110–112].

4.5.2 The FEL evolution

Independently to Colson's approach, V. N. Baier and A. I. Milstein investigated the FEL theory in considering the motion of relativistic particles in the superposition of a transverse magnetic field and a plane electromagnetic wave propagating along the direction of motion. They could find the small signal gain, its maximum, and then considered the case of a strong signal in the optical cavity configuration. They distinguished the case of the initially uniform phase distribution to the bunched one, where phase oscillations can occur and limit the output power [113,114]. Coherent radiation close to the cyclotron resonance was also discussed [115]. A. N. Kondratenko and A. I. Saldin (19th FEL Prize in 2006) considered very early the possibility of production of coherent radiation from a self-instability, without the use of an optical resonator [116–118]. This pioneering work will be discussed in the high gain section. They also developed a linear theory of free electron lasers with Fabry–Perot cavities [119].

4.6 Saturation and efficiency

4.6.1 Saturation

Different phenomena contribute to the gain reduction leading to the saturation of the output power.

4.6.1.1 Electron energy loss

If too much energy is taken by the light wave, the resonant condition is no longer fulfilled since the electron energy is reduced, the electrons consequently slow down. When the optical wave power grows, an increasing number of electrons are trapped in the ponderomotive potential. When going down in phase space, the electrons loose kinetic energy to the advantage of the light wave. When they reach the bottom of the accessible space and cannot give any more energy, the laser saturates. Indeed, the electrons can even undergo several rotations in phase space before escaping the undulator because of slippage, while alternately providing to or taking energy from the optical wave. These oscillations are called 'synchrotron oscillations' [120–122]. They induce sidebands in the radiation spectrum. Strategies for sideband suppression have been examined [123]. While the laser intensity saturates, the gain is reduced. The electron energy decreases due to the kinetic energy loss enabling FEL amplification [124], the electron beam energy can also be reduced by the accumulated effects of spontaneous emission along the undulator, given by $\frac{\Delta\gamma_{\mathrm{SR}}}{\gamma_{\mathrm{o}}} = -\frac{1}{3} r_{\mathrm{e}} \gamma_{\mathrm{o}} K_{\mathrm{u}}^2 k_{\mathrm{u}}^2 L_{\mathrm{u}}$, with r_{e} the classical electron radius.

4.6.1.2 Increase of energy spread

The electron bunching and interaction via an energy exchange with the optical wave leads to an increase of the energy spread of the beam, reducing in consequence the gain via the contribution of F_{inh} that becomes more important. Intuitively, the gain bandwidth (related to the spontaneous emission bandwidth proportional to the inverse of the number of undulator periods) gets larger because of the inhomogeneous contribution and the gain distribution flattens. In consequence, the gain bandwidth and the limits set by the energy spread, provide a maximum undulator length.

4.6.1.3 Slippage

The slippage can stop the interaction: the electrons travel slightly slower than the photons, and once at the exit of the undulator, the time difference becomes typically $\frac{N_\mathrm{u}\lambda}{c}$. For the radiation to not escape from an electron bunch of duration σ_l, one can even consider that the radiation advance should remain in the peak of the distribution: $\frac{N_\mathrm{u}\lambda}{c} < \frac{\sigma_\mathrm{l}}{10}$ or $N_\mathrm{u} < \frac{\sigma_\mathrm{l} c}{10\lambda}$. This gives 300 periods for 1 nm radiation, 10 fs electron bunch, 300 periods, or for 1 μm, 10 ps electron bunch. Slippage thus sets another limit in terms of undulator length, the electron bunch duration should be larger than $N_\mathrm{u}\lambda_\mathrm{u}$.

4.6.2 Efficiency

4.6.2.1 Efficiency increase by undulator tapering

Electrons travelling on half the width of the gain curve of $1/2N_\mathrm{u}$ can deliver a relative energy of $\frac{\Delta\gamma}{\gamma} = \frac{1}{2}\frac{\Delta\lambda}{\lambda} = \frac{1}{4N_\mathrm{u}}$) of their kinetic energy, the efficiency r becomes

$$r = \frac{1}{4N_\mathrm{u}}. \tag{78}$$

The maximum efficiency is found by considering the total width of the gain curve, which would lead to $r = \frac{1}{2N_\mathrm{u}}$. It is however less realistic because the energy spread effect can limit the process. For example, for 50 undulator periods, the efficiency is of the order of 0.5%.

If too much energy is taken by the light wave, the resonant condition is no longer fulfilled since the electron energy is reduced, the electrons consequently slow down. A way to enhance the efficiency is to control the push further the saturation, i.e. electron trapping in the slow space charge wave. Indeed, "the nonlinearity of the oscillations of the trapped particle in the potential well of the wave leads to phase scrambling and finally the particle phase distribution becomes uniform (no bunching). The space-charge Coulomb forces in the electron bunches and the ripple magnetic field strength can also contribute to beam thermalization. The wave growth vanishes if the electron distribution is uniform at the phase velocity of the wave" [125], so it was proposed to increase the intensity of the magnetic field (an exponential profile was chosen) just before the space charge wave saturates, enabling an increase in the radiation rate [125].

One can delay saturation and let the intensity grow further by adjusting the undulator magnetic field so that the resonance condition remains fulfilled. Such a configuration of undulator is called a 'tapered undulator' either by changing the period [126], proposed by P. Sprangle *et al.* (fourth FEL Prize in 1991), or by changing the amplitude of the magnetic field proposed by Kroll [120]. One introduces a magnetic field dependent on the longitudinal position, as $B_\mathrm{uz}(s)$, as shown in Fig. 30. Technically, the change of magnetic field can be done by setting an angle between the girders of the magnetic arrays on which are located the undulator magnets, or by adopting a variable period [127]. The spontaneous emission properties of a tapered undulator have been calculated [128].

For a tapered undulator provides a varying magnetic field along the longitudinal direction $B_\mathrm{uz}(s)$, the resonance condition can be maintained according to

$$\lambda = \frac{\lambda_\mathrm{u}}{2\gamma_\mathrm{s}^2}\left[1 + \frac{1}{2}\left[\frac{eB_\mathrm{z}(s)\lambda_\mathrm{u}(s)}{2\pi m_\mathrm{o}c^2}\right]^2\right]. \tag{79}$$

The efficiency then depends on the number of electrons trapped in the potential well (bucket) and on the average energy loss of the resonant electrons.

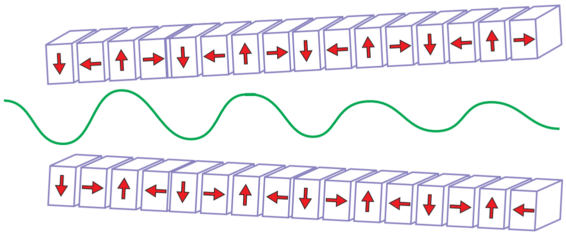


Fig. 30: Tapered undulator: the magnetic field amplitude depends on the longitudinal position. In the example here, an angle is set between the two girders supporting the magnet arrays.

4.6.2.2 Efficiency increase by energy recovery on the accelerator

There was also a strong interest to use recirculating accelerators for driving a FEL to recover the electron from one pass to another, and in particular of storage rings which exhibited good beam quality. The FEL theory was thus developed by the Italian team in particular, in considering both amplifier [129, 130] and oscillator [131] configurations. The electron beam distribution is modelled using the Fokker–Planck distribution. The energy spread enhancement and associated bunch lengthening due to the FEL interaction is kept on several turns. The competition with the anomalous bunch lengthening has also to be considered [132]. in the case of a storage ring FEL, the average power scales as

$$P \propto \left(\frac{\Delta\sigma_\gamma}{\gamma}\right)^2 P_{\text{sync}} \quad \text{with} \quad P_{\text{sync}} \propto IE^4. \tag{80}$$

It provides the limit which can be achieved on a storage ring FEL. It results from the radiative heating of the energy spread in electrons circulating in the storage ring. It is known as the 'Renieri's limit'. It has been independently found in Novosibirsk by N. A. Vinokurov [104].

4.7 FEL properties

The laser tuneability, one of the major advantages of FEL sources, is obtained by merely modifying the magnetic field of the undulator in a given spectral range set by the electron beam energy. The polarization depends on the undulator configuration.

The multimode theory was developed [133, 134] and the super-mode, defined "as the configuration of spatial modes, which reproduces itself after one passage throughout the interaction region" is introduced [135]. The evolution of the modes in the optical resonator was also used to evaluate the filling factor [88] and the multimode theory was examined considering the three-dimensional parabolic wave equation coupled to the Lorentz force equation [80, 136], enabling us to obtain different transverse mode patterns and dynamics [137]. FEL spatial and temporal behaviours were also examined as a coherent superposition of the exact Lienard–Wiechert fields produced by each electron in the beam [138]. The evolution of the free electron laser oscillator was further modelled using a Lagrangian formalism to follow the dynamics of the interaction between the electron beam and optical wave in a single pass [139].

After having a description of the gain and the saturation, theoreticians started to investigate coherence properties. R. Bonifacio [140] introduced the description in terms of electron field coherent quasi-classical states, where both the photon number and the electron momentum follow a Poisson distribution centred on the classical trajectories. G. Dattoli [141] examined the case of a given initial classical state, using a quantum description and looking at the coherent states of angular momentum and he found that "both the laser and wiggler fields are in the Glauber [142, 143] coherent state in the many mode case". The study was then continued in the classical conditions [144] and quantum analysis [94, 145] showing that in fact, strictly speaking, FEL does not exhibit Glauber coherence [145].

4.8 Low gain FEL configurations

So far we have discussed mainly the oscillator configuration where the synchrotron radiation from the undulator (spontaneous emission) is stored in an optical resonator. The electron light interaction leads to bunched electrons which emit in phase with the incident wave which gets amplified.

An external laser tuned on the undulator resonant wavelength λ can be sent in the undulator, synchronized with the electron arrival. The light wave and the electrons can interact in the undulator, leading to the external light amplification. This configuration is called the 'master amplifier'. Radiation is achieved on the same wavelength as the incident wave.

Using an external light wave tuned on the undulator resonance, the light wave interacts with the electron bunch in the undulator, inducing an energy modulation of the electrons; which is gradually transformed into density modulation at λ and leads to a coherent radiation emission at λ and λ/n, n being an integer (fundamental and harmonics) [81–83, 146–149].

Figure 31 presents the scheme of coherent harmonic generation: an external laser tuned on the fundamental of the undulator is used to modulate in energy, setting the emitters in phase for the radiation on the harmonics to emit coherently.

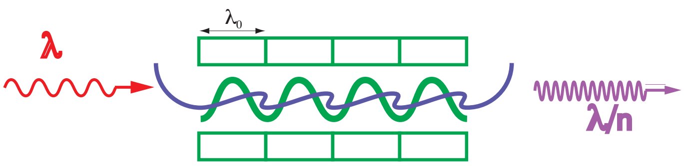


Fig. 31: Coherent harmonic generation scheme: an external laser tuned on the resonant wavelength of the undulator enables us to perform efficiently the energy exchange leading further to the density modulation and coherent emission on the harmonics.

4.9 Classical and quantum approaches: unification and quantum effects

Started with a first FEL theory discussed with quantum mechanics, FEL classical theory appeared to be very useful and applicable to the majority of the cases. While new theoretical approaches were investigated, unified models were searched. The formalism of the quasi-Bloch equations enables us to unify the quantum and classical approaches [144, 150]. Further, a unified theory of magnetic bremsstrahlung, electrostatic bremsstrahlung, Compton–Raman scattering, and Cerenkov–Smith–Purcell free electron lasers, are also proposed by A. Gover (18th FEL Prize in 2005) [151]. Quantum features and in particular coherence were also analysed using a non-linear quantum model [152, 153]. The limits of classical models were found when quantum effects start to influence the FEL process [124]. Quantum FELs have been actively studied in the R. Bonifacio's group [154–158].

4.9.1 Quantum recoil

When an electron emits a photon $\hbar\omega_{\mathrm{ph}}$, its energy is reduced by such an amount due to the quantum recoil. If the energy change due to the recoil is of the order or larger than the FEL gain bandwidth, i.e. given by the spontaneous emission width $\frac{\Delta\omega}{\omega} \approx \sqrt{(\frac{\Delta\omega}{\omega})^2_{\mathrm{h}} + (\frac{\Delta\omega}{\omega})^2_{\mathrm{inh}}}$, then the quantum recoil may significantly affect the FEL gain. Consider a typical gain bandwidth of 10^{-3}, for a short wavelength FEL, the fraction of the energy change $\frac{\hbar\omega_{\mathrm{ph}}}{E}$ is more than 10^{-6}, the quantum electron recoil is then negligible. It can then start to play a role with low energy electron beams and high energy emitted photons (for example in the X-ray range), such as in using an optical undulator (created by an optical wave).

4.9.2 Quantum diffusion

The emission of spontaneous emission radiation, if not affecting the electron energy by a significant amount, introduces an energy loss. In addition, the discrete nature of photon emission (over a wide energy spectrum) increases the uncorrelated energy spread, as for quantum excitation in a storage ring. The diffusion rate of the energy spread is given by [124]: $\frac{\mathrm{d}\langle(\Delta\gamma)\rangle^2}{\mathrm{d}s} = -\frac{7}{15} r_\mathrm{e} \bar{\lambda}_\mathrm{Compton} \gamma_\mathrm{o}^4 K_\mathrm{u}^2 k_\mathrm{u}^3 F(K_\mathrm{u})$ with $F(K_\mathrm{u}) = 1.2K_\mathrm{u} + \frac{1}{1+1.33K_\mathrm{u}+0.40K_\mathrm{u}^2}$ and $\bar{\lambda}_\mathrm{Compton} = \hbar/m_\mathrm{o}c \approx 3.86 \times 10^{-13}$ the reduced Compton wavelength [159].

5 The first FEL experimental results

5.1 The first FEL in Stanford (USA) in the infrared in 1977

5.1.1 The first FEL amplification in the infrared in 1976

After the theoretical prediction of the FEL concept in 1971, J. M. M. Madey searched how to set-up an experiment to test his idea of a FEL a [55]. Indeed, the High Energy Physics Laboratory on the Stanford campus concentrated a high knowledge on accelerator physics, both for normal conducting and superconducting devices. S-band accelerators had been developed by William W. Hansen (1909–1949) and Edward Ginzton (1915–1998) after the second World War, which led to the construction of the Stanford Linear Accelerator Laboratory (SLAC), a three kilometre S-band linac with upgraded klystrons under the direction of Wolfgang K. H. Panofsky (1919–2007). First superconducting linear accelerators were developed by William Fairbank (1917–1989) and Alan Schwettman (15th FEL Prize in 2002) [160], in order to exploit higher gradients and reduced power consumption of superconducting niobium cavities. High stability and sufficiently low energy spread beams for the low gain free electron laser exploration could be achieved on the superconducting linear accelerator.

J. M. J. Madey obtained financial support from the Air Force Office of Scientific Research (AFOSR) in 1972 in two steps: a first one to demonstrate gain, and a second one to achieve the FEL oscillation provided the first was a success. A team was gathered with J.M. J. Madey for laser physics, electronic instrumentation, cryogenic systems, superconducting undulator, Luis Elias (third FEL prier in 1990) for the optics, optical instrumentation, and conventional laser sources, superconducting undulator, and Todd Smith (third FEL Prize in 1990) for the accelerator. The composition of this new research team logically balanced the expertise between accelerators and electron tubes, and optics and lasers.

The first experimental demonstration of the FEL amplification was performed in Stanford in 1976 [161]. The electron beam has been produced by a 24 MeV electron beam generated by a superconducting undulator at 1.3 GHz. A 5.2 m long 3.2 cm period NbTi superconducting helical undulator was built (see Fig. 32) with a very high mechanical precision for the coil winding. It provided an undulator field of 0.24 T. A CO_2 laser at 10.6 μm with variable intensity and polarization was focused inside the undulator to a waist of 3.3 mm. The wavelength of the CO_2 laser being fixed, tuning was performed by changing the energy of the electrons around 24 MeV. The signal was detected with a high speed helium-cooled CuGe detector, synchronized with 1.3 GHz from the accelerator.

The CO_2 has been amplified, demonstrating a single pass gain of 7%, as shown in Fig. 33. Good agreement was found on the theoretical expectations regarding the gain. One notices that the gain is the derivative of the spontaneous emission, as understood by Madey [60, 87].

This FEL amplifier first experiment was a major step for the validation of the FEL concept with its gain medium using relativistic electrons in periodic magnetic fields.

5.1.2 The first FEL oscillator in the infrared in 1977

Because of the low value of the gain (7%), a high finesse optical resonator was necessary for attempting the oscillator experiment in order to insure cavity losses smaller than gain. An intermediate wavelength of (3.4 μm) was selected despite the small gain reduction, enabling the propagation of the desired funda-

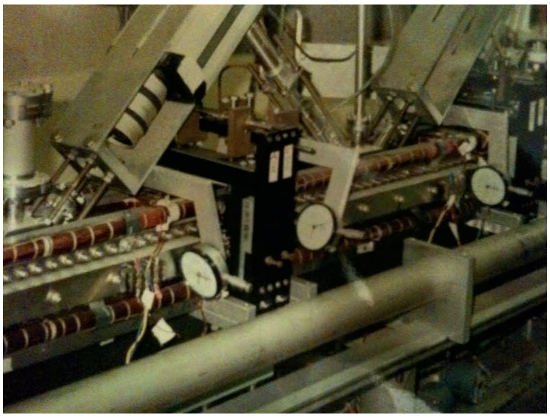


Fig. 32: Picture of the superconducting undulator used for the Madey's experiment

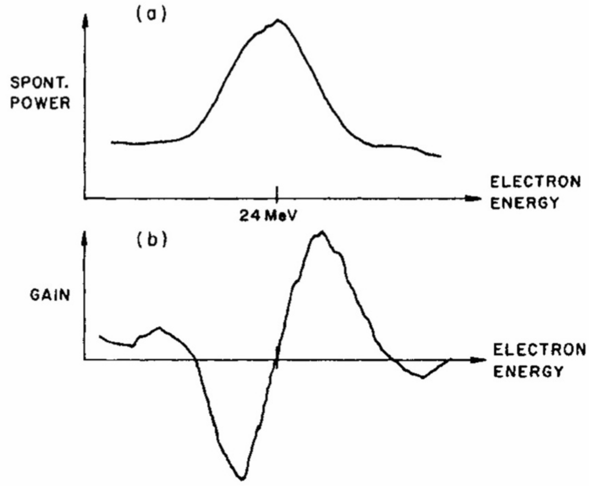


Fig. 33: (a) Spontaneous undulator emission and (b) measured gain which corresponds to the derivative of the spontaneous emission, from the first FEL amplification measurement by the Madey's group [161]. Undulator field of 0.24 T, peak current of 70 mA.

mental Gaussian resonator mode through the undulator vacuum chamber with minimal diffraction losses and to pre-align the optical resonator mirrors with an intracavity He–Ne plasma tube. Two new members joined the team: David Deacon as a PhD student, and G. Ramian for the accelerator injector since a higher peak current electron gun was required. Thanks to a gridded dispenser cathode (Eimac) driven by microwave triode amplifiers, 4 ps long electron macro-pulses with 2.6 A peak current at 11.8 MHz were achieved. Under such conditions, the expected gain reaching typically 100% appeared sufficient to overcome the cavity losses of 3% at 3.4 μm. The undulator having being damaged by an unanticipated surge in the voltage provided by its high current power supply, a second superconducting undulator had been also built but it happened that a failure of the insulation of the wire drastically limited the rate at which the magnet could be ramped up or down during operation.

The experiment was finally ready for operation in December 1976, and the FEL oscillation was rapidly observed, in January 1977 after optimizing the electron beam steering and focusing and optical cavity tuning. Figure 34 shows the FEL line, as compared to the spontaneous emission: the FEL line is sharper (relative bandwidth of 0.23% Full Width Half Maximum FWHM), and much more intense. The FEL provided a 360 mW average power, corresponding to an estimated 7 kW peak power and 500 kW intracavity peak power [162]. The output power reached nearly twice the power extracted from the electron beam in the amplifier experiment.

It can be pointed out that this first FEL result owes thanks to the quality of the electron beam delivered by the linear accelerator, together with its ability to provide rather long trains of electrons, enabling sufficient passes in the optical resonator to achieve the FEL saturation, (thanks to the high number of micro-pulses). Saturation was reached typically after 100 μs. Cryogenic operation being rather heavy, the next experiments took place in 1981 [163, 164], enabling further analysis of the FEL properties, to be compared with theoretical expectations.

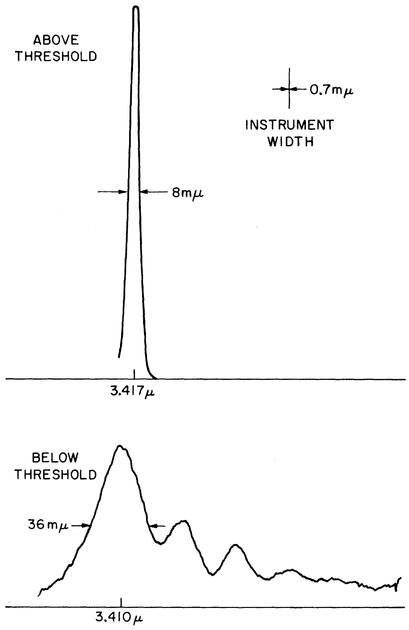


Fig. 34: First FEL: (a) FEL line at 3.4 μm. (b) Spontaneous undulator emission [162]. Electron beam from the MARK-III superconducting accelerator at Stanford, superconducting helical undulator.

This infrared FEL oscillator achieved in 1977 on the superconducting linear accelerator (Stanford, USA) established the first experimental demonstration of the FEL concept as a new laser type. It had thus evidenced that this new type of laser based on stimulated Compton backscattering could effectively work and opened bright perspectives in terms of average and peak power outputs. It indeed paved the way towards the further advent of X-ray tuneable FELs as unique sources of radiation for matter investigation. It was also the first laboratory experiment of stimulated Compton backscattering.

5.1.3 New directions after the first results

The first FEL paper terminates with the following "Because the gain falls at short wavelengths, a higher electron current will be required to support laser operation in the visible and the ultra-violet. Based on the small-signal gain equations sufficient current has been stored in existing electron storage rings to sustain laser operation at wavelengths as short as 1200 A" [162]. It thus gave directions of evolution. The first one dealt with the required improvement of electron beam parameters. The second concerned the attractiveness of the storage rings as an accelerator to drive the FEL, even in the isochronous operation [165]. Besides improvements on the electron characteristics, ways of increasing the undulator gain were investigated, such as the optical klystron [102] proposed by N. Vinokurov and Skrinsky, or the transverse gradient undulator enabling to handle a rather high level of energy spread [166]. "The transport system is designed to resolve the energy spread of the incident electrons into a transverse position and/or momentum spread at the entrance to the laser magnet. The magnet is designed to take advantage of the different trajectories followed by electrons of different energies, with the result that the optical wavelength at which gain is a maximum is far less sensitive to the electron energy than it would be in a conventional system" [166, 167].

5.1.3.1 Towards a better efficiency using storage rings or electrostatic accelerators

From the very beginning, great hope was put in recirculating accelerators, since "the RF accelerating field for the ring would have to supply only the energy actually transformed to radiation in the periodic field. The overall efficiency of such a system thus would not be limited to the fraction of the electrons' energy convertible to radiation in a single pass through the interaction region" [161]. In particular, free electron lasers on storage rings, as illustrated in Fig. 35, are considered in detail.

Specificities regarding the energy spread evolution due to the recirculation of the electrons in the storage ring after their heating by the FEL interaction have been investigated from a theoretical point of view. The energy spread can be enhanced via the FEL process but it can then be relaxed via the natural damping which takes place in the storage ring [129, 131].

One constraint comes nevertheless from the fact that the length of the straight section is limited.

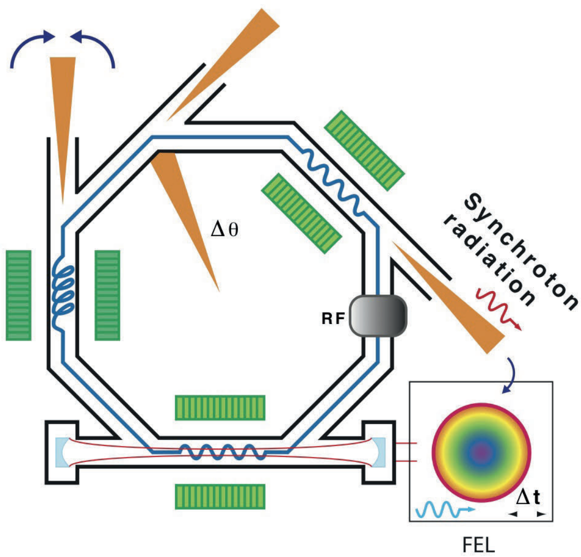


Fig. 35: Scheme of the storage ring free electron laser

J. M. J. Madey then searched for a storage ring to implement a storage ring FEL test experiment. Yves Petroff and Yves Farge, the director of LURE (Laboratoire d'Utilisation du Rayonnement Electromagnétique) in France were quite positive on the idea of using the ACO (Anneau de Collisions d' Orsay) storage ring for FEL investigations. ACO was initially build for high energy physics, and it turned to a parasitic use of synchrotron radiation in the beginning of the eighties. J. M. J. Madey came with some collaborators (D. Deacon, K. Robinson) while Y. Farge and Y. Petroff settled some team on the French side, with Michel Billardon (14th FEL Prize in 2001), Jean-Michel Ortéga (14th FEL Prize in 2001). M. E. Couprie (14th FEL Prize in 2001) and R. Prazeres joined the team later. An experiment was also set up on VEPP-3, in Novosibirsk [168], on Adone (Frascati, Italy) [169], and Brookhaven National Laboratory [170].

Besides the storage ring type of accelerator, L. Elias (Third FEL Prize in 1991) did consider the use of a DC electrostatic accelerator such as Van der Graff for the operation of high power efficient tuneable FEL [171] with low energy electron beams. Indeed, "Wall power to laser power efficiencies greater than 10% are possible" and should be compared to the 0.2% value in the case of the first Stanford experiment.

Analysis are carried out for a 9.38 MeV for 400 nm, and 3.55 MeV for 16 μm. Further analysis confirmed the possibility of highly efficient energy recovery [172] with DC electrostatic accelerators enabling to reach the required high average currents in the long pulse and CW operation modes for a FEL application [173].

5.1.3.2 Towards lower emittances and higher electron beam current

Besides the storage rings and electrostatic accelerators, realizable FELs on linear accelerators were considered [174], especially in terms of emittance, current density, electron bunch length, and stability. J. M. J. Madey nevertheless states that "As satisfying as it was to have completed two key proof of principal experiments, it was also clear that the development of useful devices based on this new gain mechanism would require both further theoretical and technical efforts. Although the experiments had established the capability of the new mechanism to operate at respectable signal levels, some significant questions remained as to the physical basis of these results. Higher electron currents and lower e-beam emittances would also clearly be required for operation of shorter wavelength and more compact systems" [55].

5.1.3.3 Following years of expectations

The great enthusiasm due to the success of the first FEL demonstration led to several experimental initiatives around the world, to extend the FEL achievements. It was followed by six years of expectations, as reported by C. A. Brau [1] as follows "Unfortunately, none of the electron-beam sources available at that time had enough electron-beam current and satisfactory electron-beam quality to make lasing easy. Although gain was measured in several experiments, it was not until six years later, in 1983, that the second free-electron laser was operated in the optical part of the spectrum. In that year, three devices began to lase".

The first was at Laboratoire pour l' Utilisation du Rayonnement Electromagnétique (LURE), in Orsay, France, where the electron beam in the storage ring ACO was used to achieve lasing in the visible [175, 176].

The second was at Stanford, a team from TRW (Thompson Ramo Wooldridge (Northrop Grumman since 2002)) used the superconducting accelerator previously used by Madey to achieve lasing in the near infrared [177].

The third was Los Alamos, where a newly constructed electron accelerator was used to achieve lasing in the mid-infrared [178].

During the same period, development of ubitron-type devices began at several laboratories. Because the threshold electron beam current at which the space charge wave can be excited increases as the third power of the electron energy, these devices were limited to low electron energy (no more than a few megaelectronvolts), and long wavelength. Nevertheless, Marshall and his co-workers at Columbia and Naval Research Laboratory achieved lasing at 400 μm with an electron beam having an energy of 1.2 MeV and a peak current of 25 kA [179]. "These devices are limited to wavelengths in the submillimetre region and beyond, where the optical radiation is transmitted through a waveguide', and "the physics in this regime involves collective oscillations (space charge waves) in the electron beam", the development of ubitron-type devices is not developed in this FEL history [69].

At the end of 1982 (September 26–October 1) the 'Bendor Free Electron Laser" conference was held in France, whose subject matter was limited to the FEL in the Compton regime. It gathered 62 participants. The atmosphere was particular, since a lot of efforts devoted towards the operation of new free electron lasers started to provide preliminary results. In the foreword of the proceeding by D. A. D. Deacon and M. Billardon, it is said: "The most striking aspect of the collection of papers contributed to this volume is the amount of experimental progresses which have been made in the 12 months since the last summary of progress of the field. Four new undulators have been brought into operation, and measurements have been made on the spontaneous emission spectra, gain, electron trapping in the linac and bunch lengthening in the storage ring, sub-threshold effects and mirror degradation, time dependent short pulse phenomena; and laser-induced harmonic generation. This sudden (and exhausting) blooming of experimental results has in fact been in the making for two or three years. The perseverance of the authors of these works during this long preparation time deserves recognition and applause.

In the short wavelength $\lambda < 1\,\mu$m range, where the storage ring is the universally favoured device because of its current density and duty factor, four projects are underway at Brookhavven, Frascati, Orsay, and Novosibirsk. The Frascati and Orsay groups have contributed to the proceedings two valuable papers which summarized a wide variety of measurements and FEL diagnostics. These two groups (along with the Novosibirsk physicists who were unable to attend the conference) have been able to probe and verify the theory of the FEL to a level of precision and complexity which is unthinkable in the linear accelerator machines.

In the long wavelength $\lambda > 1\,\mu$m range, a lower current density is required to drive the interaction, and a wide variety of electron beam sources are now being put into use at 11 experimental centres. There are two induction linac sources, two microtrons, six RF linacs, one storage ring, and one van de Graaf sources being set up or used for FEL work at the following respective centre, Lawrence Livermore Laboratory, Naval Research Labs., Bell Labs., Frascati (ENEA), Los Alamos national Laboratory, Math. Sciences Northwest, NRL, Stanford (H. E. L. P.), TRW, the UK Collaboration, Berlin, and Santa-Barbara. At present, all the experimental work in this wavelength region has been done with the RF linacs. As is the case for short wavelength work, in the previous 12 months an unprecedented flow of new research results have been produced in the infrared devices. During the conference, these results were described by the Stanford group, who had succeeded in measuring the time structure of their picosecond laser pulses, and by the Los Alamos, the MSNW (Mathematical Sciences NorthWest), and the TRW groups, who have been able to measure the electron trapping in tapered wigglers."

Next we discuss the three new FEL operations achieved in 1983.

5.2 The second FEL in Orsay (France) in 1983

5.2.1 The second FEL oscillator in Orsay (France) in the visible in 1983

At that time, storage rings appeared as suitable accelerators because of the electron beam performance. A picture of the ACO (Anneau de Collisions d' Orsay) storage ring (LURE, Orsay, France) is shown in Fig. 36.

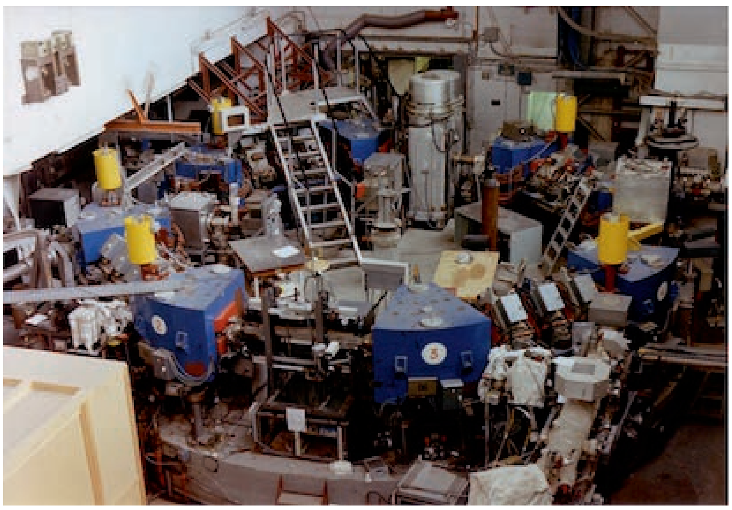


Fig. 36: ACO storage ring used for the second world wide FEL in Orsay (France), dipoles in blue

First measurements started with a superconducting undulator (23 $\times 40$ mm periods, maximum field of 0.45 T, $K = 1.68$ T) with an inverse T-shape vacuum chamber [180, 181], enabling to observe visible radiation between 140 and 240 MeV. The gain has been measured [182] and found to be very small, it thus required mirrors of extremely low losses. It appeared that significant imperfections in the magnetic field led to a broadening of the line and a gain reduction of 50%.

It was followed by the construction of a $SmCo_5$ permanent magnet-based undulator [183, 184] (17 mm $\times 78$ mm), using the configuration proposed by Halbach [32] with magnets rotated by $\pi/2$ from one position to another. The radiation produced by such an undulator was observed in the Vacuum Ultra Violet (VUV) for the first time using the ACO electron beam at 536 MeV. However, the straight section length being limited to 1.3 m, the gain was limited to a few 10^{-4} [185] and made the laser oscillation impossible despite the efforts concerning high reflectivity mirrors [186]. The gain could be enhanced by a factor of 2 up to 7 by turning the undulator to the optical klystron configuration by replacing the three central periods by a three pole wiggler [187]. The radiation has been measured and analysed, as shown in Fig. 37 [184].

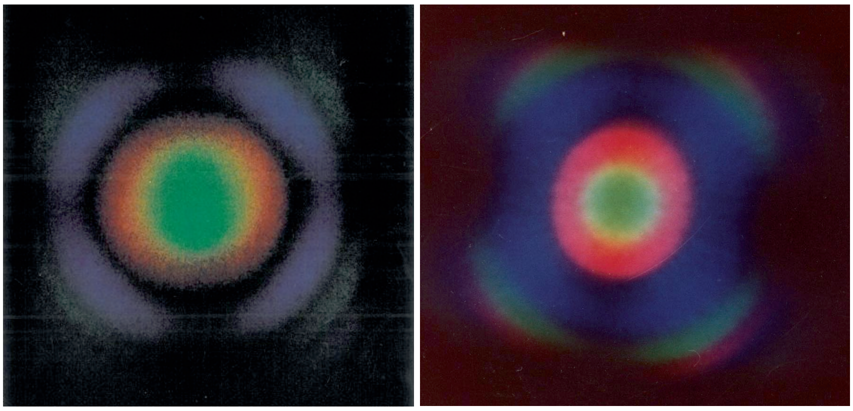


Fig. 37: ACO optical klystron spontaneous emission for different undulator gaps

Because of the mirror reflectivity degradation induced by the harmonic content of the undulator, the electron beam energy has been set between 160 and 166 MeV to minimize the undulator harmonic content. The optical cavity has a length of 5.5 m with round trip cavity losses of 7×10^{-4}. ACO (Orsay, France) [175] provided the second worldwide FEL (first visible radiation) in 1983. Figure 38 shows the laser tuneabllity achieved on ACO by changing the optical klystron gap. Getting the level of the cavity losses smaller than the gain was at time very challenging, and issues with mirror degradation induced by synchrotron radiation and mirror measurements were investigated [186].

Lasing was occurring on different lines of the optical klystron spectrum, as shown in Fig. 39. When the electron bunches circulating in the ring are not synchronized with the optical pulses bouncing between the mirrors, i.e. in the optical cavity detuned configuration, the energy exchange could not take place, and the measured spectrum corresponds to that of the spontaneous emission. When the cavity is properly tuned and the gain is larger than the cavity losses, then the optical klystron lines are growing and lead to the laser effect. Because of the fringe structure of the optical klystron, three lines are simultaneously lasing with the most intense one at 647.6 μm, each wavelength being located at a maximum of the gain versus wavelength curve, fulfilling properly the Madey theorem. This ACO FEL can be considered as the first multi-colour FEL.

Various studies were carried out after the first laser oscillation on ACO [185, 188, 189]. The FEL dynamics involves an interplay between the electron energy heating induced by the FEL interaction and the synchrotron damping. The FEL was exhibiting a naturally pulsed macro-temporal structure for perfect synchronism (synchronization between the electron circulating in the storage ring and the optical pulses bouncing around the mirrors of the optical resonator) [190], as shown in Fig. 40. Due

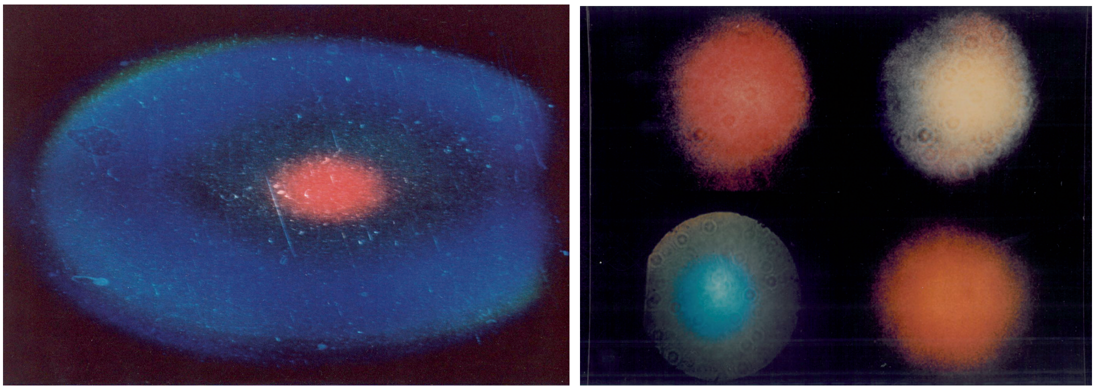


Fig. 38: First visible FEL on ACO storage ring in France in 1983. Left : Laser oscillation with red central wavelength, right : free electron laser tuneability in the visible

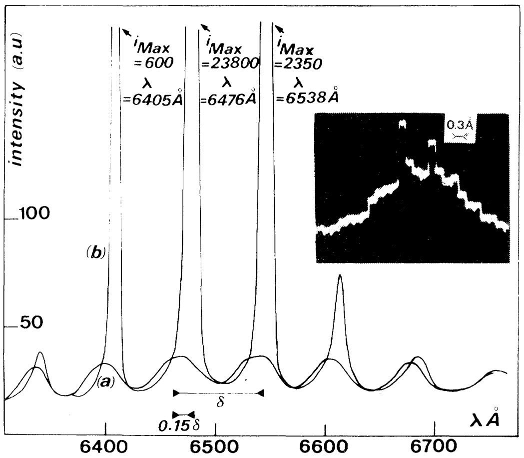


Fig. 39: First visible FEL on ACO storage ring in France in 1983 Spectra of the cavity output radiation under two conditions: a) cavity detuned without amplification and b) cavity tuned (laser on), from [175]. Insert: zoom on one laser line.

to the electron beam recirculation, the electron bunch heating induced by the FEL interaction leads to an electron bunch lengthening [191] and even to a modification of the shape of the electron bunch longitudinal distribution [192]. The FEL was operated in the Q-switching mode, in cancelling the optical gain by a small variation of the RF frequency, trigged by an external pulsed low frequency generator or by applying a modulation of the transverse position of the electron beam with the electric field of a pick up electrode. During a few milliseconds, the optical resonator being tuned, the FEL pulse can develop. Then, the pulse naturally decays, and afterwards, the cavity is detuned enabling the electron beam to be cooled down, and the FEL pulse to restart with the maximum power starting from non-heated electron beam.

5.2.2 Coherent harmonic generation in the VUV on the ACO storage ring Orsay (France) in the UV and VUV

Coherent harmonic generation [193] was achieved in the UV and VUV on the ACO storage ring. a Nd–Yag laser (1.06 μm wavelength, 20 Hz repetition rate, 15 MW peak power, 12 ns pulse duration) was tuned on the optical klystron first harmonic. The coherent third and the fifth harmonic of Nd–Yag laser were observed, with a spectral ration of 6000 for the third one, and 100 for the fifth one. It was then followed by further measurements in the VUV [194, 195].

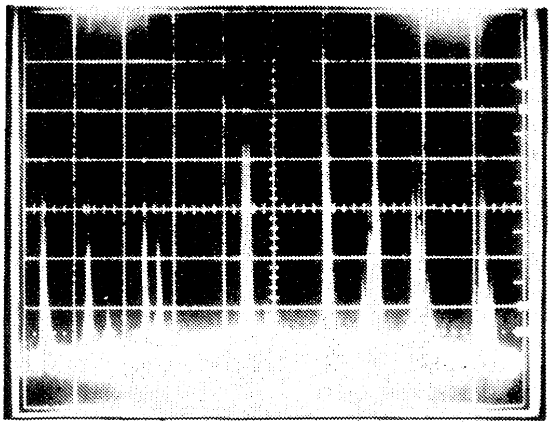


Fig. 40: Temporal structure at perfect synchronism on the ACO storage ring FEL, from [175]

5.3 The next two FEL oscillators in 1983

5.3.1 The new developments on the Stanford HEPL FEL

Led by a team of TRW, Space and Technology Group, the Stanford FEL has been operated with a tapered undulator in order to enhance the efficiency [177]. The FEL has been operated above threshold at the wavelength of 1.57 pm. The employed undulator is a permanent magnet one in the Halbach configuration [196] using $SmCo_5$ magnets with a period of 36 mm reaching up to 0.29 T. The peak field is varied longitudinally by controlling the spacing between the magnet planes. Inspired from the optical klystron, the undulator consisted of different segments: a prebuncher with a constant undulator field providing energy modulator of the electron beam, a dispersive section, a tapered section to improve trapping and higher extraction efficiency, and a radiator with constant undulator field. The laser was operated with 0%, 1%, and 2% tapers in energy, achieving laser efficiency of 0.4%, 1.1%, and 1.2%, respectively. The efficiency of the 2% taper was three times the untapered case. T. Smith converted the FEL to a user facility.

5.3.2 The Los Alamos FEL

Los Alamos National Laboratory had set up a FEL programme, aiming at demonstrating high power at 10.6 μm.

A first objective was to demonstrate the advantages offered by permanent magnet-based undulators, by developing a permanent magnet undulator in the Halbach configuration [196]. An undulator of 1 m long, 27.3 mm period, 8.8 mm gap, 0.31 T peak field, thanks to a funding from the Department of Energy, was built by R. Warren (sixth FEL Prize in 1993). It was decided to set the magnets directly in a vacuum in order to reach a higher magnetic field. Three fluorescent targets were installed on the electron path, for a proper alignment of the vacuum chamber and overlap of the electron / photon beams [197].

A linear accelerator of 20 MeV energy at 1.3 GHz was built on purpose for the experiment. It provided a peak current of 20 A, with an emittance of 2 mm mrad, an energy spread of $\pm 1\%$, and a total length of 10 m. It is the first FEL experiment with a dedicated accelerator, whereas the competing experiments are sharing the use of the electron beam.

The FEL operation by itself required the use of highly reflecting dielectric mirrors placed directly in vacuum and with a good resistance to high laser power. Besides, the electron beam macro-pulse should be sufficiently long to enable the growth of the FEL power. The optical resonator had a length of 9.92 m, matching the electron micro-pulse separation of 46.15 ns and cavity losses of 3%. Mirrors were compatible with an alignment with a He–Ne laser.

The FEL gain using a CO_2 laser was measured. FEL oscillation was obtained in 1983 in the 9–11 μm spectral range, with an intra-cavity peak power of 20 MW and an average output power of 1 kW in 70 μs macro-pulses [178, 198]. Nine orders of magnitude of power growth were observed with a net growth of 17%. A Germanium detector was used for the analysis of the optical signal. Harmonic lasing was observed on the second and third harmonics and characterized by different decays in the optical cavity. The FEL was then studied in details [199, 200]. The dependence of gain and saturation on cavity length, alignment, beam parameters, and other critical variables were compared with theory.

The Los Alamos team was eager to improve the FEL efficiency and first demonstrated an extraction efficiency larger than 3% [201] in the amplifier configuration with a tapered undulator. An efficiency larger than 4% [202] in an oscillator configuration was then obtained, as shown in Fig. 41.

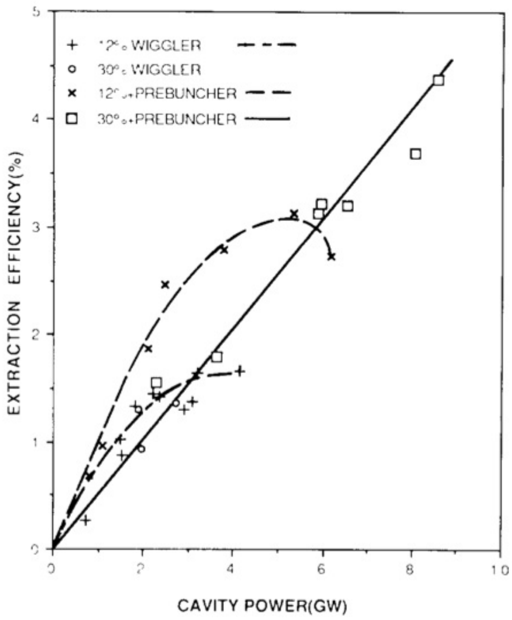


Fig. 41: Efficiency in the case of a tapered undulator in the Los Alamos oscillator experiment from [200]

5.3.3 The Santa Barbara FEL

Besides the interest of storage rings for electron beam recirculation, the use of energy recovery on an electrostatic accelerator was studied [171]. An experiment was thus set at the University of California, Santa Barbara, using an electrostatic accelerator [173], as shown in Fig. 42.

The electron beam is produced in an electron gun, then goes through a pelletron charging chain so that it is charged to a negative high voltage with respect to the ground, and it is accelerated in the accelerator column. It is transported and matched to the FEL line with the undulator and the optical cavity thanks to a set of achromatic bends and quadrupoles, and then sent back to the electrostatic tube entrance where it is captured by the low voltage multistage collector and restored the lost energy by FEL interaction. Energy recovery is of interest for FEL since only typically one percent of electron beam energy is converted to the FEL. It can also lead to a more stable electron beam. The FEL was operated with a permanent magnet Halbach type undulator (36 mm period, 38 mm gap, 0.46 T peak field) with a waveguide resonator [203] using electrons of 2.98 MeV providing a peak current of 1.25 A, with macro-pulses of 50 μm long and a 90% recovery efficiency. The optical beam was produced in the far infrared at 750 GHz with 30% gain per cycle for 11% losses per cycle, with 10 kW estimated peak power. The saturation was reached in 4 μs [204]. Single-mode operation in a free electron laser was then observed [205].

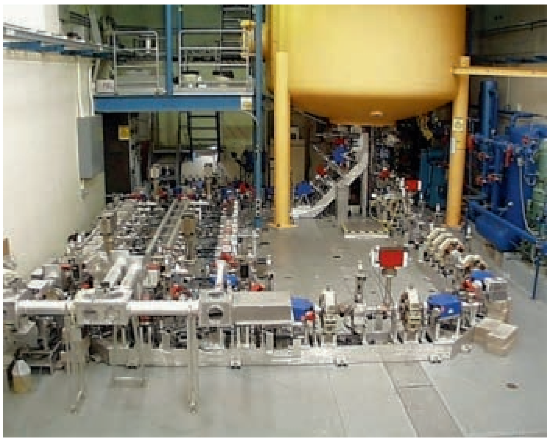


Fig. 42: Picture of the Santa Barbara (USA) FEL

5.4 Following developments on low gain FELs

Following these new FEL operations, the field was in rapid expansion, and FEL experiments were installed on various types of accelerators. There was a quest for wavelengths of operation, which depended on the electron beam, the mirror performance for the oscillator configuration and the ways to increase the gain. In addition, coherent harmonic generation using an external laser was continued. However, not all the projects have been successful as states C. Brau in his book in 1990: "It should be pointed out that of the free electron lasers which have been constructed thus far, considering only those built to operate in the optical regime, only a few have worked. The reason for failure, in most cases, has been that the available electron accelerator was not satisfactory for free-electron laser experiments and could not, within time and budget restrictions, be suitably modified. One or two of the lasers which have not yet worked may yet be brought into operation, but free-electron lasers remain subtle, expensive devices" [1]. Ten years after the first Stanford demonstration, there were still less than 10 Compton FEL under operations [2] but there was already a great interest for user application. Are given below examples of FEL launched in the continuity of the first FELs and which contributed significantly to the FEL field growth. These examples are not intended to be exhaustive. W. B. Colson was, during FEL conference, collecting the information on the different FELs under operation and progress [206, 206–208]. In the report of the FEL conference held in 1994 a national research Council FEL Committee report [209] is mentioned, chaired by D. Levy, recommending building infrared FELs as user facilities, developing the technology for UV FELs, and research and development for X-ray FEL.

5.4.1 FEL oscillators

In the early years of FELs, the type of accelerator, especially because of electron beam performance, defined somehow the reachable spectral range, as illustrated in Fig. 43. The main progress on the FELs built after the results of 1983 are described below, classified per accelerator type.

5.4.1.1 Storage ring based FELs

The activity continued on storage rings [210, 211] especially for the quest towards short wavelength of operation, since high quality electron beams were still produced. In addition, the interplay between the beam dynamics in the ring and the FEL interaction was of great interest.

VEPP-3 FEL IN RUSSIA

The second storage ring FEL following the one on ACO was achieved on VEPP3 (Novosibirsk, USSR) [212–215] in 1988. The team led by N. Vinokurov (fourth FEL Prize in 1991) included in particular

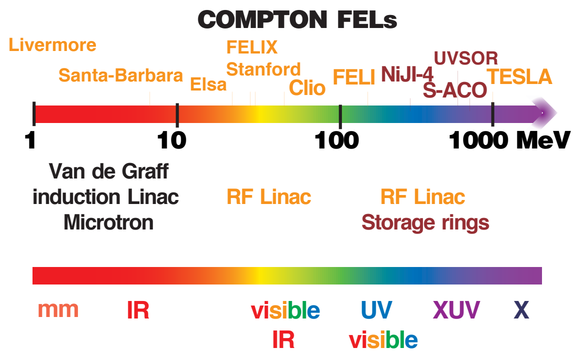


Fig. 43: Accelerator type used for FEL and corresponding spectral range. Examples of FELs

V. Litvinenko (17th FEL Prize, in 2004), I. Pinaiev, V. Popik, N. G. Garvilov, A. S. Sokolov, as younger scientists at that time, and senior ones with A. N. Skrinski and G. N. Kulipanov. Mutual coherence of spontaneous radiation from two undulators separated by an achromatic bend was observed [216, 217]. The experiment was renewed, with an electromagnetic optical klystron implemented [218] on a bypass [219], as shown in Fig. 44.

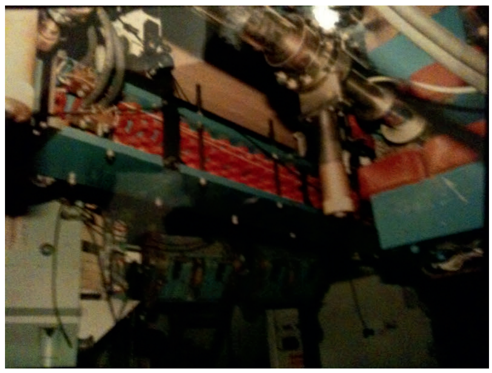


Fig. 44: Picture of the VEPP3 FEL optical klystron (the FEL was installed on the ceiling)

Because of the very low gain, despite the enhancement close to the optical klystron, the cavity losses were very critical. A method of mirror measurement with a reflectivity close to unit was proposed [220]. The confocal configuration for the optical cavity was studied [221]. The laser covered from the visible to the UV down to 240 nm with three sets of mirrors. Linewidth narrowing was achieved with a Fabry–Perot etalon [222–224], reaching a relative band width of 10^{-5}, as shown in Fig. 44.

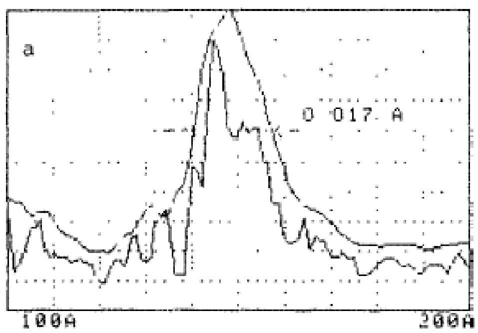


Fig. 45: Linewidth narrowing on the VEPP3 FEL with an etalon installed inside the optical resonator, from [223]

VEPP3 FEL kept for a while the record of the shortest FEL wavelength at 200 nm. Measurements of FEL spectra and temporal structures were compared to the theory [225].

SUPER-ACO FEL IN FRANCE

Then, the third storage ring FEL was obtained on the Super-ACO (Orsay, France) [226–228] following the successful results achieved on ACO. Super-ACO was a storage ring built on purpose for synchrotron radiation use. Super-ACO FEL was operated at 600 MeV, and then 800 MeV, which was the highest electron beam energy for a storage ring FEL. The Super-ACO optical klystron spectrum and associated gain are shown in Fig. 46. The gain is the derivative of the spontaneous emission, as given by the Madey's theorem.

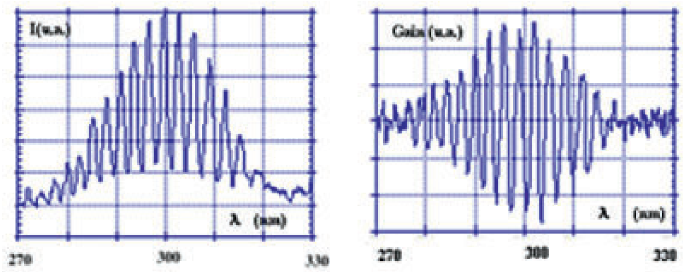


Fig. 46: Super-ACO optical klystron spontaneous emission for different undulator gaps

The undulator synchrotron radiation led to even more drastic conditions of degradation of the multilayer mirrors in the optical cavity [229] requiring specific mirror characterizations [230]. The FEL was then obtained in the UV [231, 232]. The FEL was fully characterized. Transverse modes can be controlled via the optical resonator [233] as shown in Fig. 47.

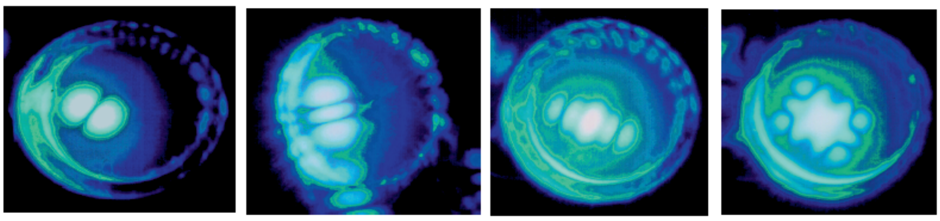


Fig. 47: Super-ACO transverse mode resulting from a misalignment of the optical cavity axis with respect to the magnetic axis of the undulator.

The temporal profile was studied [234, 235]. The zero detuning regime with CW operation of the FEL was stabilized using longitudinal feedback [236]. Extensive studies on longitudinal dynamics were carried out [237]: the coupled dynamics of the electrons in the storage ring [238], mutual influence of the coherent synchrotron oscillations [239], local energy exchange between the FEL and the electron beam [240], FEL-induced suppression of the sawtooth instability [241], control of the pulsed zones versus detuning [242], and advection-induced spectro-temporal defects [243]. The super-ACO ring was shut down in 2003.

TERAS (TSUKUBA ELECTRON RING FOR ACCELERATING AND STORAGE) AND NIJI-IV FEL IN JAPAN

Storage ring FEL oscillation was then obtained on TERAS (Tsukuba, Japan) [244], NIJI-IV (Tsukuba, Japan) (Niji is the Japanses word for "rainbow") [245].

UVSOR (ULTRAVIOLET SYNCHROTRON ORBITAL RADIATION) FEL IN JAPAN

A storage ring FEL was developed on UVSOR (Okazaki, Japan) [246]. The first developments of the UVSOR FEL were led by H. Hama (17th FEL Prize in 2004). One of its specificities is the use of an helical optical klystron [247]. Temporal dynamics was studied [248, 249], longitudinal feedback was developed to stabilize the temporal position of the FEL micropulse [250].

DUKE FEL IN THE USA

The DUKE (Duke, North Carolina, USA) FEL, shown in Fig. 48, was first operated in the visible [251], and then in the UV–deep UV [252] below 200 nm [253] with a distributed optical klystron implanted on a long straight section, enabling a reasonable gain. The DUKE FEL conducted also various dynamical studies [254], such as the observation of giant pulses [255], self-induced harmonic generation [256], time structure [257], and output power limitations [258]. Micropulses are Fourier limited [259]. The DUKE FEL was first developed by V. Litvnenko (17th FEL Prize in 2004).

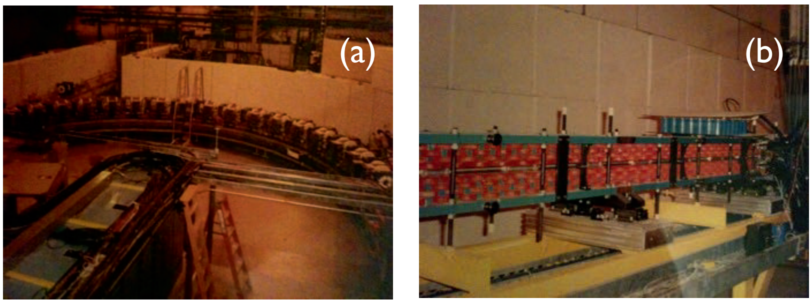


Fig. 48: DUKE FEL pictures: (a) the ring, (b) the multiple optical klystron

The DUKE storage ring is a dedicated accelerator for FEL operation and it is still under operation.

DELTA (DORTMUND ELECTRON ACCELERATOR) FEL IN GERMANY

The FEL was also achieved on the DELTA (Dortmund, Germany) storage ring in the visible and UV [260].

ELETTRA FEL IN ITALY

The FEL on the ELETTRA (Trieste, Italy), using a helical optical klystron, could provide sufficient gain as well [261]. After its first operation [262], it enabled operation at shorter wavelengths [263, 264] in the VUV at 190 nm, setting the record of the shortest wavelength achieved on a FEL oscillator. For short wavelengths, mirror degradation due to undulator synchrotron radiation appeared to be critical [265–268].

Even though the ELETTRA FEL was quite successful, it had to be operated in dedicated shifts [269]. A storage ring FEL is usually operated at a lower beam energy than the one employed for conventional synchrotron radiation users, setting a limit in the development of the storage rings FELs. This was also the main reason for the withdrawal of the SOLEIL FEL [270].

5.4.1.2 Linac based FELs

MARK III FEL STANFORD AND THEN DUKE UNIVERSITY, USA

The MARK III FEL (started in 1986 at Stanford University and then moved to Duke University, USA) used a normal conducting linear accelerator (26–45 MeV, 10π mm mrad emittance, 0.7% energy spread, 20–40 A peak current) operated in the infrared (1.4–8 μm) and enabled different FEL studies such as harmonic lasing [271], coherent harmonic emission [272], pulse compression using energy chirp [273, 274], and master oscillator power amplifier (MOPA) configuration [275]. User applications were then developed.

Dedicated linac-based FELs were then built for user applications.

VANDERBILT UNIVERSITY FEL, USA

The Vanderbilt University FEL (Nashville, USA), led by C. Brau (ninth FEL Prize in 1996) used a MARK III-type linear accelerator and a 1.08 m length, 23 mm period, 0.4 T field undulator [276] and was established to serve the medicine and material science fields [276, 277]. It operated in the infrared (2–10 μm).

FELIX, THE NETHERLAND

FELIX (The Netherland), led by A. Van der Meer (12th FEL Prize, 2002), uses a normal conducting linear accelerator (15–45 MeV, 50π mm mrad emittance, 0.25% energy spread) 38 $\times$ 65 mm period, 0.22 T field undulator first operated in the 6–100 μm spectral range [278]. Various studies have been carried out, such as phase locking [279], limit-cycle operation [280], and single mode selection [281]. FELIX has been operating for 25 years.

CLIO (CENTRE LASER INFRAROUGE D'ORSAY), FRANCE.

CLIO (Orsay, France), led by Jean-Michel Ortéga (11th FEL Prize, 2001), uses a normal conducting linear accelerator (30–70 MeV, 50π mm mrad emittance, 0.2% energy spread, 100 A peak current) 1.08 m length, 48 mm $\times$ 23 mm period, 0.4 T field undulator [282] first operated in the 2–17 μm [283] and then 3–120 μm [284] spectral range for users. Various operating modes were investigated, such as two colour operation [285], efficiency improvement [286], sub-picosecond pulse regimes [287]. CLIO has been in operation for 25 years.

ELSA (ETUDE D'UN LASER ACCORDABLE, STUDY OF A TUNEABLE LASER), FRANCE

Another infrared FEL was built and operated in France, on the linear accelerator ELSA, enabling lasing [288] and studies on high efficiency [289].

FELBE, GERMANY

FELBE (FEL at the Electron Linear Accelerator with High Brilliance and Low Emittance) (Forschungszentrum Dresden-Rossendorf, Germany) [290], uses a superconducting linear accelerator consisting of two 20 MV superconducting units operating in CW mode with a pulse repetition rate of 13 MHz, with 1 mA average current. It serves two free electron lasers (U27-FEL (until end of 2016) and U100-FEL), produces coherent electromagnetic radiation in the mid and far infrared (4–250 μm). Pulse energies are in the few 100 nJ range with pulse durations of a few picoseconds. The typical operation mode offers a 13 MHz micropulse repetition rate in macropulses of a few 100 μs at up to 25 Hz or, alternatively, FEL operation in a continuous 13 MHz mode.

FELI (FREE ELECTRON LASER RESEARCH INSTITUTE), JAPAN

FELI (Japan) was built in Japan and serves different user beamlines [291–293].

5.4.1.3 Induction linac based FELs

High power FELs were developed both at Livermore (USA) [294] and at CESTA (France) [295]. These results will be described further in the high gain section.

5.4.1.4 Microtron based FELs

A Cherenkov FEL was built in ENEA (Frascati, Italy). The electron beam at 5.3 MeV with a 200 mA current, with emittances of 6π mm mrad in the vertical, 18π mm mrad in the horizontal, and a 0.5% energy spread is focused to travel close and parallel to the dielectric (polyethylene). A quasi-optical resonator using a mirror and an output coupler provides feedback to the radiation. The FEL was emitted at 1660 μ.mrad [296, 297].

A FEL in the mm wavelength range was achieved on the Pahra microtron in the Lebedev Physical Institute (Moscow, Russia) with an electron beam of at 7 MeV, 50 mA current, 40π mm mrad (6π mm mrad) radial (vertical) emittances, a 0.1% energy spread, an undulator (6 $\times$ 168 mm periods, 0.26 T) and a waveguide, mylar mirrors [298, 299].

KAERI (Korea) has developed a compact far infrared (FIR) FEL driven by a 7 MeV microtron [300].

5.4.1.5 Energy recovery accelerator based FELs

Besides electrostatic accelerators, energy recovery can also be performed using linear accelerators. In energy recovery linacs (ERL), the electron beam is recirculated in a loop so that is enters again the accelerating sections, but dephased by π, so that the beam energy is given back to the accelerating sections [301]. It thus provides a high electron beam efficiency and reduces the radiation hazard by setting the beam dump at low energy. ERL-based FELs are suitable for high average power output [302].

JEFFERSON LABORATORY FEL, USA

An ERL-based FEL was first operated in the infrared at Jefferson Laboratory (Virginia, USA) by the team of G. Neil (13th FEL Prize in 2000) and S. Benson (13th FEL Prize in 2000). The superconducting energy recovery linac provides an electron beam at 18.7 MHz of 48 MeV with a 5 mA current, 80 pC charge, 60 A peak current, 7.5π mm mrad emittance, and transforms 75% of the beam power back to RF power. The FEL has been operated at 3.1 μm [303] with an undulator of 40 $\times$27 mm period, 1.4 deflection parameter and an 8 m long optical cavity with infrared mirrors of reflectivity of 99.85%. Laser damage could be an issue [304]. The average power reached 1 kW [305]. The FEL was operated in the tapered configuration [306] In July 2004, 10 kW of CW operation was achieved at a wavelength of 6 μm [307], and then extended in 2006 to 14.2 kW at 1.6 μm in a CW mode of operation [308]. After a machine modification, the spectral range has been extended to the UV in 2010 down to 363 nm with 100 W average power level [308].

JAERI FEL, JAPAN

E.J. Minehara (13th FEL Prize in 2000) led the team of the JAERI FEL installed on the superconducting energy recovery linac in JAEA in Japan [309, 310]. Superradiance [311] sustained saturation [312] was studied. A 1.7 kW operation was achieved [313].

NOVOSIBIRSK ENERGY RECOVERY FEL, RUSSIA

The only FEL room temperature energy recovery linac is located in the Budker Institute (Novosibirsk, Russia) [314, 315].

KAERI ENERGY RECOVERY FELS, KOREA

KAERI has developed three types of FELs since 1992: a millimetre wave driven by a 0.4 MeV electrostatic accelerator [316], a compact FIR FEL driven by a 7 MeV microtron, and an infrared FEL with average power of 1 kW driven by a 40 MeV superconducting accelerator [317]. The research is led by Y. U. Jeong (24th FEL Prize in 2013).

5.4.1.6 FEL oscillator performance and limits

Figure 49 reports on the spectral range covered by several FEL oscillators in the visible and UV–VUV. Reaching shorter and shorter wavelengths was getting difficult since the gain, that had to overcome the mirror losses, was typically decreasing for equivalent beam parameters. Some new developments were carried out for the UV dielectric multilayer mirrors [266–268] whereas the conditions of mirror degradation became even worse [265]. The shortest wavelength was obtained on the ELETTRA FEL at 190 nm. The figure shows as well the output power, which is of course, larger for larger electron beam energies, but which makes the obtention of a sufficient gain more difficult.

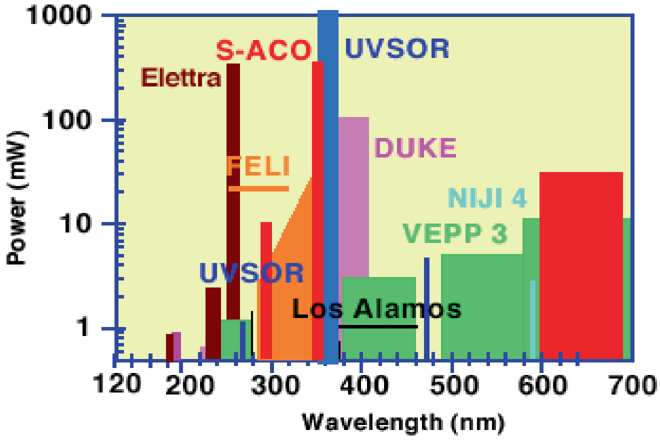


Fig. 49: Short wavelength FEL oscillators

Because of the oscillator configuration, the multi-pass in the optical resonator enables us to increase the coherence of the FEL. Figure 50 displays the pulse duration versus wavelength for different FEL oscillators. They are all operating close to the Fourier limit.

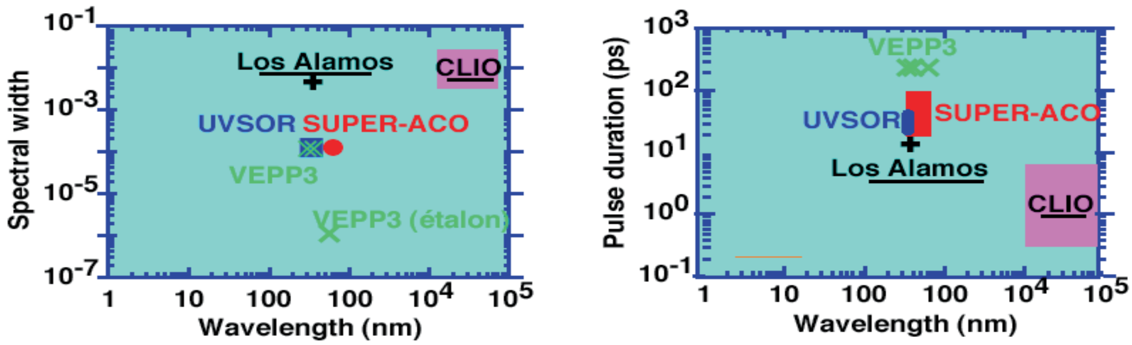


Fig. 50: Pulse duration versus wavelength for several FEL oscillators

Presently, thanks to the performance of Bragg reflectors such as diamond crystals in the X-ray, X-ray FEL Oscillators (XFELO) driven by a CW superconducting linac or an Energy Recovery Linac

(ERL) are under study [318–320]. Such an XFELO will be fully coherent, providing spectrally pure X-ray pulses.

5.4.2 Coherent harmonic generation

Following the VUV radiation obtained on the ACO storage ring, coherent harmonic generation was successfully continued on the Super-ACO storage ring FEL [321, 322]. At DUKE, the storage ring FEL in OK-4 with its sufficiently powerful super-pulses enabled us to generate second, third, fourth, fifth, and seventh coherent harmonics in the range from 37 to 135 nm [323]. The tuneability of the FEL wavelength provided for natural wavelength tuneability of the harmonic radiation. Coherent harmonics were also generated in ELETTRA [324]. At the UVSOR storage ring FEL, various features of the coherent harmonics [325] were studied, such as the influence of the synchrotron sidebands [326], the undulator, and injected laser helicity [327]. A test experiment was set-up in Sweden [328]. It used a photoinjector, with two accelerating sections, in which the beam is recirculated to reach 375 MeV and compressed in a dog-leg. The generation of circularly polarized coherent light pulses at 66 nm by seeding at 263 nm in a first modulator (planar undulator of 30 $\times$ 48 mm, 13.2 mm gap, deflection parameter of 3.52) and an APPLE-II type elliptical radiator (30 $\times$ 56 mm, 15.2 mm gap, deflection parameters for horizontal, circular, and vertical polarizations of 4.20, 3.44, and 2.98). Coherent pulses at higher harmonics in linear polarization have been produced up to the sixth order (44 nm), with 200 fs pulse duration [329].

Self-induced coherent harmonic generation was also produced in MARK-III [272], and enabled us to achieve radiation down to 36.5 nm [256]. The FEL was obtained simultaneously on the fundamental and third harmonic at Los Alamos [330]. Harmonic lasing was also performed at the Jefferson Laboratory FEL [331].

5.5 Exotic FELS

Variants of the FEL concept can be proposed by modifying the gain medium.

In the case of a gas-loaded FEL [332–334] with a gas of refractive index n_{g}, the resonance condition is modified according to

$$\pm\lambda_{\mathrm{r}} = \frac{\lambda_{\mathrm{u}}}{2n\gamma^2}\left(1 + \frac{K_{\mathrm{u}}^2}{2}\right) - \lambda_{\mathrm{u}}(n_{\mathrm{g}} - 1). \tag{81}$$

In such a case, tuneability can then be also adjusted with the gas pressure. For a given wavelength, the required electron beam energy is smaller than in the case of a conventional FEL. Experiments were carried out [335–338].

Radiation observed from electrons skimming over a diffraction grating was observed [339].

Cherenkov radiation can also be produced in dielectric loaded waveguides: a relativistic electron beam passes at grazing incidence above the surface of a dielectric loaded waveguide and excites TM-like surface waves. The longitudinal component of the evanescent electric field induced electron bunching, leading to coherent emission [340, 341]. The synchronism condition for a single-slab geometry is given by

$$\lambda_{\mathrm{r}} = \frac{2\pi d\gamma(\epsilon - 1)}{\epsilon} \tag{82}$$

with d the film thickness and ϵ the dielectric constant.

Cherenkov-based FEL radiation has been observed using a 2.5 MeV electron beam [297].

In a Smith–Purcell FEL [342–345] proposed by J. E. Walsh (1939–2000, 11th FEL Prize in 1998), the radiation is emitted when an electron passes close to the surface of a grating of period λ_{u}. The wavelength of the emitted radiation depends on the radiation observed at the angle θ from the normal is:

$$\lambda_\mathrm{r} = \lambda_\mathrm{u} \left(\frac{1}{\beta} - \sin\theta \right). \tag{83}$$

With a sufficiently high current, the electrons interact with the fields above the grating and get bunched, so the Smith–Purcell radiation is enhanced [345]. This has been experimentally demonstrated [346–348].

5.5.1 First user applications

First user applications were started [2, 349, 350] slightly more than 10 years after the first FEL oscillation, so in fact rather rapidly. First FEL applications were conducted in the infrared on the Stanford Superconducting Linear Accelerator [337, 351–354] with a strong impulse given by A. Schwettmann, on MARK III FEL [355]. Human surgery had even started, thanks to the possibility to get the required wavelength [356]. Different user activities were developed on CLIO [357–359], on FELIX [360], on FELI [361], and on FELBE [362,363]. The Jefferson Lab. infrared FEL carried out various types of user applications, such as vibrational modes in myoglobin [364], clusters [365] and industrial applications of kW UV [366]. Imaging is carried out with mm range FEL in Russia [367] and Korea [368].

In the UV, user applications started on the Super-ACO FEL in France [369], first in biology [356, 370] in 1993 for the study of the time-resolved fluorescence of the coenzyne NADH (nicotinamide adenine dinucleotide). Then, the FEL was coupled to the VUV synchrotron radiation produced in a beam line to perform a pump-probe two-colour experiments [371], enabling us to excite the system with the UV FEL and to probe the excited state with synchrotron radiation. It was first applied to surface photovoltage effects studies [372, 373]. User experiments were also carried out on various storage ring FELs, such as DUKE [374], UVSOR [375, 376], ELETTRA [377].

Generation of short wavelength radiation on small accelerators coupled to a laser by Compton backscattering is becoming attracting again nowadays in order to deliver X-rays in a compact installation. Indeed, radiation generated by Compton backscattering was easily produced with the electron beam at the origin of the FEL and the FEL itself. The two beams are already transversally overlapped and synchronized, for the FEL generation (indeed, the FEL being itself a stimulated Compton backscattering process). X-rays were generated on CLIO [378], then gamma-rays on UVSOR FEL [379], on the Super-ACO FEL [232, 380, 381], and on the DUKE FEL [382, 383] which developed a unique gamma-ray facility [384, 385].

These encouraging results in the early days of FELs let the community envision prospects for XUV (Xray Ultra Violet range) FELs [386]. The high gain regime, for which theory was actively developed, appeared to be quite suitable. First results were achieved in Livermore.

6 The high gain FEL

The study of the high gain FEL started rather early. After the small-signal low gain studies, considerations for the strong-signal case were explored in order to understand saturation. Modelling moved then towards a self-consistent theory, taking into account the evolution of both the electromagnetic field and electronic distribution during the interaction [62]. The self-consistent theories were then naturally applied to the high gain case.

6.1 Plasma type studies

The work by P. Sprangle *et al.* [64] was continued using a plasma approach in deriving the general FEL dispersion relation and in applying it to both low and high gain limits [387]. Saturation was analysed in terms of electron trapping in the space charge and ponderomotive potential and efficiencies were deduced. Scaling laws were found. A practical two stage case has been considered: a first radiation generated in a first undulator (called a 'magnetic pump' in the paper) is reflected to act as a seed (called 'pump' in the paper) of wavelength 2 cm for the second pass of the electron beam in the undulator leading to radiation at 190 μm in the second stage, for a 3 MeV energy. Further theoretical developments along such lines were carried out [151, 388–390] including the role of collective effects of the space charge [391]."The equations of the free-electron laser amplifier are generalized to include higher order modes. The density and velocity fluctuations in the entering electron beam cause noise excitation in the amplifier. The electron beam fluctuations have been studied extensively, both theoretically and experimentally, in travelling wave tubes, and hence the well-tested formalism developed for this purpose is conveniently applied to the present problem. It is found that the fluctuations put a severe constraint on the achievable exponential gain in a proposed Raman-type free-electron laser operating at optical frequencies" [65]. The gain degradation has been investigated [392].

6.2 Instability type studies: emission of coherent radiation from a self-modulated electron beam in an undulator

In Novosibirsk, at the Institute of Nuclear Physics of the former USSR Academy of Sciences, A. I. Saldin (19th FEL Prize in 2006) and A. N. Kondratenko considered very early the possibility of producing coherent radiation from a self-instability, without the use of an optical resonator [116–118]. The work was first presented at the Russian Academy of Sciences [116], then in 1979 as a preprint of the Institute of Nuclear Physics [117], and then in an international particle accelerator conference [118].

They considered a situation without an optical cavity and investigated the question of "radiative instability of the beam in an undulator". The required initial level of density oscillation for the instability to occur could result from statistic density fluctuations. They said "For sufficient length of the undulator, the resonant harmonics of density fluctuations become large enough during a pass that the modulated beam radiates from a definite section of the undulator. Such a scheme may be used as an independent source of coherent radiation or as an amplifier". A first analysis of the self-modulation in the single pass regime was first discussed by Kroll *et al.* [62] in a plasma physics context or by using high frequency device models. The instability growth was analysed using methods applied for those of storage rings, using canonical conjugate variables of the Hamiltonian describing the relative motion of electrons with radiation, and particle density. It was found that in the case of a wide beam, the amplitude of the modulation can grow exponentially and coherent power is calculated with L_g the characteristic length in which the amplitude becomes e times larger. Conditions of applications of the results were examined. Besides, "To reveal the effect of self-modulation, a knowledge of the initial level of the harmonics of beam density is necessary. In a realistic situation, if the initial conditions are not prepared in a special manner, there exists a continuous spectrum of fluctuations of density harmonics that arise from the fact that there is a finite number of particles in the beam. Hence all the harmonics in a given band width ($\Delta K/K \sim \lambda_\mathrm{u}/(2\pi L_g)$) will become unstable and grow by a few times in the length L_g". The evolution of the width of the harmonics after the pass of a given length is analysed, and its corresponding correlation length is estimated (of the order of $L_g/(2\gamma_{\|})$.

The reduction down to shorter wavelengths has also been analysed [393] using the Russian approach. The possibility of using a high gain FEL amplifier to start from noise was first considered on a storage ring to produce X-ray radiation. Examples were given for a 2.8 MeV energy, 100 A current, 50 μm rad emittance electron beam, helical undulator ($\lambda_\mathrm{u} = 20$ mm, $B_\mathrm{u} = 0.3$ T) leading to radiation at 450 μm with a growth length of 14 cm, or for a 10.2 MeV energy, 30 kA current, 50 μm rad emittance electron beam, helical undulator ($\lambda_\mathrm{u} = 60$ mm, $B_\mathrm{u} = 0.1$ T) leading to radiation at 100 μm with a growth length of 22 cm. Also considered was the case of a 20.4 MeV energy, 14 kA current, 5 nm.rad emittance electron beam in a storage ring, helical undulator ($\lambda_\mathrm{u} = 70$ mm, $B_\mathrm{u} = 0.2$ T, length of 15 m) leading to radiation at 5 nm with a growth length of 2 m and an output peak power of the order of 1 TW. Conditions are given for the energy spread (that should be less than 10^{-3}) and angular spread (that should be less than $3 \times\ 10^{-5}$). These parameters are in fact not so different from the ones of the present LCLS (Linac Coherent Light Source) X-ray FEL, as discussed later. Instead of a storage ring, a linear accelerator is presently used on LCLS. It took also several decades to achieve, technologically speaking, the required electron beam parameters.

This pioneering work indeed was considered for the first time to start from the spontaneous emission to amplify it in the high gain regime until saturation, in the case of an infrared FEL with a 10 MeV electron beam [117, 118]. It is now usually called self-amplified spontaneous emission, in reference to the amplified spontaneous emission in conventional lasers. Its sketch is shown in Fig. 51.

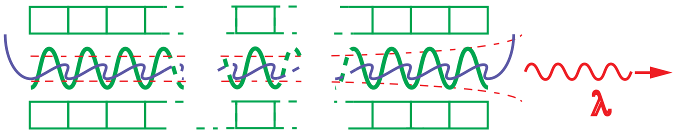


Fig. 51: FEL self-amplified spontaneous emission (SASE) configuration: the spontaneous emission emitted in the beginning of the undulator is amplified in one single pass. Operation at short wavelengths requires high beam energies for reaching the resonant wavelength, and thus long undulators (0.1–1 km for 0.1 nm) and high electron beam density (small emittance and short bunches) for ensuring a sufficient gain.

6.3 Hamiltonian-type studies

The Hamiltonian description was also applied to the high gain case [93]. Using the Hamiltonian description in the moving frame [394], R. Bonifacio (1940–2016) *et al.* investigated cooperative and chaotic transition (a kind of phase transition from a regime of small gain amplifier to that of a large gain amplifier): “below this threshold value the electrons radiate weakly and almost independently whereas above threshold the electrons strongly interact via the emitted radiation field. In the latter case the particles exhibit strong self-bunching and give rise to cooperative emission of radiation” [395]. “The exact threshold value w_T of the coupling parameter is analytically obtained by investigating, for any number of electrons N, the stability property of an initial condition with zero field excitation and totally unbunched electrons (i.e. electrons uniformly spread over an optical wavelength)”. Results are compared to simulations. It is pointed out that the transition can be stimulated by noise within the interaction volume.

The analysis was continued by the introduction of the notion of collective instability [396]. The electrons communicate with each other through the radiation and the space charge field. Thus, they ‘self-bunch’ on the scale of the radiation wavelength periods. The electrons have nearly the same phase and emit collectively coherent synchrotron radiation. R. Bonifacio, C. Pellegrini (12th FEL Prize, in 1999) and L. M. Narducci introduced the plasma frequency and the Pierce parameter. The exponential growth is found as a solution for the cubic equation which has one real and two complex conjugate roots. The instability condition could then be derived in terms of Pierce parameter and spectral detuning. The notion of lethargy, the “time required for the initial pulse to build up” was also discussed. It is also found that in the high gain case, the maximum growth is found for zero detuning. This regime is called the self-amplified spontaneous emission (SASE) regime. SASE has been studied in detail [397,398], in particular with issues regarding the transient behaviour of the system, using the Maxwell–Vlasov equations [399]. Effects of harmonics, space charge, and electron energy spread on the collective instability are discussed [400]. The model covers both Compton and Raman regimes [401]. The regime of superradiance for a high gain FEL was analysed [402]. Prospective SASE sources were designed [398]. The diagram block in the SASE case is shown in Fig. 52. Semi-analytical models were developed [403, 404]. High gain single-pass free electron laser dynamics and pulse propagation effects were also considered [405].

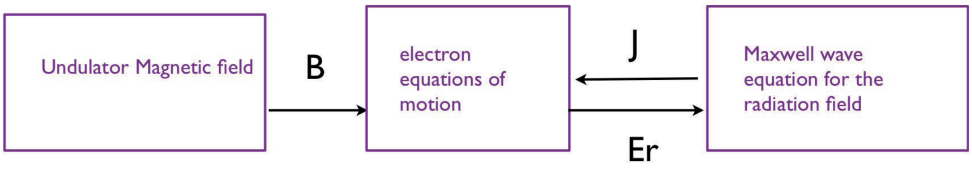


Fig. 52: SASE diagram block with the process of collective instability arising from communication from the neighbouring electrons.

6.4 One-dimensional high gain FEL modelling

6.4.1 The coupled system of equations

With high density electron beams and long undulators, a strong bunching takes place (space charge) and the change in electric field can no longer be neglected. Thus, the FEL is treated via a set of coupled equations [396, 398, 401]:

- the coupled pendulum equation, describing the phase space evolution of the particles under the combined undulator magnetic field and electric field of the optical wave;
- the evolution of the optical field in the presence of an electronic density and current;
- the evolution of the bunching coupled to the longitudinal space charge forces, enabling us to evaluate the electronic density and current.

The electronic density and the current resulting from the electrons in the undulator are first evaluated to treat the light wave evolution.

6.4.1.1 Radiation field evolution

The radiated field now depends on the longitudinal coordinate as

$$E_x(s,t) = E_x(s)\exp\left[-\mathrm{i}(ks-ct)\right]. \tag{84}$$

Its evolution is ruled by the Maxwell equation

$$\left[\frac{\partial^2}{\partial s^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]E_x(s,t) = \mu_0\frac{\partial j_x}{\partial t} + \frac{1}{\epsilon_\mathrm{o}}\frac{\partial \rho}{\partial x}$$

with j_x the average over the electron beam cross-section of the transverse electron peak current. In one-dimensional FEL theory, the electronic density is assumed to be independent of x, so $\frac{\partial \rho}{\partial x} = 0$. The transverse current is mainly due to the electrons in the wiggler, and to a small extent to the radiation field.

The phase of E_x may vary with s, i.e. the FEL phase velocity, may differ slightly from that of a plane electromagnetic wave at the speed of light c. Inserting the electric field expression in the Maxwell equation, it becomes

$$\left[2\mathrm{i}kE_x'(s) + E_x''(s)\right]\exp\left[-\mathrm{i}(ks-ct)\right] = \mu_0\frac{\partial j_x}{\partial t} + \frac{1}{\epsilon_\mathrm{o}}\frac{\partial \rho}{\partial x}. \tag{85}$$

Under the slowly varying envelope approximation (SVEA) the electric field does not change much over a few undulator periods even though its increase over the whole undulator length is large. The change in one wavelength is even smaller, so is its first-order derivative. It can be written as $|E_x'(s)|\lambda \ll |E_x(s)|$ and $|E_x'(s)| \ll k|E_x|$. However, one should keep the first-order derivative in order to describe the FEL growth as a function of the distance in the undulator. One can neglect the change of slope over one undulator period λ_u, as $|E_x''(s)|\lambda \ll |E_x(s)|$ and $|E_x''(s)| \ll k|E_x|$. The second-order derivative can be neglected. So, in the paraxial approximation, one has $\frac{\mathrm{d}}{\mathrm{d}t} = c(\frac{\partial}{\partial s} + \frac{1}{v_{\mathrm{s}}}\frac{\partial}{\partial t})$. The field evolution then becomes

$$\frac{\mathrm{d}E_x(s)}{\mathrm{d}t} = -\mathrm{i}\frac{\mu_o}{2k}\frac{\partial j_x}{\partial t}\exp\left[-\mathrm{i}(ks - ct)\right]. \tag{86}$$

6.4.1.2 Current sources

The current comes first from the displacement of the electron because of the undulator field. Then, a bunching takes place and, if important, an electron can be affected by the neighbouring electrons due to space charge forces.

The transverse current source due to the electron movement in the undulator is expressed using the electron transverse velocity as

$$\overrightarrow{j} = \rho_{\mathrm{e}}\overrightarrow{v} \quad \text{so} \quad j_x = j_s\frac{v_x}{v_{\mathrm{s}}} \simeq j_s\frac{K_{\mathrm{u}}}{\gamma}\sin\left(k_{\mathrm{u}}s\right) \tag{87}$$

The field evolution becomes $\frac{\mathrm{d}E_x(s)}{\mathrm{d}t} = -\mathrm{i}\frac{\mu_o K_{\mathrm{u}}}{2k\gamma}\frac{\partial j_s}{\partial t}\exp\left[-\mathrm{i}(ks - ct)\right]\sin\left(k_{\mathrm{u}}s\right)$. While the electrons and the light wave interact, periodic density modulation (micro-bunching) is taking place, and the current can be developed as a function of the ponderomotive phase ψ:

$$\tilde{j} = \tilde{j_o} + \tilde{j_1}\exp\left(\mathrm{i}\psi\right) \tag{88}$$

where the term $\exp\left(\mathrm{i}\psi\right)$ represents the bunching,

$$\frac{\partial j_s}{\partial t} = \frac{\partial j_s}{\partial \psi}\frac{\partial \psi}{\partial t} = -\mathrm{i}\omega\tilde{j_1}\mathrm{e}^{\mathrm{i}\psi} = -\mathrm{i}\omega\tilde{j_1}\exp\left(\mathrm{i}k(s - ct) + \mathrm{i}k_{\mathrm{u}}s\right).$$

Combining the expression of the current and the electric field derivative terms, the field evolution becomes

$$\frac{\mathrm{d}E_x(s)}{\mathrm{d}t} = -\frac{\mu_o c K_{\mathrm{u}}}{2\gamma}\tilde{j_1}\exp\left[\mathrm{i}k(s - ct) + \mathrm{i}k_{\mathrm{u}}s\right]\exp\left[-\mathrm{i}(ks - ct)\right]\frac{\mathrm{e}^{(\mathrm{i}k_{\mathrm{u}}s)} - c.c.}{2} = -\frac{\mu_o c K_{\mathrm{u}}}{4\gamma}\tilde{j_1}[1+\exp\left[\mathrm{i}2k_{\mathrm{u}}s\right.$$

with *c.c.* meaning complex conjugate. As the complex field amplitude is slowly varying on the scale of the wavelength λ, it can be driven by a current averaged longitudinally over several wavelengths. One can thus average the transverse current longitudinally. The phase factor carries out two oscillations per period and averages to zero. Thus, one gets

$$\frac{\mathrm{d}E_x(s)}{\mathrm{d}t} = -\frac{\mu_o c K_{\mathrm{u}}}{4\gamma}\tilde{j_1}[1 + \exp\left[\mathrm{i}2k_{\mathrm{u}}s\right]]. \tag{89}$$

Then, let's consider the space charge term due to the longitudinal field. The electric field is created by the modulation ρ_e of the charge density in the electron bunch, as illustrated in Fig. 53.

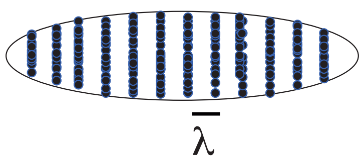


Fig. 53: Schematic representation of the slices of length λ_{r} along a bunched electron beam

According to Maxwell's equation, $\nabla.\overrightarrow{E} = \frac{\rho_e(\psi,s)}{\epsilon_\mathrm{o}}$. Similarly to the current, one can develop the electronic density as

$$\rho_e(\psi, s) = \rho_{eo} + \tilde{\rho_1} \exp{(\mathrm{i}\psi)}. \tag{90}$$

It then becomes $\frac{\mathrm{d}E_s}{\mathrm{d}t} = \frac{\tilde{\rho_1}}{\epsilon_\mathrm{o}} \exp{(\mathrm{i}[(k + k_\mathrm{u})s - \omega t])}$, so, the amplitude of the induced space charge longitudinal field is

$$E_s = -\mathrm{i}\frac{1}{\epsilon_\mathrm{o}(k + k_\mathrm{u})}\tilde{\rho_1} \approx -\mathrm{i}\frac{\mu_o c^2}{\epsilon_\mathrm{o}\omega}\tilde{j_1}. \tag{91}$$

6.4.1.3 Energy change due to the field

In the low gain low field case (in neglecting the field variation over one undulator pass), one has

$$\frac{\mathrm{d}\eta}{\mathrm{d}s} = -\frac{eE_\mathrm{l}K_\mathrm{u}}{2\gamma^2 m_\mathrm{o}c^2}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]\sin{(\xi + \phi)}.$$

In the high field case with a bunched beam, one gets

$$\frac{\mathrm{d}\eta}{\mathrm{d}s} = -\frac{eK_\mathrm{u}}{2\gamma^2 m_\mathrm{o}c^2}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]\mathrm{Re}(E_x \mathrm{e}^{(\mathrm{i}\psi)}) - \frac{e}{\gamma_\mathrm{r} m_\mathrm{o}c^2}\mathrm{Re}(E_s \mathrm{e}^{(\mathrm{i}\psi)}).$$

The first term corresponds to the electron motion and the second one to the space charge contribution. One has

$$\frac{\mathrm{d}\eta}{\mathrm{d}s} = -\frac{e}{m_\mathrm{o}c^2\gamma_\mathrm{r}}\mathrm{Re}\left[\left(\frac{K_\mathrm{u}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]E_x}{2\gamma_\mathrm{r}} + E_s\right)\mathrm{e}^{(\mathrm{i}\psi)}\right]. \tag{92}$$

Due to the bunching process, the electrons are grouped periodically in the electron bunch with a spatial modulation equal to the wavelength λ_r and its harmonics of order n λ_{rn}, and the electron bunch is described as a series of slices of length λ_r and number of electrons N_slice, which correspond to slices of length 2π in the phase ψ representation. The shape function $S(\psi)$ is given by

$$S(\psi) = \sum_{j=1} N_\mathrm{slice}\delta(\psi - \psi_j(t)), \tag{93}$$

j being the current electron number. By developing in Fourier series, one gets

$$S(\psi) = \frac{c_o}{2} + \mathrm{Re}\Big(\sum_{k=1}^{\infty} c_k \exp{(\mathrm{i}k\psi)}\Big), \quad c_k = \int_0^{2\pi} S(\psi)\exp{(\mathrm{i}k\psi)}\mathrm{d}\psi. \tag{94}$$

In the case of the first harmonic, the current becomes

$$j_1 = -ecn_\mathrm{e}\frac{2}{N_\mathrm{slice}}\sum_{j=1} N_\mathrm{slice}\exp{(\mathrm{i}\psi_j)}. \tag{95}$$

6.4.1.4 High signal set of equations

The FEL dynamics is now ruled by a set of coupled equations. The source current, depending on the electron bunching, evolves as $j_1 = -ecn_\mathrm{e}\frac{2}{N_\mathrm{slice}}\sum_{j=1} N_\mathrm{slice}\exp{(\mathrm{i}\psi_j)}$. The transverse electric field evolution is ruled by the current source term, depending on the electrons in the undulator, as $\frac{\mathrm{d}E_x(s)}{\mathrm{d}t} = -\frac{\mu_o c K_\mathrm{u}}{4\gamma}\tilde{j_1}$. The phase evolution is related to the energy exchange as

$$\frac{\mathrm{d}\psi_j}{\mathrm{d}s} = 2k_\mathrm{u}\eta_n, \quad j = 1, \ldots, N_\mathrm{slice}. \tag{96}$$

The energy exchange is governed by the current term due to the electron movement and to the space charge induced electric field:

$$\frac{\mathrm{d}\eta}{\mathrm{d}s} = -\frac{e}{m_\mathrm{o}c^2\gamma_\mathrm{r}}\mathrm{Re}\left[\left(\frac{K_\mathrm{u}\left[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)\right]E_x}{2\gamma_\mathrm{r}} + E_s\right)\mathrm{e}^{(\mathrm{i}\psi)}\right]. \tag{97}$$

These equations are usually solved numerically. One can obtain however an analytic solution under certain approximations.

6.4.2 Evolution of the light wave in the high gain regime of FEL

6.4.2.1 The FEL cubic equation

In the case of a rather 'small' periodic density modulation, a normalized particle distribution function, obeying the Vlasov equation, is defined. After mathematical manipulation [396, 402], one can show that the radiation amplitude E_x satisfies

$$\frac{\dddot{E}_x}{\Gamma^3} + 2\mathrm{i}\frac{\eta}{\rho_\mathrm{FEL}}\frac{\ddot{E}_x}{\Gamma^2} + \left[\frac{k_\mathrm{p}^2}{\Gamma^2} - \frac{\eta^2}{\rho_\mathrm{FEL}^2}\right]\frac{\dot{E}_x}{\Gamma} - \mathrm{i}E_x = 0 \tag{98}$$

ρ_FEL is the so-called Pierce parameter, or FEL parameter. It depends on the electron beam density and energy and on the undulator characteristics (deflection parameter, Bessel function term, undulator wavenumber):

$$\rho_\mathrm{FEL} = \left[\frac{K_\mathrm{u}[JJ]\omega_\mathrm{p}}{4\omega_\mathrm{u}}\right]^{2/3} = \frac{1}{2\gamma k_\mathrm{u}}\left(\frac{\mu_\mathrm{o}e^2K_\mathrm{u}^2[JJ]^2k_\mathrm{u}n_\mathrm{e}}{4m_\mathrm{o}}\right)^{1/3}, \tag{99}$$

where Γ, the gain parameter, is proportional to the Pierce parameter,

$$\Gamma = 2k_\mathrm{u}\rho_\mathrm{FEL}, \tag{100}$$

and k_p, the space charge parameter, is given by $k_\mathrm{p} = \frac{\omega_\mathrm{p}}{c\gamma}\sqrt{\frac{2\lambda}{\lambda_\mathrm{u}}}$ with ω_p the plasma pulsation $\omega_\mathrm{p} = \sqrt{\frac{4\pi e^2 n_\mathrm{e}}{m_\mathrm{o}}}$. In the specific case of $\eta = 0$ (on resonance) and for $k_\mathrm{p} = 0$, i.e. for negligible space charge, the cubic equation takes its simplest form, as

$$\frac{\dddot{E}_x}{\Gamma^3} - \Gamma^3\mathrm{i}E_x = 0. \tag{101}$$

Considering the electric field expressed as $\approx \mathrm{e}^{(\mathrm{i}\kappa s)}$, it becomes $\kappa^3 = \mathrm{i}\Gamma^3$ with three solutions:

$$\kappa_1 = -\mathrm{i}\Gamma, \quad \kappa_2 = (\mathrm{i} + \sqrt{3})\Gamma/2, \quad \kappa_3 = (\mathrm{i} - \sqrt{3})\Gamma/2.$$

κ_2 leads to an exponential growth of the electric field.

6.4.2.2 The FEL power growth and evolution of the light wave in the high gain regime

The power grows as

$$E_x(s) = E_{x0}\exp\left(s/L_\mathrm{go}\right), \quad L_\mathrm{go} = \frac{1}{\sqrt{3}\Gamma} = \frac{1}{\sqrt{3}}\left(\frac{4m_\mathrm{o}\gamma^3}{\mu_\mathrm{o}\mathrm{e}^2K_\mathrm{u}^2[JJ]^2k_\mathrm{u}n_\mathrm{e}}\right)^{1/3}. \tag{102}$$

The bunching factor B evolves similarly. It is noticeable that there is amplification at resonance, this feature differs from the small signal gain case. At the beginning of the undulator, the three terms of the cubic equation do contribute to the change in the field intensity and the exponential growth is not dominant, the bunching takes place. This regime is called the 'lethargy'. Solving the cubic equation for a non-zero detuning (slightly off resonance) provides the dependence of the imaginary solution with detuning, i.e. the gain bandwidth. It has a maximum for zero detuning and decreases for both positive and negative detunings. From the analysis of the behaviour, one can deduce that the FEL bandwidth is given by the Pierce parameter:

$$\frac{\Delta\lambda}{\lambda} = \rho_{\mathrm{FEL}}. \tag{103}$$

Adopting the same type of evaluation as in the small signal gain case, one estimates that the saturation power of the radiated field is the electron beam power multiplied by the gain bandwidth:

$$P_{\mathrm{sat}} = \rho_{\mathrm{FEL}} E I_{\mathrm{p}}, \tag{104}$$

with E the electron beam energy and I_{p} the peak current. Since the radiation pulse duration is close to that of the electron bunch, the Pierce parameter gives the efficiency of the FEL, i.e. the fraction of the beam energy given to the radiation field. Typically, the saturation power is reached after roughly 20 gain lengths, at the saturation length L_{s}.

$$L_{\mathrm{s}} \approx 20 L_{\mathrm{go}} \approx \frac{1}{\sqrt{3}\Gamma} = \frac{20\lambda_{\mathrm{u}}}{4\pi\sqrt{3}\rho_{\mathrm{FEL}}} = \frac{5\lambda_{\mathrm{u}}}{\pi\sqrt{3}\rho_{\mathrm{FEL}}}. \tag{105}$$

So the saturation can be achieved with N_s, given by

$$N_s = \frac{L_{\mathrm{s}}}{\lambda_{\mathrm{u}}} = \frac{5}{\pi\sqrt{3}\rho_{\mathrm{FEL}}}. \tag{106}$$

The FEL parameter defines the growth rate, measured in undulator periods.

The start-up comes from the spontaneous emission noise. It is followed by an exponential growth due to a collective instability (self-organization of the electrons from a random initial state). When the power saturates, there is a cyclic energy exchange between the electrons and the radiated field and a consequent change of power which corresponds to rotations in phase space. Growth and bunching also occur on the harmonics of the fundamental wavelength. The number of radiated coherent photons per electron at saturation $N_{\mathrm{coh.ph}}$ is given by $N_{\mathrm{coh.ph}} \sim \frac{\rho_{\mathrm{FEL}} E}{E_{\mathrm{ph}}}$, with E_{ph} the photon energy. For photons of 10 keV with a beam of 15 GeV, a Pierce parameter of 0.001, at 10 keV, $N_{coh.ph} \sim 1500$!

6.4.2.3 The SASE spectral and temporal properties

The uncorrelated trains of radiation, which result from the interaction of electrons progressing jointly with the previously emitted spontaneous radiation, lead to spiky longitudinal and temporal distributions, apart from single spike operation for low charge short bunch regime [406, 407]. The emission usually presents poor longitudinal coherence properties. There is some particularity of the temporal structure of the SASE pulse. Because the photons move faster than the electrons, the radiation emitted by one electron moves ahead and slips by one wavelength per undulator period, so for the total undulator length by $N_{\mathrm{u}}\lambda$. The analysis [402] of the effect of slippage for an electron bunch of finite length, when the slippage effect cannot be neglected, shows that the interaction between the electrons is only effective over a cooperation length, the slippage in one gain length. In a one-dimensional model the cooperation length can be written as

$$L_{\mathrm{coop}} = \frac{\lambda}{2\sqrt{3}\rho_{\mathrm{FEL}}}. \tag{107}$$

Since the initial noise varies along the bunch length, the output radiation pulse consists of a series of spikes of random intensity separated by a distance proportional to the cooperation length [408]. In the case of spontaneous radiation, the intensity along the pulse varies randomly in each wavelength. For SASE at saturation, the interaction between electrons and their emitted radiation generates a number of spikes of random intensity and duration proportional to the cooperation length. The number of spikes in a pulse is given by the ratio of the bunch length to the cooperation length. The intensity in each spike fluctuates from pulse to pulse. There is no correlation between the phases of different spikes. The statistical distribution of the total intensity, summed over all spikes, is given by a gamma distribution function [407]. The line width, in a SASE FEL, is inversely proportional to the spike length, and not to the bunch length. The width is of the order of the FEL parameter. In consequence, a SASE radiation pulse is not Fourier transform limited, except for the case of an electron bunch length shorter than the cooperation length, when a single spike is produced. Examples of spikes are shown in Fig. 54.

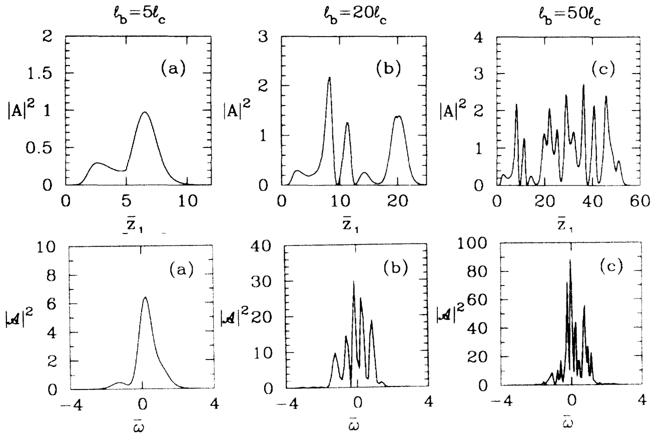


Fig. 54: SASE spikes for different bunch lengths. Temporal (up) and spectral (down) distributions, from [408]

On a single pass FEL, transverse coherence results from the electron beam emittance (which should be of the order of the emitted wavelength) and from possible optical guiding. The optical guide can arise from gain guiding (quadratic gain medium [45]) or from the contribution of a refractive index [409]. As a consequence, the undulator can even be longer than a few Rayleigh lengths!

6.5 Three-dimensional analysis

The role played by diffraction is analysed: exponentially growing modes, which have a profile independent of the longitudinal coordinate, exist and a dimensionless parameter, which is proportional to the radius of the electron beam and independent of the interaction length, determines whether diffraction is important [410]. Then, the small-signal gain of the fundamental exponentially growing mode of the high gain free electron laser is calculated, taking both diffraction and electron energy spread into account. As the electron beam radius is reduced, the gain bandwidth increases by a large amount [411]. A two-dimensional analysis using the properties of optical fibres shows that optical guiding can take place in a free electron laser [409, 412, 413].

Three-dimensional analysis of coherent amplification and self-amplified spontaneous emission in free electron lasers was carried out [414, 415] using the three-dimensional Maxwell–Klimontovich equation. The Klimotovich distribution function takes into account the discreteness of the electrons.

The radiation field, represented by a complex amplitude, which is the slowly varying part of the full amplitude, satisfies Maxwell equation. Slowly varying and high- frequency components of the electronic distribution are treated separately. It is found that transversally, optical guiding takes place for high gain FEL [410, 411]. Electron correlation, transverse radiation profiles, spectral features, transverse coherence, and intensity characteristics are analysed, as shown in Fig. 55. The results, which agree with recent microwave experiments, are applied to proposed schemes for generation of short-wavelength coherent radiation. Corrections terms (from the three-dimensional theory) can be introduced [416, 417].

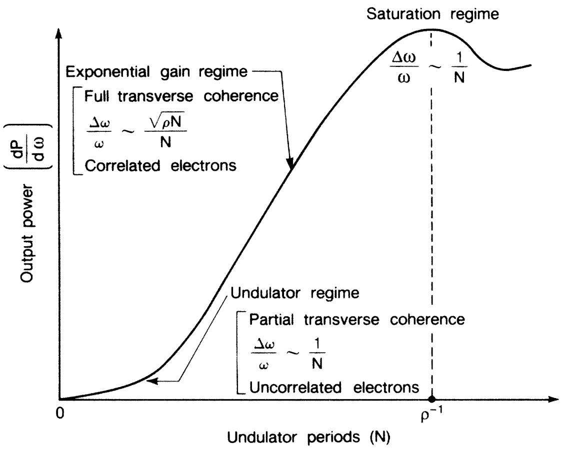


Fig. 55: Schematic representation of SASE properties from the three-dimensional analysis by K. J. Kim in [414]

Semi-analytical models of SASE FEL based on the logistic FEL equation [404] are developed. They include diffraction, beam quality and pulse propagation.

6.6 Conditions for SASE amplification

Conditions for SASE amplification are detailed below.

6.6.1 Emittance requirement

There should be a proper transverse matching (size, divergence) between the electron beam and the photon beam along the undulator for insuring a proper interaction. It means that the emittance should not be too large at short wavelength. The FEL gain increases with the beam current provided that

$$\frac{\epsilon_n}{\gamma} < \frac{\lambda}{4\pi}. \tag{108}$$

High power short wavelength FELs require thus low emittance electron beams (much smaller than 100π mm mrad and peak currents of the order of 100 A.

6.6.2 Energy spread requirement

The electron beam should be rather 'cold', its energy spread should be smaller than the bandwidth, i.e.

$$\frac{\sigma_\gamma}{\gamma} < \rho_{\mathrm{FEL}}. \tag{109}$$

6.6.3 Rayleigh length requirement

The radiation diffraction losses should be smaller than the FEL gain, i.e. the Rayleigh length should be larger than the gain length:

$$Z_r > L_{\mathrm{go}}. \tag{110}$$

For long undulators, intermediate focusing is then put between undulator segments.

Reviews of the free electron laser theory are presented in [124, 418]. Different numerical codes can be used for FEL calculations, such as GINGER [419] by W. Fawley (25th FEL Prize in 2014), GENESIS [420] written by S. Reiche (21st FEL Prize in 2010), PERSEO [421] by L. Giannessi (24th FEL Prize in 2013), MEDUSA [422] by H. Freund, PUFFIN [423] by B. Mac Neil and L. T. Campbell, TDA [424] by J. S. Wurtele etc.

6.7 High gain up-frequency conversion

6.7.1 High gain harmonic generation theory

As for the case of the low gain FEL, harmonic generation can take place, as schematized in Fig. 56.

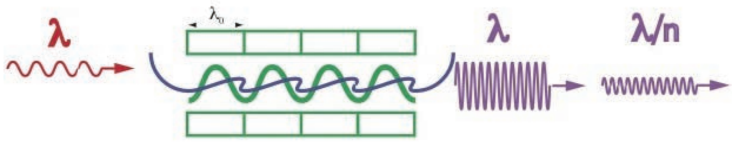


Fig. 56: High gain harmonic generation

Three-dimensional analysis of harmonic generation in high gain free electron lasers has been carried out [425] in the case of a planar undulator using the coupled Maxwell–Klimontovich equations that take into account non-linear harmonic interactions. “Strong bunching at the fundamental wavelength can drive substantial bunching and power levels at the harmonic frequencies”. “Each harmonic field is a sum of a linear amplification term and a term driven by nonlinear harmonic interactions. After a certain stage of exponential growth, the dominant nonlinear term is determined by interactions of the lower nonlinear harmonics and the fundamental radiation. As a result, the gain length, transverse profile, and temporal structure of the first few harmonics are eventually governed by those of the fundamental. Transversely coherent third-harmonic radiation power is found to approach 1% of the fundamental power level for current high-gain FEL projects” [425].

Non-linear harmonic generation in high gain free electron lasers [426] can also be treated semi-analytically using a theoretical ansatz and fitting methods, providing “the most significant aspects of the high-gain free-electron laser dynamics” [426]. Expressions are found for the growth of the laser power, of the e-beam-induced energy spread, and of the higher-order non-linearly generated harmonics. They are applied to treat pulse propagation and non-linear harmonic generation in free electron laser oscillators [148], two harmonic undulators, and harmonic generation in high gain free electron lasers [427].

Different variants have been considered.

In the high gain harmonic-generation (HGHG) configuration [428–432], a small energy modulation is imposed on the electron beam by its interaction with a seed laser in a first undulator (the modulator) tuned to the seed frequency, it is then converted into a longitudinal density modulation thanks to a dispersive section (chicane) and in a second undulator (the radiator), which is tuned to the nth harmonic of the seed frequency, the microbunched electron beam emits coherent radiation at the harmonic frequency of the first one, which is then amplified in the radiator until saturation is reached. By some means, it recalls the optical klystron scheme. The HGHG configuration is shown in Fig. 57, The seed signal should then overshot the shot noise from the start-up SASE radiation.

In such a way, the higher-order harmonic components of the density modulation induced by the FEL process are exploited in a ‘harmonic converter’ configuration, to multiply the frequency, and extend the original spectral range of operation of the FEL. To be more efficient, it is combined to the ‘fresh

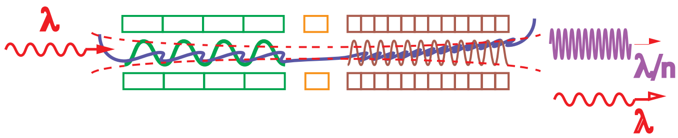


Fig. 57: High gain harmonic generation: seeding is performed in two stages, the first stage is seeded with an external laser, where a density modulation of the electron bunch takes place, whereas the second stage is seeded by the FEL from the first stage while the undulator is set on a harmonic of the radiation from the first stage; and the electron bunch radiates coherently after passing through a dispersive magnetic chicane.

bunch technique' [433] where a proper delay is applied on the electron bunch with respect to the optical path so that a 'non-heated' part of the electron bunch is used for the second stage.

The HGHG can be put in cascade with a series of modulator/radiator undulators, enabling potentially effective frequency conversion. L. H. Yu considered that "the cascading of several HGHG stages (35) can provide a route for x-ray generation using current near-ultraviolet seed laser performance. In this approach, the output of one HGHG stage provides the input seed to the next undulator. Each stage is composed of a modulator, dispersion section, and radiator. Within a single stage, the frequency is multiplied by a factor of 3 to 5. For each stage, the coherent radiation produced by the prebunched beam in the radiator at the harmonic of the seed is many orders of magnitude higher in intensity than the SASE generated. In a specific example (35), after cascading five HGHG stages, the frequency of the output is a factor $5 \times 5 \times 5 \times 4 \times 3$, i.e. 1500 times the frequency of the input seed to the first stage. Dispersion sections are placed between stages to shift the radiation to fresh portions (36) of the electron bunch to avoid the loss of gain due to the energy spread induced in the previous stage" [429]. Shot noise at the different stages can then become an issue [434]. Schemes for reducing this shot noise are proposed [435].

In the harmonic cascade configuration (see Fig. 58), the wavelength ratio of the two stages is a ratio of integers [436, 437].

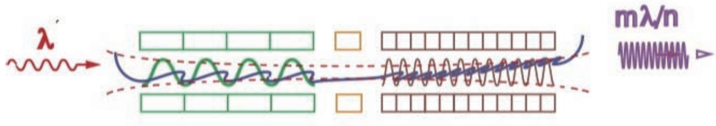


Fig. 58: Harmonic cascade configuration: seeding is performed in two stages, the first stage is seeded with an external laser, whereas the second stage is seeded by the FEL from the first stage while the undulator is set so that the wavelength ratio of the two stages is a ratio of integers.

In particular seeding cases, the seeded FEL can become super-radiant [436], leading to further pulse shortening and intensity increase. Depending on the respective electron bunch and slippage length, complex spatio-temporal deformation of the amplified pulse can lead ultimately to a FEL pulse splitting effect [438].

6.7.2 Echo enabled harmonic generation

In the echo enabled harmonic generation (EEHG) [439] scheme (see Fig. 59), two successive laser–electron interactions are performed, using two undulators, in order to imprint a sheet-like structure in phase space. As a result, higher-order harmonics can be obtained in an extraordinary efficient way.

Figure 60 shows the imprinted modulation applied in the echo scheme.

Schemes derived from EEHG, such as the triple mode chicane, open perspectives for very high up-frequncecy conversion for short wavelength (nm) light of short duration at moderate cost [441]. The echo concept can also be applied to storage ring based light sources [442]. EEHG opens the way to

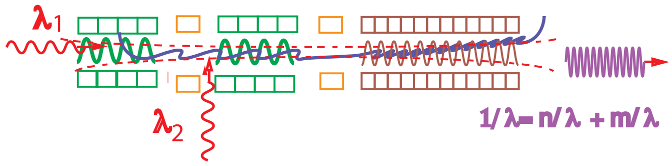


Fig. 59: FEL EEHG: a coherent source tuned on the resonant wavelength of the undulator applies a first energy modulation, electrons move according to their energy in the chicane where a second energy modulation is applied, imprinting a fine structure in phase space.

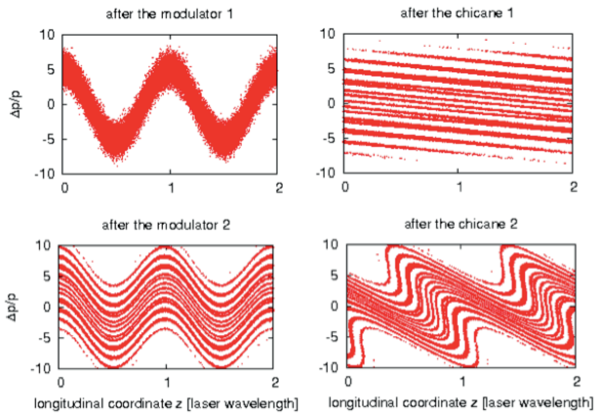


Fig. 60: Evolution of the particle phase space along the EEHG stages. Phase space of the beam after the first modulator (top left), the first chicane (top right), the second modulator (bottom left), and the second chicane (bottom right). Horizontal axis: phase, vertical axis: relative energy, from [440].

shorter wavelengths when operating on a high-order harmonics of the seed wavelength. Echo has no equivalent in classical optics.

7 Single-pass short wavelength FEL experimental results

These encouraging results in the early FEL research let the community envision prospects for XUV FELs [386] and soft X-ray FELs [443–445] to be installed on storage rings, the accelerators providing the best performance at that time (energy spread of $\sim 0.1\%$, peak current of a few hundreds of ampere).

The high gain regime, for which theory was actively developed, appeared to be quite suitable. First, high gain FEL experiments took place on oscillators, then experiments aiming at demonstration of SASE at intermediate wavelengths were undertaken. However, the decrease in wavelength was accompanied by an improvement of accelerator technology, enabling us to fulfil the requirements for SASE. The advent of the photoinjector became really crucial for ensuring the development of FELs at shorter wavelengths. In addition, the requirements in terms of linear accelerator performance for future colliders met the needs of X-ray FELs, and the technological developments were fruitfully applied within the FEL community. In particular, the high electron beam density also suited for getting a short gain length.

7.1 Towards VUV X-ray FELs?

7.1.1 Limits of storage ring driven FEL for short wavelength FELs

The electron beam quality is an essential contribution to the success of a given FEL. Indeed, the energy spread should be sufficiently small to enable a proper bunching. At the end of the twentieth century, the shortest FEL wavelength on an oscillator has been achieved on a storage ring FEL [263, 264]. However, the electron beam recirculation is limiting the output power according to the Renieri limit, and compatibility of the use of the storage ring with normal synchrotron radiation use became an issue.

7.1.2 Early development of photoinjectors

With respect to conventional thermoionic guns, photoinjectors [446, 447] in which a laser illuminated photocathode is located directly in the high gradient accelerating cavity, can enable us to provide a high quality electron beam. Compared to a thermionic gun, the current density can be very high so that bunching is not necessary. The time structure can also be controlled by the laser beam, and matched into the RF accelerators without degrading the emittance. The electrons produced on the photocathode surface are quickly accelerated in a RF cavity in order to limit the emittance blow-up due to the space charge force. Several laboratories have initiated the development of photoinjectors.

The first photocathode-driven electron beam enabling FEL was achieved at Stanford on the MARK III linear accelerator [448]. The gun used a LaB_6 cathode, illuminated by a tripled Nd:Yag laser, leading to an energy spread of 0.8% and an horizontal (vertical) emittance of $8(4)\pi$ mm mrad.

At Los Alamos, the facility has been modified to target FEL oscillation in the visible. For this purpose, the thermoionic gun was replaced by a photoinjector [449–452]. The pioneering work of the Los Alamos team on photo-injectors was recognised by the FEL prize 2017 awarded to Bruce Carlsten and Richard Sheffield (27^{th} FEL prize 2017). The photoinjector (26 MeV/m at the CsK_2Sb cathode at 1.3 GHz in the $\pi/2$ mode) produced 6 MeV, 300 A, 15 ps electron pulses at 22 MHz repetition rate. The drive laser was a Nd-YLF laser at 527 nm with very low phase ($<$ 1 ps) and amplitude ($<1\%$) jitters. Changing from the thermoionic gun to the photoinjector enabled to reduce the emittance by a factor of 4 and the energy spread by nearly a factor of two. Typically, the electron phase space density could be larger by one order of magnitude. B. Carlsten proposed the idea of emittance compensation [453], leading to a significant reduction of the normalized emittance with respect to usual thermoionic guns.

Another photoinjector was developed at Brookhaven National Laboratory (Center for Accelerator Physics) [454]. In the frame of a SLAC/BNL, UCLA collaboration, a research and development (R&D) effort was launched on the development of a photoinjector. A 4.5 MeV 1.5 cell standing wave RF (2.856 GHz) photoinjector gun based on the Brookhaven design, using a copper cathode, was completed at UCLA [455]. Driven by sub 2 ps pulses of UV (266 nm) light (up to 200 μJ/pulse) and powered by a SLAC XK5 type klystron (24 MW, 4 μs), it could generate up to 3 nC charge. Accelerating gradients of up to 100 MV/m were achieved. A 0.25 kA peak current (with 9 ps duration pulses) could be produced with emittance in the $1 - 10\pi$ mm mrad range.

A gun test facility at SLAC was then implemented with a 3 m S-band linac section [456] and the design was improved. Four copies of the gun were fabricated.

The CANDELA photoinjector was also developed at Orsay [457].

The development of photoinjectors continued, and became crucial for single-pass FELs, because it permitted to provide electron beams with higher performance.

7.1.3 Considerations for short wavelength single-pass FEL in the SASE regime

Because of the limited performance of mirrors in terms of reflectivity, short wavelength FEL are usually operated in the so-called SASE set-up, where the spontaneous emission at the input of the FEL amplifier is amplified, typically up to saturation in a single pass after a regime of exponential growth. In the beginning of the twentieth century, several authors started to design X-ray FELs in the SASE regime [458]. A workshop on prospects for a 1 Å FEL in Sag Harbor in 1990 [459] aiming at answering the questions: "*What are the prospects for a* 1 Å *Free-Electron Laser? Can we obtain electron sources bright enough to get down to the* 1 Å *region ?*" "To focus the workshop, the initial discussion by R. Palmer defined three canonical 1 Å FEL cases as possible alternatives, i.e. with 1.6, 5, and 28 GeV electron beam sources. Each is a loose optimization of conflicting requirements needed to achieve $\lambda = 1$ Å on the electron beam quality, brightness, peak current focusing properties and its incidence on the wiggler period and total length."

L. H. Yu [460] introduced the problematics as such: "The Free Electron Laser (FEL) holds great promise as a tuneable source of coherent radiation. At the present, the shortest wavelength achieved by a FEL is 2500 Å. However, as recent progress in the development of laser driven photocathode electron guns has provided electron beams with lower and lower emittance and higher and higher current, it has become clear that FEL's with much shorter wavelength can be achieved. A FEL operating below 1000 Å will yield important advances in fields such as photochemistry, atomic and molecular physics. A FEL with wavelength of 30 Å will bring new era to the development of holography of living cells. And, if a FEL with 1 A wavelength can be developed, its impact on solid physics, molecular biology, and many other fields can hardly be exaggerated. Is it possible for a FEL to achieve 1 Å? What are the difficulties and the challenges to the present technology to build a 1 Å FEL? What are the requirements on electron beam quality and the wiggler magnets required to build a 1 A FEL? To lase at 1 Å, the FEL must operate in the high gain regime. For oscillator configuration, aside from the difficulties associated with the requirements on the mirrors which must stand high intensity 1 A radiation, we need high gain to overcome the loss in the cavity mirrors. The difficulties with the mirrors make the single pass FEL a more likely solution. For single pass configuration we also need high gain to minimize the total length of the wiggler. To achieve high gain for 1 Å FEL, the electron beam must have high peak current, low normalized emittance, and small energy spread. Strong focusing of the electron beam becomes necessary for such a short wavelength. In order to achieve short gain length, the wiggler should have high magnetic field on axis and short wiggler period. The requirements for a 1 Å FEL should be determined by the gain calculation for these various system parameters. It is usually carried out by numerical simulations. However, to explore the large parameter space for a possible FEL configuration, an analytical tool to calculate the gain would be much more convenient than the simulation". C. Pellegrini [461] concluded with the following words "The FEL in the SASE regime offers an attractive route to an X-ray laser. To make this laser a reality it is necessary to solve many problems; produce electron beams with very high quality and refine the understanding of the physics of FELs. We also need to produce long, short-period undulators with good field quality. To reach these goals we need an extensive experimental and theoretical effort on electron guns, accelerators and FEL with a number of intermediate steps that will take us from the present region of 240 nm and 1 W to 0.1–1 nm and 1 GW". J. C. Golstein [462] examined more particularly the undulator errors and concluded more generally as "All of the separate, requirements on the electron beam and the wiggler for this sort of one- Angstrom SASE FEL amplifier seem to substantially exceed achievements in existing devices. To achieve all of these requirements simultaneously, as is required for this device, would appear to require many years of development". K. J. Kim examined emittance and current density achievable in RF photo-cathode guns, and investigated the effect of space charge and RF curvature induced emittance growth.

The work in such a direction was continued during discussions held during fourth generation light source workshops [463, 464].

7.2 Historical observations of high gain single-pass SASE FEL

7.2.1 SASE observations at long wavelength

The activity continued on storage rings [210, 211] especially for the quest towards a short wavelength of operation, since high quality electron beams were still being produced. In addition, the interplay between the beam dynamics in the ring and the FEL interaction was of great interest.

Following theoretical development on high gain FEL, various experiments were carried out.

LIVERMORE NATIONAL LABORATORY, USA

Saturated high gain amplification has first been observed in the mm waves (34.6 GHz) in the mid eighties in a collaboration between Lawrence Livermore National Laboratory and Lawrence Berkeley Laboratory (USA) [294]. The electron beam from the Electron Laser Facility (ELF) (Lawrence Livermore National Laboratory) provided 6 kA, -3.3–3.8 MeV beam with a normalized emittance of 1500π mm mrad, which goes through a slit, bringing to beam to a current of approximately 500 A with 15 ns pulses with a normalized edge emittance of 470π mm mrad. The 3 m long wiggler of 98 mm period was composed of specifically shaped solenoids with independent power supplies providing a peak field of 0.5 T surrounded by a stainless-steel waveguide. The experiments were first carried out in the amplifier configuration where saturation was observed, before moving to the SASE one, for which saturation was also achieved after 2 or 3 m of undulator, depending on the experimental conditions. The power growth is shown in Fig. 61.

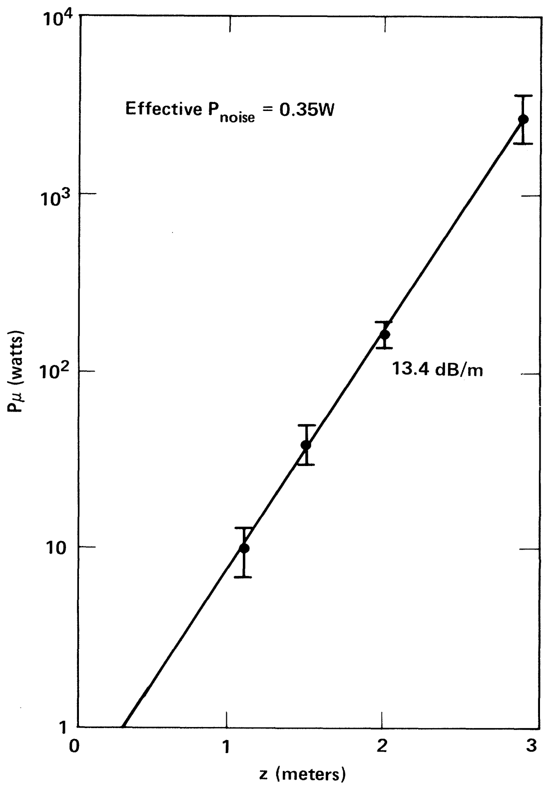


Fig. 61: Power growth in the Livermore–Berkeley experiment, figure taken from [294]

The extraction efficiency first reached 5% and then 34% by undulator tapering, leading to an output signal of 1 GW [465]. This was also an experimental demonstration of the undulator tapering for improving efficiency and it provided a very important result for the community. However, the radiation being propagated in a waveguide, it did not provide a full test of the diffraction effects that can affect the FEL, especially at shorter wavelengths, when the radiation is propagating in vacuum.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT), USA

A superradiant emission (18 MW) at 640 μm with a 4% relative bandwidth has been observed with a 2 MeV 1 kA electron beam on the PULSERAD accelerator with a helical undulator (31.4 mm period) [466]. It would correspond to an efficiency of 7%.

CENTRE D' ETUDES SCIENTIFIQUES ET TECHNIQUES D'AQUITAINE (CESTA), BORDEAUX, FRANCE

Bunching has been demonstrated at CESTA (France) at 8 mm (35 GHz) with the LELIA induction linac [295] delivering a 1 kA electron beam at 2.2 MeV, leading after transport to 800 A [467], as illustrated in Fig. 62. The 12 cm period 3.12 m long helical undulator was fed by a capacitor discharge providing a peak field of 1.1 T [468]. Cherenkov radiation was produced and measured with a picosecond streak camera. A 40 MW SASE has also been observed [469].

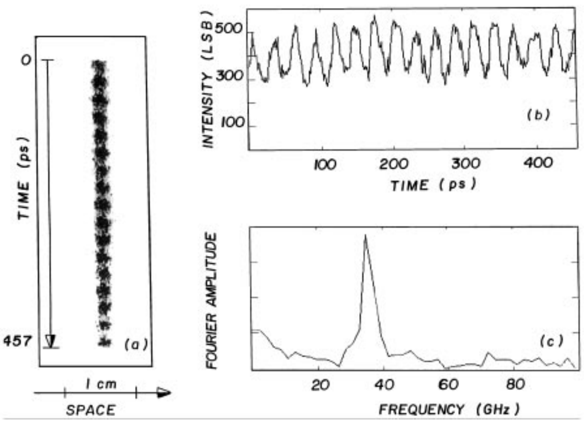


Fig. 62: Bunching observed on Cherenkov radiation observed with a streak camera (ARP) from the CESTA SASE FEL experiment from [467]. Radiation observed on a narrow rectangular slit 10 mm wide and 0.3 mm high. The slit was then displaced in time to provide a photographic record of the light intensity. Sweep speed of 25 ps/mm at a position 27.5 cm after the wiggler exit: (a) streak camera recording; (b) digitized intensity of (a) plotted vs time, and (c) frequency spectrum of (b).

7.2.2 SASE observation in the near infrared, visible and UV

The progress of the SASE observations is discussed.

INSTITUTE OF SCIENTIFIC AND INDUSTRIAL RESEARCH (ISIR), OSAKA UNIVERSITY, OSAKA, JAPAN

An increase of undulator radiation intensity by 5–100 times has been observed using the 38 MeV electron beam from a L-band linac (28 nC charge, 30 ps pulse length, $0.7 - -2.5\%$ energy spread), a 2 m long undulator (60 mm period) in the 20–40 μm spectral range, as shown in in Fig. 63 [470].

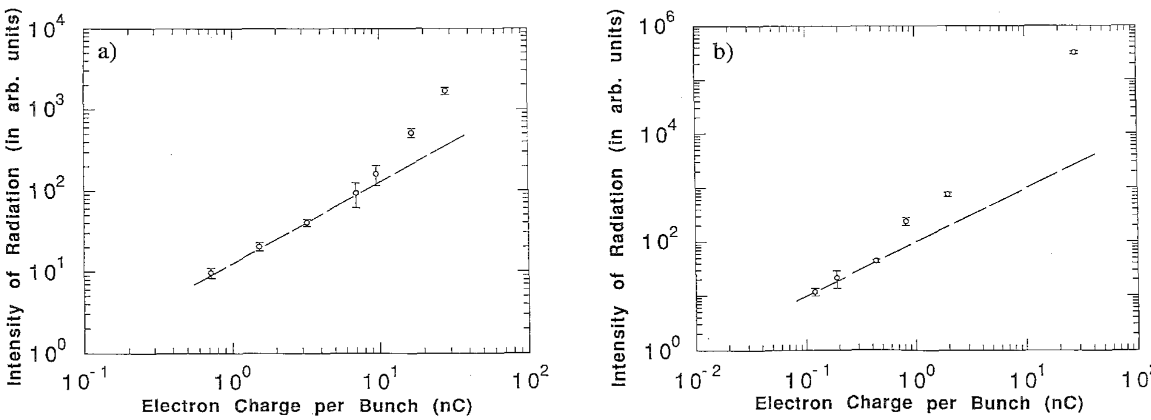


Fig. 63: Undulator radiation signal versus charge, solid line: spontaneous emission, from [470]

SUNSHINE, STANFORD UNIVERSITY, STANFORD, JAPAN

Coherent, far-infrared undulator radiation from sub-picosecond electron pulses consistent with SASE predictions with a gain length of 45.4 cm have been observed using electrons from the 16 MeV SUNSHINE S-band linac (350 ps) travelling in a 26 $\times$77 mm period permanent magnet undulator ($K =$ 0.3–0.2) [471].

CLIO, ORSAY, FRANCE

SASE at start-up in the mid infrared (5–10 μm) has been observed on CLIO (France) [472] at the Laboratoire d' Utilisation du Rayonnement Electromagnétique as shown in Fig. 64.

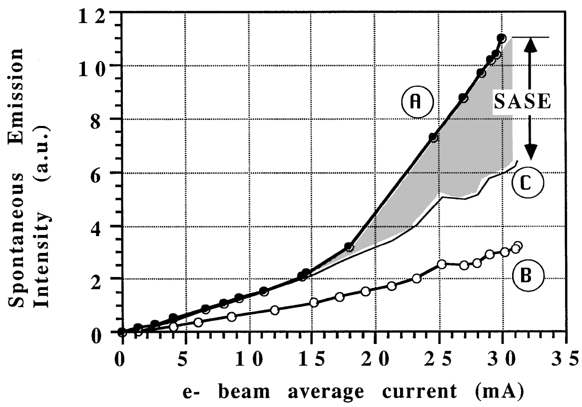


Fig. 64: Undulator radiation signal for (A) two undulator segments, i.e. 38 periods, (B) one undulator segment, i.e. 19 periods, (C) twice the intensity for one undulator segment, from [472].

A 50 MeV 3 GHz linear accelerator was providing a peak current of 100 A with an emittance of 150π mm mrad to two planar undulators of 19 $\times$ 50.4 mm period. Up to 500% gain has been measured, with a Pierce parameter of 1.9×10^{-3}. The growth in (A) is clearly non-linear, as an evidence of the SASE regime, and it differs from (B) corresponding to coherent synchrotron radiation for an equivalent number of undulator periods. SASE spectra were compared to spontaneous emission ones and present "a noisy intensity from bunch to bunch, with about 100% fluctuations", corresponding to the "spiky regime of the SASE which is, intrinsically, not a stable process" [472].

BNL, LONG ISLAND, USA

SASE was then achieved at 1064 and 633 nm, using a 61 8.8 mm period pulsed electromagnet Massachusetts Institute of Technology (MIT) micro-undulator with a peak field of 0.45 T [473]. The electron beam is produced at the Accelerator Test Facility at BNL using a photocathode RF gun: a single micropulse at 34 MeV with a variable charge of 01 nC and less than 5 ps full width at half maximum bunch length is used. Undulator radiation at 1064 nm is amplified from 2 to 6 times with respect to the spontaneous emission. SASE gain at a wavelength of 633 nm at a beam energy of 48 MeV was also observed, as illustrated in Fig. 65.

Then, SASE high gain and intensity fluctuations have been measured at 16 μm using a photocathode RF gun, a half-cell linear SATURNUS accelerator at Brookhaven National Laboratory with an emittance of 8–10π mm mrad, an energy spread of 0.08–0.14%, a transport line and an undulator from the Kurchatov Institute (40 $\times$ 15 mm period, 0.75 T field) leading to a Pierce parameter of $1 \ \times 10^{-2}$. One should note at this point the significant reduction of the electron beam emittance thanks to the RF photoinjector. First statistical analysis of SASE radiation was performed [474].

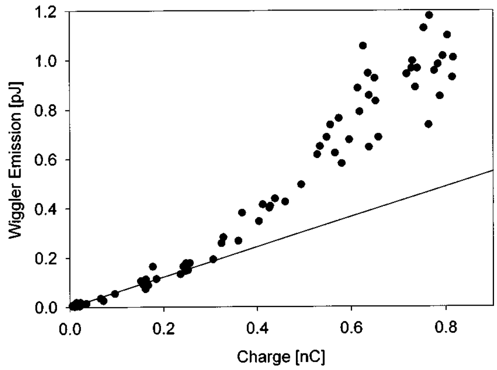


Fig. 65: Undulator radiation signal versus charge, solid line: spontaneous emission, from [473]

LOS ALAMOS NATIONAL LABORATORY, LOS ALAMOS, USA

The Los Alamos high brightness photoinjector integrated into a L-band linac 1.3 GHz at 17 MeV coupled to a 20 mm period 2 m long undulator generating a 0.7 T magnetic field and tapering [475] enabled high gain SASE 15 μm [476]. The pionnering work was recognised by the FEL prize award to Bruce Carlsten, Dinh Nguyen and Richard Sheffield for application of RF photo-injector to first high gain SASE FEL in 2017, as one of the keys for the success of present X-ray FELs. Then five orders of magnitude of amplification and saturation in the mid-infrared have been achieved in the frame of a UCLA, L. Alamos, Stanford, Kurchatov collaboration [474]. The experiment has been performed on the Advanced Free Electron Laser (AFEL) linac at the Los Alamos National Laboratory with a $CsTe_2$ photocathode at the Los Alamos National Laboratory (18 MeV, 0.25%) with the Kurchatov undulator. It led to more than five orders of magnitude amplification at 12 μm [477], as shown in Fig. 66.

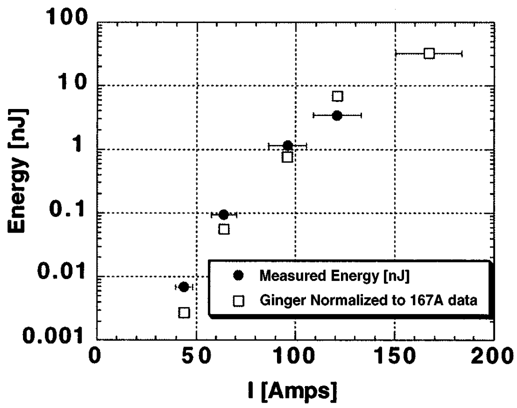


Fig. 66: Undulator radiation signal versus current and comparison with theory, from [477]

This result showed that the high gain single-pass FEL potentialities at intermediate wavelengths opened new perspectives. The electron beam micro-bunching at the exit of the SASE FEL was measured by observing coherent transition radiation, presenting a narrowband frequency spectrum [478].

LEULT, ARGONNE, USA

SASE saturation was then achieved in 2000 at Argonne National Laboratory in the visible at 530 nm, 385 nm [479,480]. The electron bunch from the Low-Energy Undulator Test Line (LEUTL) is initially accelerated to 5 MeV (8π mm mrad emittance), then injected into the linear accelerator, and further accelerated to the desired energy (up to a maximum of 650 MeV), and compressed to increase the peak current. Nine undulators 2.4 m long undulator segments (72 $\times$ 33 mm period, 1 T field, deflection parameter of 3.1) are used. They are separated by 38 cm in order to insure a proper phase matching and accommodate diagnostics, a quadrupole magnet, and steerers. An effective gain length of about 1.5 m was first measured [479] while saturation was then observed [480] as shown in Fig. 67.

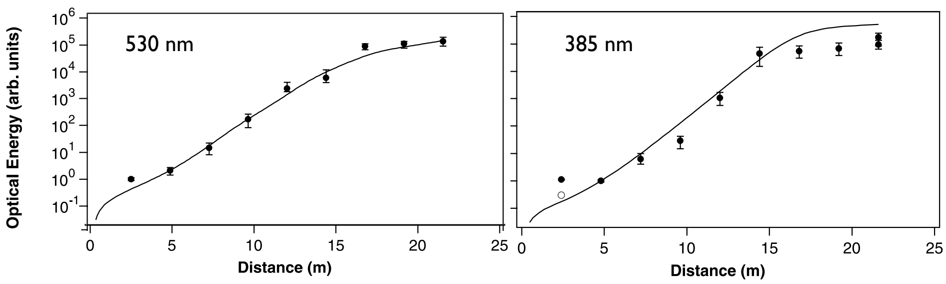


Fig. 67: SASE versus undulator length at 530 and 385 nm, from [480]

VISA, BNL-SLAC-LLNL-UCLA, USA

Saturation has been observed on the VISA (Visible to Infrared SASE Amplifier) experiment on the Accelerator Test Facility (ATF) at Brookhaven National Laboratory [481,482]. The electron beam from an S-band, 1.6 cell photocathode RF gun is accelerated to 72 MeV (200 A peak current, 2π mm mrad emittance, 0.17% energy spread) and sent into the 4 m long VISA planar undulator (18 mm period, 6 mm gap) [481]. Non-linear harmonic radiation of 845 nm at 423 nm and 281 nm was observed using the VISA SASE FEL at saturation [481], as shown in Fig. 68. The measured non-linear harmonic gain lengths decreased with harmonic number, as expected. Both the second and third non-linear harmonics energies are about 1% of the fundamental energy. This result was the first observation of non-linear harmonic SASE FEL radiation which demonstrated its potentialities to produce coherent, femtosecond X-rays.

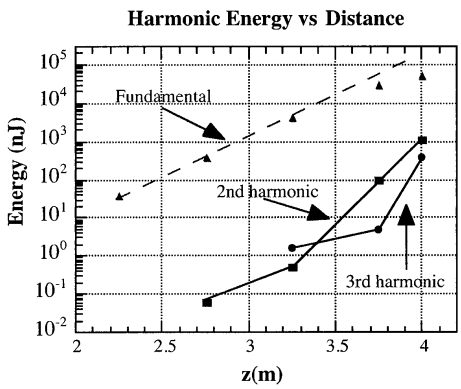


Fig. 68: Energy growth on the first, second, and third harmonics on the VISA experiment, from [481]

7.2.3 SASE observation in the VUV-soft X-ray

In the same years, a major step was achieved with the observation of SASE radiation in the VUV spectral range. The test facilities were then used as a source for scientific applications. Indeed, the pioneering users had also to learn about these new sources with their high peak power, spiky structure, pulse to pulse jitter, and short pulses.

FLASH, HAMBURG, GERMANY

B. Wiik, director of DESY (Deutsches Elektronen-Synchrotron), after a sabbatical at SLAC in 1992, considered, with G. Materlik, the possibility to build a short wavelength FEL using the TESLA accelerator, developed for the future collider. Indeed, the electron beam requirements were similar. A first Review Committee was set in 1995 and gave a positive feedback on the FEL proposal. TTF (TESLA Test Facility) was built in 1997 to test the superconducting technology for the planned linear collider TESLA, which has been replaced by the International Linear Collider. J. Schnieder reported about this time as: “Based on the good experience with superconducting technology at the large hadron lepton collider HERA at DESY and the need for high luminosity at the linear collider, the challenge was accepted to realize the accelerator in superconducting RF technology in a large international effort. The so-called TESLA collaboration was founded, which by the end of 2002 included 50 institutions from 12 countries. The ambitious goal was to increase the gradient of the accelerator by a factor of 5 and to reduce the cost of the cryomodules by a factor of 5, which has now been achieved” [483]. The FEL on TESTA-TTF [484, 485] was seen as a test-bench for the technology and physics for a future European XFEL project. The first design considered a 1 GeV electron beam with an emittance of 2π mm mrad, a peak current of 2 kA, a relative energy spread of 0.1%, a 25 m effective undulator length (27.3 mm period, 12 mm gap, 0.5 T field) for generating a FEL at 6.4 nm with a 3 GW saturation power, up to 7200 pulses can be present in the 800 μm long macropulse, at 10 Hz repetition rate thanks to the choice of a superconducting linear accelerator. Later, the TTF-FEL was renamed FLASH (free electron laser in Hamburg).

A UV (7 ps pulses) laser-driven 1.5 cell RF gun at 1.3 GHz with a Cs_2Te cathode delivers electrons which are sent in superconducting TESLA-type accelerator modules for reaching a 1 GeV electron beam, with a compression chicane in between before reaching fixed gap undulator segments (0.46 T magnetic field). First experiments were carried out with 233 MeV electron beam (emittance of 6π mm mrad, a peak current of 0.4 kA, a relative energy spread of 0.13%), enabling a gain of 3000 at 109 nm and studies of statistical properties, as shown in Fig. 69 [486], in 2000 and then saturation [487] in 2001, i.e. 25 years after the FEL invention. Tuneability in 80–120 nm range was demonstrated, and a very high degree of photon beam transverse coherence was observed. This result competed the shortest wavelength achieved on a FEL oscillator (on a storage ring).

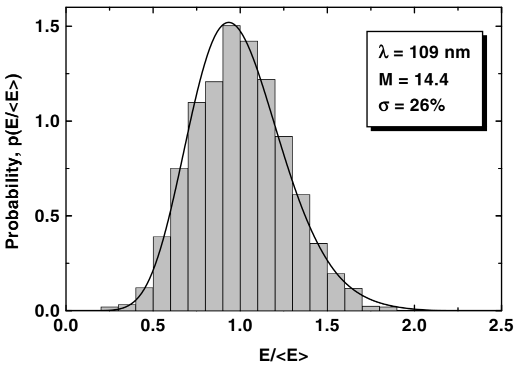


Fig. 69: SASE probability distribution, from [486]

With higher peak current, the GW level (close to 1 μJ energy) had been reached in the 95–105 nm spectral range [488] with a gain length of 67 cm, leading to a cooperation length of 5 μm.

These results constitute a major step in the SASE history. First user experiments were started quickly afterwards [489] and it became a user facility since summer 2005. Then, with some improvements on the accelerator, the 32 nm wavelength was reached with GW level power, ultra-short pulses (25 fs FWHM, and a high degree of transverse and longitudinal coherence [490] in 2006, at 13.7 nm with up to 170 mJ/pulse, 10 fs pulse duration, leading to peak powers of 10 GW [491]. With 700 pulses per second, the average power reached 20 mW. Harmonics were also quite intense with one or two orders of magnitude of power reduction (0.6% for the third (4.6 nm) and 0.03% and the fifth (2.75 nm) harmonics) in the water window of interest for biological samples. With an upgrade of the linac enabling us to reach 1 GeV, the spectral range was extended down to 6.5 nm [492]. Then a third harmonic RF cavity for phase space linearization coupled to an energy increase up to 1.25 GeV led to FEL operation down to 4.1 nm [493], i.e. in the water window on the fundamental wavelength.

Thanks to the high repetition rate of the superconducting linac, two FEL branches can be operated simultaneously, as shown in Fig. 70, with the development of FLASH-II branch with variable gap undulators [494].

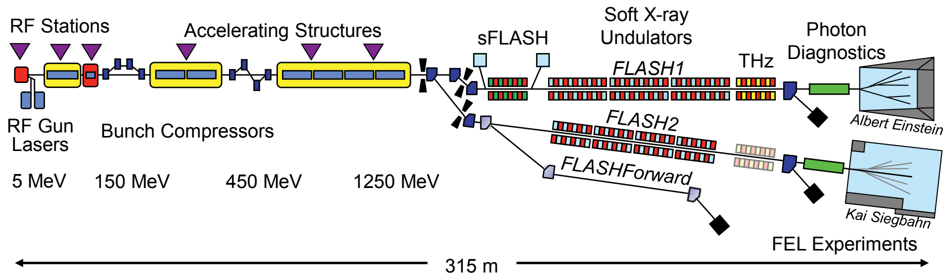


Fig. 70: Present FLASH layout from https://flash.desy.de/

While electron beam requirements from colliders and FEL met and considering the growing demand of synchrotron radiation, DESY had developed further light sources with respect to high energy physics after the shutdown of the accelerator HERA in 2007, with third generation light sources (PETRA III a very low emittance ring, the synchrotron research lab HASYLAB), and FEL-based fourth generation ones (FLASH, European XFEL). The light sources became the most important facilities of DESY. E. Saldin (19th FEL Prize in 2006), M. Yurkov (26th FEL Prize in 2015), E. Schneidmiller (26th FEL Prize in 2015) for the theoretical solid basis of the project and J. Rossbach (19th FEL Prize in 2006) for project lead brought a significant contribution to the success of FLASH. FLASH results made also confident the scientific community about the development of even shorter wavelength FELs.

SCSS TEST ACCELERATOR, HARIMA, JAPAN

The idea to combine the high level expertise on high density electron beams generated by linear accelerators in C band technology associated with a specific thermoionic gun developed following the work of T. Shintake (22nd FEL Prize in 2011), on in-vacuum undulators [495, 496] by the H. Kitamura group and on use of X-ray synchrotron radiation at SPring-8 by T. Ishikawa *et al.* led to a draft of a compact and low cost XFEL development concept in April 2000 [497]. R & D was launched on specific hardware and led to the completion of an in-vacuum undulator with a shorter period and higher magnetic field in 2002, an electron gun and the achievement of a very low emittance in 2003 [498]. After a first RIKEN symposium held in July 2003, a R&D group for XFEL was established in 2004. The SPring-8 SASE Source XFEL project was included in January 2005 as an important research objective for future R & D in the Ministry of Education, Culture, Sports, Science and Technology (MEXT) policy report on promoting science and technology of light and photons. An international Review Committee (M. E. Couprie, J. Galayda, J. Hastings, S. I. Kurokawa, W. Namkung, J. Schneider, chaired by K. J. Kim) underlined the specificities of the project: “The SPring-8 Compact SASE Source (SCSS) is an innovative project for generation and use of intense, coherent, short-pulse X-ray beams. Although its goals are similar to other X-ray free-electron laser (FEL) projects in the USA and Europe, the SCSS is unique in its compactness of design and in its co-location with the SPring-8, the world’s leading third-generation X-ray synchrotron radiation facility” and underlined the more critical components or aspects of the project. Some choices were quite original, such as the use of thermo-ionic gun operating at 1450^o C with a graphite heater instead of a photoinjector (see Fig. 71), the C-band (5.7 GHz) accelerator technology enabling a 35 MeV/m with high precision high purity oxygen-free copper structures, short period in-vacuum undulators.

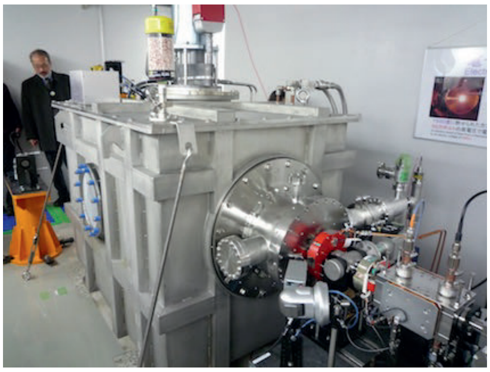


Fig. 71: Thermoionic gun of SCSS Test Accelerator, with T. Shintake

In 2005, the construction of SCSS test accelerator at 250 MeV was launched, recognized by MEXT as "critical technology of national importance" and a XFEL user group was established and an electron beam was successfully transported at the end of the year. The key components such as the CeB_6 electron gun [499] with an ultra-low emittance 0.6π mm mrad, the C-band structures and power sources, leading to an electron beam of 0.3 nC charge (0.3 kA), the in-vacuum undulators (600 $\times$15 mm period, maximum deflection parameter of 1.5 with a 3 mm minimum gap), were successfully developed. In 2006, 49 nm FEL radiation was obtained and then extended in the 60–40 nm spectral range with an energy of 30 mJ [500, 501].

7.2.4 SASE observation in the X-ray

Reaching the SASE regime experimentally took one decade after the achievement of the VUV storage ring FEL oscillator and typically 25 years after the SASE in the far infrared measured on the Livermore experiment. The FEL in the X-ray in the 0.1 nm spectral region was indeed obtained at Stanford in 2009, i.e. at the same place where the first FEL was successfully demonstrated in 1977, i.e. 32 years later. Several X-ray FEL facilities followed.

7.2.4.1 The first observation of SASE radiation in the X-ray domain at Stanford

LCLS, STANFORD, USA

The first considerations to use the SLAC linac combined to a low emittance photoinjector goes back to the fourth generation workshop held in 1992 [463], the consecutive FEL studies [502], and progress on high peak current low emittance linear accelerators [503]. A study group was formed in 1992 by H. Winick at SLAC to study FEL design, performance, and optimization, accelerator (gun, acceleration up to 70 GeV) and associated beam transport, undulators, lay-out, and scientific applications. The targeted spectral range was the water window. It lead to the study [502] for a 2–4 nm wavelength FEL using a 10 MeV, S-band photoinjector; part of the SLAC linac to accelerate the beam up to 10 GeV (at 7 GeV with an emittance of 3π mm mrad, peak current of 2.5 kA, an energy spread of 0.04% (uncorrelated), 0.1% (correlated); two longitudinal bunch compressors to increase the peak current and reduce the bunch length to 0.16 ps; and an undulator (83 mm period, 0.78 T field) enabling a 4 nm FEL with 11 mJ energy per pulse and a 60 m saturation length. The considered photoinjector was consisting of a 3.5 cell π-mode standing wave 100 MV/m accelerating structure, with a metal photocathode illuminated by a 2 ps laser pulse, providing a 3π mm mrad emittance, 250AA peak current, 1.6 ps RMS pulse duration, 0.15% relative energy spread. "In April 1992, it was considered to submit a proposal for a 2 to 4 nm

FEL to the US Department of Energy for construction starting in 1995, and development work to be done between 1992 and 1995. The name LCLS, introduced by Winick, appears for the first time in a memorandum dated June 13, 1992" [3]. Applications were discussed in the "Scientific applications of short wavelengths coherent light sources" workshop (chairs: W. Spicer, J. Arthur, H. Winick, Oct. 1992) and concerns were expressed about the sample damage induced by the high intensity of such a FEL. The working group studies were presented to a review committee (J. Bisognano, L. Elias, J. Goldstein, B. Newnam, K. Robinson, A. Sessler, R. Sheffield, chair: I. Ben Zvi) that concluded that "there is no physical principle saying that the device would not work" and R&D was recommended (electron density via electron source emittance and longitudinal pulse compression, beam alignment in the undulators at 20 μm). In 1994, the Department of Energy, upon the request for funding, asked for a review by the National Research Council, which ended up with the recommendation from the 'FEL and other Advanced Coherent Light' Committee [209] to continue the research and development towards an X-ray FEL in order to improve the technology and thus to reduce the cost. A second workshop on 'Scientific Applications of Coherent X-rays' in 1994 (J. Arthur, G. Materlik, H. Winick) [504] pointed out the advantage to use the SLAC linac and the existing building to limit the cost of X-ray FEL, and the required R&D to reduce the wavelength from 4 nm to 0.15 nm. It envisioned the new paths that could be opened by a X-ray FEL: "Such an x-ray laser should in fact lead to the same sort of revolutionary developments in x-ray studies of matter that was produced in optical studies by the introduction of the visible/UV laser". The feasibility of accelerating and compressing electron beams for reaching high peak currents (several 100 A) while keeping emittance constant was assessed [503]. A Conceptual Design Report for a XFEL in the 0.15–1.5 nm range was completed [505], it was reviewed in 1997. The Basic Energy Science of US Department of Energy [506] recognized as top priorities funding for LCLS R& D in the frame of a national effort and the importance of the first SASE experimental results.

The LCLS design [398] considered the use of the existing SLAC accelerating sections, combined to a photoinjector and undulators. The photoinjector was developed [507, 508] relying on preliminary results obtained in the frame of the BNL/UCLA/SLAC collaboration. Funds were given following the recommendations from a panel chaired by S. Leone, and the work has been distributed between different laboratories: SLAC, UCLA, Livermore Nat. Lab., Argonne, Brookhaven, Los Alamos. A formal project management structure has been established, with a Scientific Advisory Committee (co-chaired by J. Stohr and G. Shenoy). A new Conceptual design report was issued in 2002 [509], and additional funds were provided. A view of LCLS is shown in Fig. 72.

Fig. 72: View of LCLS, from portal.slac.stanford.edu/sites/lcls_public/Pages/Default.aspx

LCLS consists in a photoinjector, derived from the one of the BNL/SLAC/UCLA collaboration (0.4π mm mrad emittance for 0.4 nC charge, or 0.15π mm mrad emittance for 0.02 nC, the SLAC accelerating sections leading to an electron beam of 13.6 GeV, 250 pC (respectively 20 pC) for 0.5π mm mrad (respectively 0.14π mm mrad) normalized emittance, a relative energy spread of 0.01% and a 3 kA peak current, thanks to two bunch compressors. A view is shown in Fig. 73. A laser heater is also implemented to cure from microbunching instability [510].

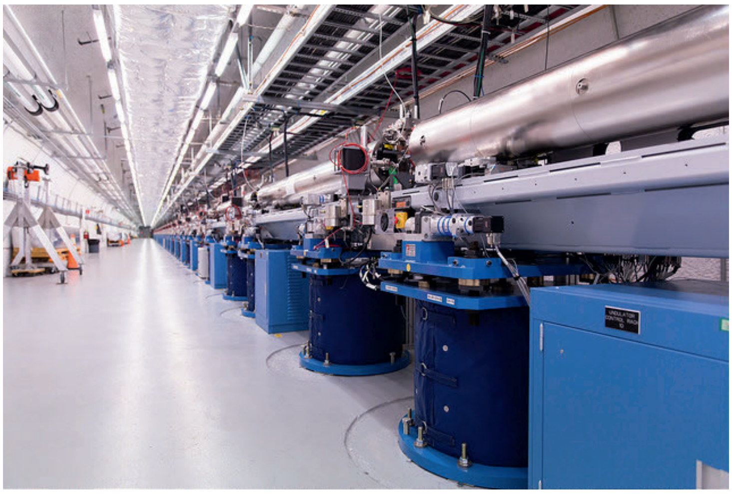


Fig. 73: View of LCLS, from portal.slac.stanford.edu/sites/lcls_public/Pages/Default.aspx

The fixed gap undulator 3.4 m long segments (30 mm period, 130 m total length) were built by Argonne National Lab [511]. Some canting of the magnetic poles (5.5 mrad) was introduced so that the resonant wavelength can be adjusted by a horizontal translation of the undulator segment. Beam position monitor and focusing was installed between the undulator modules.

After the commissioning of the injector led by David Dowell (twentieth FEL Prize in 2009), SASE radiation was achieved in April 2009 at 0.15 nm, very rapidly after sending the electron beam in the undulators. Saturation was obtained without using the total number of undulator segments [512]. LCLS was also adjusted with a lows charge and shorter electron bunches [513]. The commissioning of LCLS was led by Paul Emma (twentieth FEL Prize in 2009), under the project management of John Galayda (23rd FEL Prize in 2012). Beam based alignment was required [514]. Microbunching instability and effect of the laser heater were also studied [515]. LCLS now operates in the 0.25–9.5 keV spectral range, with a 120 Hz repetition rate, with several mJ and pulses as short as 5 fs.

The success of LCLS, the first tuneable X-ray FEL, was a major advance in FEL history. It opened the way to explore new areas of matter investigation with such a high energy per pulse. There are now six experimental stations.

7.2.4.2 The second observation of SASE radiation in the X-ray domain in Japan

SACLA, HARIMA, JAPAN

Following the success of the test facility SCSS Test Accelerator in 2006, a review working group concluded that “the XFEL plan should be actively carried forward and the project should be started at an early date”, and 2.3 billion yen was allocated for construction and research on use of the XFEL facility by the government at the end of the year. XFEL Project Head Office was established in 2006, and the project was jointly promoted by RIKEN and Japan Synchrotron Radiation Research Institute (JASRI). The construction of the XFEL facility in Japan began in July 2009. In 2009, the accelerator and undulator buildings were completed, the thermoionic gun operating at 1450°C with a graphite heater, the acceleration tubes (see Fig. 74) were installed, undulators started, the experimental building was built in 2011. A fifth XFEL symposium was held in 2009.

The FEL was achieved on June 7, 2011. SASE is operated at SACLA down to 0.08 nm [516]. For the final adjustments, the undulator segments have been aligned using the photon beam itself [517].

SACLA has also some specificities: it gathers XFEL and SPring-8 in such a way that radiation from both light sources can be combined on a sample, or the electron beam from the linear accelerator

Fig. 74: SACLA C-band linac

of SACLA can serve to inject the storage ring for short pulse operation as shown in Fig. 75.

Fig. 75: SACLA FEL top view from http://xfel.riken.jp/eng/. In the long building is installed SACLA with the different FEL branches, whereas in the rear of the picture, one of the third generation storage ring beamlines, a very intense laser, and SACLA photon beam can be coupled. The configuration also enables to use the linac electron beam to inject in the storage ring for very short pulse synchrotron radiation production.

It is the only X-FEL not having a photoinjector, but a thermoionic gun. It is the first X-ray FEL to have adopted the C-band accelerator technology and in-vacuum undulators (18 mm period).

SACLA now operates with two hard X-ray beamlines. In addition, SCSS Test Accelerator has been moved, the electron beam energy has been raised and undulators have been added, providing an additional soft X-ray beamline, presently open to users [518].

7.2.4.3 The next X-ray SASE FELs in the X-ray domain

PAL FEL, POHANG, KOREA

Pohang Accelerator Laboratory launched the study of a XFEL [519] in the beginning of the twentieth century. Studies were carried out and reviewed. PAL-XFEL uses a 10 GeV S-band linac (0.4π mm mrad emittance, 60 Hz), two series of undulators (for the hard X-ray line at 10 GeV: 20 segments of 26 mm period, 0.81 T peak field, 8.3 mm minimum gap; and for the soft-ray beamline at 3.15 GeV: 20 segments of 35 mm period, 1 T peak field, 9 mm minimum gap) [520] to cover the 0.1–4.5 nm spectral range. The site of the facility is shown in Fig. 76.

Fig. 76: PAL FEL site, from http://pal.postech.ac.kr/paleng/Menu.pal?method=menuView-&pageMode=paleng&top=7&sub=5&sub2=0&sub3=0

The installation of the linac and undulator was completed in January 2016. After approval from the Radiation Safety Control Agency, commissioning of the accelerator was started from a 135 MeV injector on April 14, 2016: the first beam from the RF-gun [521] was achieved, and in 11 days, the beam was accelerated up to 10 GeV and was transported 715 m away from the gun.

The electron beam was then sent at the entrance of the undulator lines, 794 m away from the gun. The beam was compressed to 3 kA peak current. The first FEL was obtained at 0.5 nm on June 14. Korea Bizwire made the following announcement on June 30, 2016: "Following the United States and Japan, Korea became the third country to successfully produce an 'x-ray free-electron laser' (XFEL), often referred to as the 'dream light'. According to the Ministry of Science, ICT and Future Planning, the POSTECH (Pohang University of Science and Technology) Pohang Accelerator Laboratory has succeeded in producing an XFEL with a wavelength of 0.5 nm. The lab started a trial run of the PAL-XFEL on April 14. The laser was first observed in the early morning of June 14, and the facility was visited on June 29 by an external verification committee to confirm the laser's successful production" [522].

Saturation was achieved at 0.144 nm on 27 November 2016, with an energy per pulse of 132 μJ for a 8 GeV electron beam and an undulator deflection parameter of 1.87, as shown in Fig. 77. The gain length is 3.43 m. Undulator tapering applied for the last eight undulators led to a further increase of the FEL intensity [523]. PAL-FEL has two hard X-ray beamlines and one soft X-ray one.

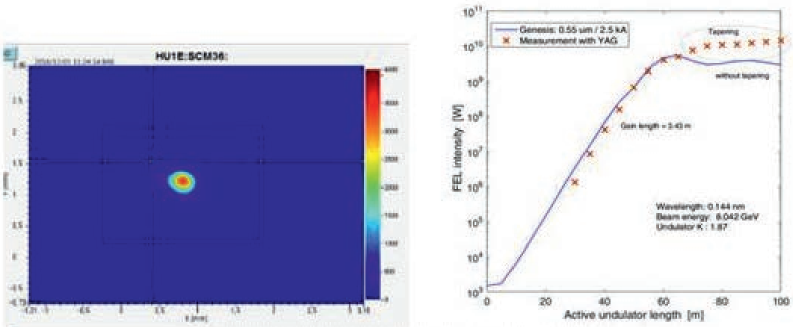


Fig. 77: X-ray FEL at PAL: X-ray spot and power saturation curve, from [523]

SWISSFEL, VILLIGEN, SWITZERLAND

SwissFEL is developed at Paul Scherrer Institute. Following the conceptual design report [525], the construction started. A special care was dedicated to the preservation of nature, of the site of the Swiss FEL facility (see lay-out displayed in Fig. 78.

SwissFEL [526] consists of a very low emittance injector, C-band accelerating sections leading to 5.8 GeV for a 100 Hz repetition rate, and 15 mm period in-vacuum undulators [527], for two different FEL beamlines: ARAMIS (0.1–0.7 nm) and ATHOS (0.7–7 nm). First, the injector was commissioned [528]. Then, the full installation was completed. A first lasing was achieved in December 2, 2016 with a 377 MeV electron beam at 24 nm, in May 2017 at 4.1 nm [524] and in October 2017 at 1.2 nm [529].

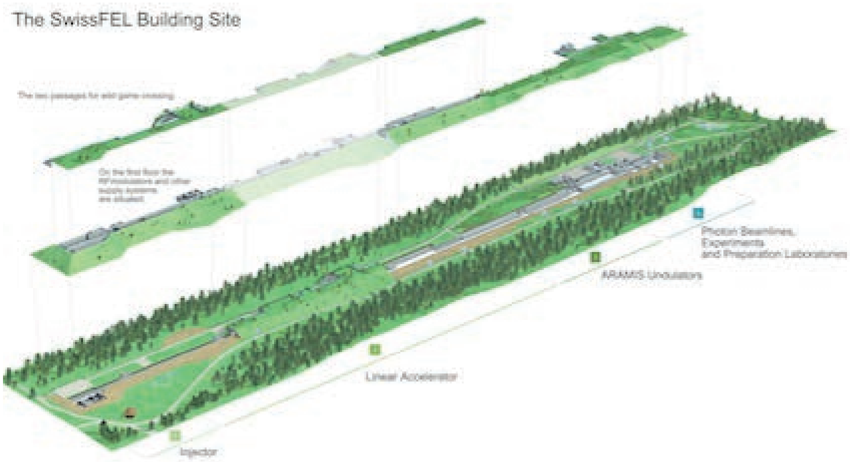


Fig. 78: SwissFEL layout, from https://www.psi.ch/swissfel/

EUROPEAN FEL, HAMBURG, GERMANY

The European XFEL has a long history. It was already discussed here with the introduction of the TESLA-TTF FEL. The German Federal Ministry of Education and Research granted permission to build the XFEL in 2007 at a cost of 850 M euros, under the provision that it should be financed as a European project. The FEL on TESTA-TTF [484, 485] was seen as a test-bench for the technology and physics of the future European XFEL project. In 2004–2007 a ‘Science and Technology Issues’ group chaired by F. Sette was created. In 2007, the European XFEL project was officially launched. The European XFEL GmbH company, has been founded in 2009 for building and operating the facility has been founded in 2009. It gathers a consortium of different countries: Denmark, France, Germany, Greece, Hungary, Italy, Poland, Russia, Slovakia, Spain, Sweden, Switzerland, bringing financial and/or in-kind contributions. The 3.4 km long X-ray free electron laser facility extends from Hamburg to the neighbouring town of Schenefeld in the German federal state of Schleswig-Holstein. Technically, European XFEL [530] uses a superconducting linac of extremely good electron beam parameters, enabling operation at high repetition rate. The electrons will be accelerated up to 17.5 GeV over 2.1 km. There are 101 accelerator modules. EFEL will provide three different SASE sources for six experimental stations, as shown in Fig. 79.

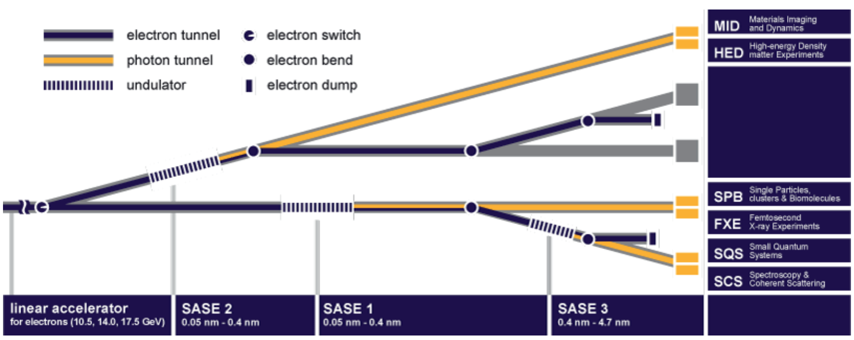


Fig. 79: Sketch of the European X FEL facility, from [531]

SASE 1 and SASE 2 cover 0.4–0.05 nm spectral range, with a 175 m undulator length, whereas SASE 3 covers 4.7–0.4 nm spectral range, with a 105 m undulator length. Pulse duration will be shorter than 100 fs. The flux reaches 10^{12} ph/s. The specificity of the EXFEL is its repetition rate of 27 000 pulses per second, leading to a peak brilliance a billion times higher than that of the best synchrotron X-ray radiation sources (5×10^{33} ph/s/mm^2/mrad$^2/0.1\%$ BW).

Civil construction of the facility started in 2009, and continued with the completion of the 3.4 km tunnel in 2012, and underground in 2013. The overall cost for the construction and commissioning of the facility is as of 2015 estimated at 1.22 B euros. First electrons have been guided from the injector into the first four 2 K superconducting accelerator modules at -271^{o}C and compression chicane in January 2017 [531] in the cooled main accelerator, as presented in Fig. 80. First lasing was achieved on May $2nd$ 2017 at a moderate energy of 6.4 GeV at 0.9 nm [532], and after beam based alignment and systematic tuning of the electron beam properties, at 0.2 nm on May $24nd$ 2017 with a energy of 1 mJ achieved three days later.The SASE 1 beam was transported to the experimental hutch in June [533].

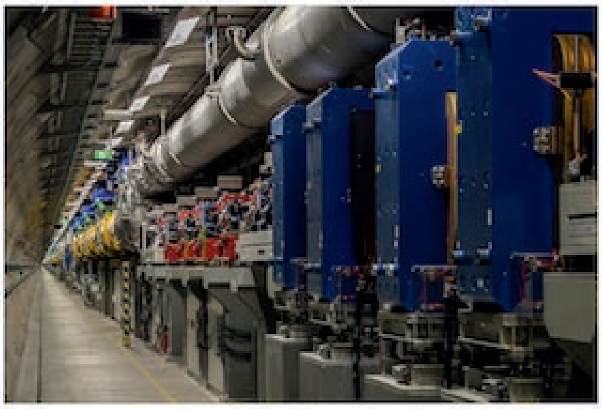


Fig. 80: View into the accelerator tunnel: electrons guided into the first four superconducting accelerator modules (yellow) and in a chicane (in front, blue and red), from [531].

European XFEL is the next world's brightest source of ultrashort X-ray pulses and will open up new research opportunities for scientists and industrial users. Thanks to its ultrashort X-ray flashes, the facility will enable scientists to map the atomic details of viruses, decipher the molecular composition of cells, take three-dimensional images of the nanoworld, film chemical reactions, and study processes such as those occurring deep inside planets.

LCLS-II, STANFORD, USA

LCLS-II [534] will move to the use of a 4 GeV superconducting accelerator technology, in the CW mode of operation. It will provide a major jump in capability, moving from 120 pulses per second to 1 million pulses per second. The electron beam properties will be of high quality: normalized slice emittance of 0.45π mm mrad, slice energy spread of 0.12×10^{-4}. The project will also incorporate variable gap hybrid undulators to cover soft (0.2–1.3 keV, i.e. 0.95–6.2 nm) and hard (1–5 keV, i.e. 0.25–1.2 nm) X-ray photons at up to MHz rates; hard X-ray above 25 keV (i.e. below 0.05 nm at 120 Hz), with performance comparable to or exceeding that of the existing LCLS. The project is conducted in the frame of a collaboration between SLAC, Fermilab, Jefferson Lab., Argonne, Cornell University.

SHANGHAI XFEL, SHANGHAI, CHINA

The construction on the Shanghai Coherent Light Facility (SCLF) for a high repetition rate X-ray Free Electron Laser at Shanghai relying on a 8 GeV superconducting accelerator technology has been approved. The super- conducting electron accelerator, undulators and photon beamlines and endstations are all installed in 3.1 km under-ground tunnels. Using 3 phase-I undulator lines, the SCLF aims at generating X-rays between 0.4 and 25 keV at rates up to 1MHz [535].

7.3 SASE properties

7.3.1 SASE longitudinal properties

The SASE emission starts from of the amplification on spontaneous emission and presents generally spikes in the temporal and spectral distributions, because of non-correlated trains of pulses, resulting in a partial longitudinal coherence. The SASE spectra observed on FLASH are shown in Fig. 81. They illustrate the SASE fluctuations: the number of spikes (wave packets) M is typically 2.5, leading, in using the value of the cooperation length, to FEL pulses of about 50 fs. It can be understood in terms of statistical properties [536].

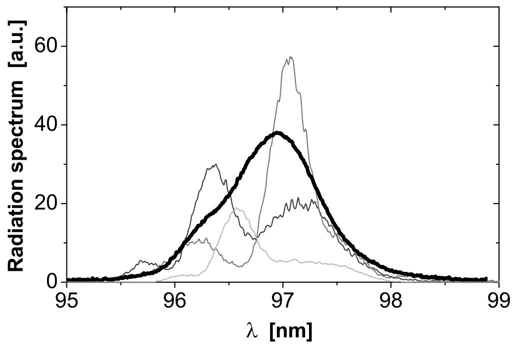


Fig. 81: SASE spectra, from [488]

There are particular cases where this spiky spectral and temporal structure of the SASE can be handled, such as seeding, as developed in the next solution. Alternatively, the FEL can be operated in a low-charge short electron bunches [537] as demonstrated in LCLS [513], or in combining an electron beam energy chirp combined with an undulator taper [538], as shown in SPARC (Test facility in Italy) [539]. Proper combinations of chicanes and undulator segments can enable to phase lock the radiation [540].

Different schemes have been proposed and/or tested for achieving extremely short pulses [541–551] with a selective amplification, modulation, phase locking of the radiation from different segments, superradiance [552]. The partial FEL coherence can be taken into account in the pulse duration measurement [553].

7.3.2 SASE transverse properties

Thanks to the rather low emittance of the electron beam and eventually to gain guiding, SASE FEL presents generally a good transverse coherence [554] and a proper wavefront [555].

7.3.3 SASE polarization

The polarization of the FEL mainly results in the choice of the undulator. Whereas the majority of the SASE FEL started with planar undulators leading to linear polarization, there is a recent trend to provide more polarization flexibility for users. For example, LCLS recently operated with a DELTA undulator [556] leading to hundreds of microjoules of circulator polarization in the 1–2.5 nm spectral range [557].

7.4 Seeding

7.4.1 External laser seeding

Following coherent harmonic generation carried out on low gain FEL, the idea of sending an external laser tuned on the undulator fundamental wavelength was developed, in the so-called 'seeding configuration' as shown in Fig. 82. The seed provides a sufficiently intense input field that generates an efficient bunching even in a short undulator, leading to coherent emission of undulator radiation and of its higher-order harmonics. The FEL may then operate as an amplifier of the initial seed, capable of increasing the peak power of a light source to approximately the same saturation power level as for the SASE case. The seed should overcome the initial shot noise.

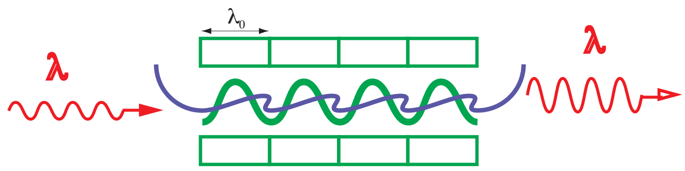


Fig. 82: FEL seeding: a coherent source tuned on the resonant wavelength of the undulator enables to perform efficiently the energy exchange leading further to the density modulation

Concerning the temporal properties, seeding could somehow enable to manipulate the FEL properties. The temporal and spectral distributions of the pulse result from the seed and the FEL intrinsic dynamics and could be modified by the interaction with an external laser. Seeding offers a good strategy for suppressing the spikes inherent to the SASE process and thus for improving the longitudinal coherence, and for reducing the intensity fluctuations, and jitter. In addition, since the electron bunch modulation is controlled by the external laser source, the saturation length can become shorter, the cost can be reduced [558]. Seeding can be used also to efficiently generate harmonics.

Experimentally, the electron beam and the seed should be synchronized, the radiation should overlap transversally and spectrally.

The progress of seeding on high gain single pass FELs is described below, starting with the use of conventional lasers first in the mid infrared, then with that of high-order harmonics generated in gas for a seed at shorter wavelength.

7.4.2 External conventional laser seeding

BNL FEL EXPERIMENT/NSLS DUV FEL, BNL, USA

The first experimental demonstration was carried out by L. H. Yu (16th FEL Prize in 2003) at Brookhaven national Laboratory [429]. The set-up was composed of a 10.6 μm with a 0.5 MW seed CO_2 laser, a 40 MeV electron beam with 120 A peak current (6 ps FWHM, with 5π mm mrad and 0.6% energy spread); a 0.76 m long first modulator (80 mm period, 0.16 T magnetic field), a 0.3 m long dispersive section, and a 2 m long radiator (33 mm period, 0.47 T magnetic field). It led to the saturated, amplified free electron laser second harmonic at 5.3 μm, as shown in Fig. 83.

The experimental results showed that the SASE output was multiplied by six orders of magnitude in the HGHG spectrum. The measured FWHM HGHG bandwidth was of 20 nm, i.e. six times smaller than the SASE one. The spectral bandwidth was significantly reduced by seeding, and longitudinal coherence improved. A single shot HGHG shown in B shows a nice spectral profile, quite different from the spiky SASE pulse.

Such a result has been a major contribution in 2000 for the FEL community, since the use of a laser-seeded free electron laser enabled to produce amplified, longitudinally coherent, Fourier transform

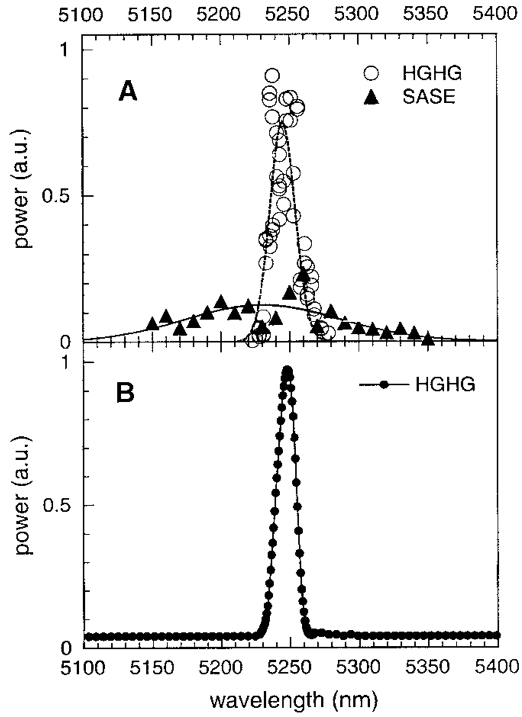


Fig. 83: High gain harmonic generation demonstration using a 800 nm laser, from [429]. A SASE point: an average of 10 shots, HGHG points: single shots normalized to the total HGHG pulse energy. B: Single shot HGHG pulse recorded with a thermal imaging camera at the exit plane of the spectrometer

limited output at the harmonic of the seed laser. "The experiment verifies the theoretical foundation for the technique and prepares the way for the application of this technique in the vacuum ultraviolet region of the spectrum, with the ultimate goal of extending the approach to provide an intense, highly coherent source of hard x-rays" [429]. "The HGHG approach offers an alternative and attractive FEL scheme that combines the benefits of the coherence properties of a laser with the short-wavelength capabilities of an accelerator based light source. A future X-ray HGHG FEL could use the best advances in short-wavelength tabletop lasers as seeds for amplifying and pushing toward shorter wavelengths" [429]. The measurements were in good agreement with the theoretical expectations.

The next step in shorting the wavelength [559]. The beam (4 ps FWHM, with 4.7π mm mrad) from the DUV FEL is produced with the BNL photoinjector where the cathode is illuminated at 266 nm, accelerated with four SLAC accelerating section bringing the energy to 177 MeV, with a chicane between the second and third accelerator module. The modulator was the same as for the previous experiment, whereas the radiation was the 10 m long NISIUS undulator with 38.9 mm period, 0.31 T peak field ($K = 1.13$) with focusing in both planes thanks to canted poles. The seed was taken from the Ti–Sa laser at 800 nm with 30 MW used for the photoinjector. The seeded FEL spectrum is shown in Fig. 84 at 266 nm, the third harmonic of the laser seed. It exhibits a nice line, of 0.1% bandwidth, as compared to the broad spiky SASE spectra. The HGHG width is close to one single SASE spike. An estimate of the pulse length of 0.9 ps was found, close to the 1 ps electron beam duration after compression. These results provided evidence of the high temporal coherence in the HGHG output and significant improvement due to the seeding, with respect to the SASE.

Figure 85 shows the output energy measured for different seed levels versus the wiggler length, by kicking the electron beam at different locations in the undulator. A 0.8 m long gain length was found. The total length of the NISIUS undulator was not sufficient for getting a saturated SASE whereas saturation can be reached in the seeded configuration. Saturation was also more rapidly reached than in the SASE

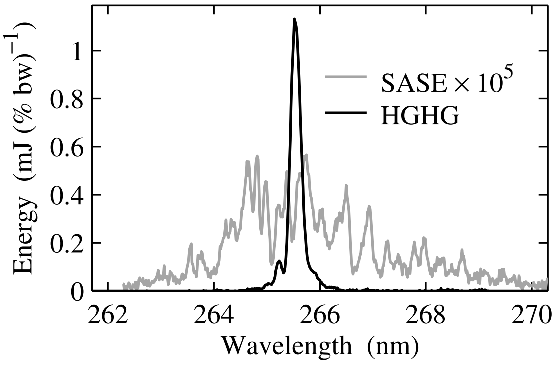


Fig. 84: High gain harmonic generation demonstration using a 800 nm laser, from [559]

case, which makes the system more compact.

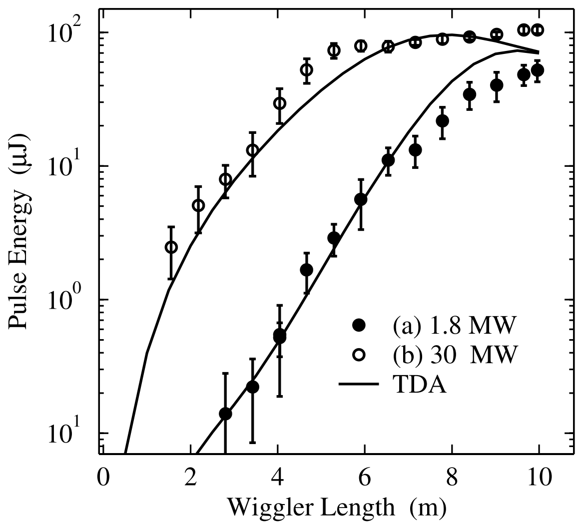


Fig. 85: High gain harmonic generation demonstration using a 800 nm laser, from [559]

Together with the fundamental radiation at 266 nm (100 μJ), significant signal was found on the second (0.1 μJ) and third (0.3 μJ) harmonics [560]. The harmonic radiation at 89 nm of the seeded FEL was successfully used for a first scientific applications in molecular physics [561]. Tuneability is achieved by applying a chirp on the electron beam [562].

A super-radiant seeded FEL was experimentally demonstrated at BNL [563].

SPARC TEST FACILITY, FRASCATI, ITALY

In Italy, a budget dedicated to FEL research has been implemented. Two proposals have been funded, the SPARC FEL test facility, and the FERMI@ELETTRA seeded FEL facility. The SPARC FEL amplifier [539, 564] is driven by a high brightness accelerator providing energies between 80 and 180 MeV and an undulator composed by six modules of variable gap. A super-radiant seeded FEL was experimentally demonstrated on SPARC up to the 11th order [565] and in the cascade configuration [566].

7.4.3 External seeding with high-order harmonics generated in gas

Conventional lasers are limited in terms of the short wavelength they can provide, even though frequency mixing schemes can be used. However, in the landscape of available light sources in the VUV and soft X-ray [567, 568], Harmonic generation in gas (HHG) is one of the most promising methods to generate radiation at short wavelengths in the vacuum and extreme ultraviolet region of the spectrum [569, 570], and is currently in operation for user applications. The high-order harmonics result from the strong non-linear polarization induced on rare-gas atoms, such as Ar, Xe, Ne, and He, by the focused intense electromagnetic field of a pump laser. As the strength of the external electromagnetic field is comparable to that of the internal static field of the atom in the interaction region close to laser focus, atoms ionize by tunnelling of the outer electrons. The ejected free electrons, far from the core, are then accelerated in the external laser field and gain kinetic energy, they can be driven back close to the core and be scattered or recombine to the ground state emitting a burst of XUV photons every half-optical cycle. Correspondingly in the spectral domain, the harmonic spectrum includes the odd harmonics of the fundamental laser frequency. The characteristic distribution of intensities is almost constant for harmonic order in the 'plateau' region, where, depending on the generating gas, the conversion efficiency varies in the range 10^{-4}–10^{-7}. For higher orders, the conversion efficiency decreases rapidly, in the 'cut-off' region, which is determined by the gas ionization and the ponderomotive energies. The lighter is the gas (i.e. the higher is the ionization energy the higher is the cut-off energy). High-order harmonics are linearly polarized sources from hundreds of nm to nm, of good temporal and spatial coherence, emitting very short pulses (fs to as), at rather high repetition rate (up to a few kilohertz). The radiation spectrum is completely tuneable in the VUV–XUV region. The harmonic radiation is emitted on the axis of the laser propagation with a small divergence (1 to 10 mrad). Fraction of a microjoule of energy can be obtained at wavelengths down to tens of nm [571]. It was then thought that HHG could suit for being considered as a seed for a high gain FEL [111] for the ARC-EN-CIEL project in France, and on the SCSS Test Accelerator [572] in the frame of a French–Japanese collaboration.

SCSS TEST ACCELERATOR, HARIMA, JAPAN

The HHG chamber has been prepared in France and sent to Japan, whereas an existing laser has been upgraded with a delay line added for such an experiment. The HHG seeding chamber, with a Xe gas cell was located inside the accelerator tunnel. A second chamber handled the transverse focus of the seed in the first undulator. An injection chamber, containing a set of steering mirrors, was located in a magnetic chicane. HHG seeding has been first performed on SCSS Test Accelerator at 160 nm [573]. The HHG seed was strongly amplified in the first undulator segment and the unseeded signal was amplified by three orders of magnitude. The saturation length was reduced by a factor of 2, making the system more compact. The fundamental wavelength was accompanied by the non-linear harmonics (NLH) at 54 nm and 32 nm [574]. Figure 86 shows that light up to the seventh harmonic of the FEL resonance can be measured while in presence of the seed. The seventh harmonic could not be detected when the FEL amplifier was operated with no seed, in SASE mode. A significant increase of the non linear harmonics signal, as compared to the unseeded case, was also observed at the third and fifth harmonics, which were amplified by factors 312 and 47 for the third (0.3 nJ at 53.55 nm) and for the fifth (12 pJ at 32.1 nm) respectively. Spectral narrowing was also observed at the harmonics (from 2.66% to 0.84% for the third harmonic and 2.54% to 1.1% for the fifth harmonic). The seed level required to overcome the shot-noise [571, 575] was studied.

The HHG layout was modified to use a SiC harmonic separator mirror, set at the Brewster angle (69^o), for the Ti–Sa pump laser. By introducing a pair of Pt-coated, nearly normal incidence mirrors, both the collimation and the focusing of the HHG radiation were achieved. HHG seed FEL was then obtained at 60 nm [576] with a seed energy of 2 nJ/pulse (i.e. 40 kW, with 50 fs pulse duration). The pulse energy of the seeded FEL (1.3 μJ), was twice larger than in SASE mode (0.7 μJ) and 650 times larger than the HHG seed level (2 nJ).

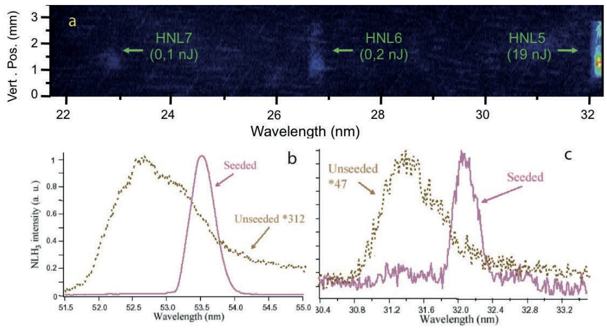


Fig. 86: Non0linear harmonics of SCSS Test Accelerator FEL seed with HHG at 60 nm (a) fifth to seventh harmonic image of the spectrometer. Comparison of SASE and seeded FEL harmonics of third order (b) and fifth one (c), from [571].

The synchronization was then improved with electro-optical sampling [577, 578], leading to a better hit rate. A few tens of microjoule could then be obtained [579] EUV-FEL.

SCSS test accelerator components have now been moved to SACLA for providing a HHG seeded FEL in the soft X-ray region down to 3 nm, with additional accelerating sections and undulator modules. It could then be combined with HGHG.

SPARC TEST FACILITY, FRASCATI, ITALY

With the flexibility of SPARC for the HGHG configuration, it appeared also to be a good candidate for testing HHG seeding. It was performed in the frame of an Italian (ENEA, INFN)–French (CEA /synchrotron SOLEIL) collaboration. The Ti–Sa laser delivering up to 2.5 mJ at 800 nm with a pulse duration of about 120 fs was focused by a 2 m focal length lens to an in-vacuum cell, where a synchronized valve injected Argon gas at 15 bar. Seeding at 160 nm was performed. Then, for a 50 nJ 266 nm seed, the resonance can be set both on the fundamental and second harmonics. The six 2.1 m long undulator segments could be independently tuned at the seed wavelength, operating as modulators, or at its second harmonic, 133 nm, as radiators of a frequency-doubling cascade. A beam of 176 MeV with a 50–70 A peak current with 0.9π mm mrad emittance was employed. Figure 87 shows the comparison between the experimental data and the results of a statistical study made with 100 random shots, simulated by GENESIS 1.3. An output energy of 1 mJ at 133 nm was obtained with four modulators and two radiators at 133 nm. The estimated gain length in the modulator of 1.1 m was sufficient to increase the input seed to a level close to saturation, and up to 4×10^{12} photons were produced at 133 nm [580].

SFLASH, HAMBURG, GERMANY

The sFLASH seed laser system producing 800 nm, 50 mJ adjustable pulse length (down to 30 fs), connected to the accelerator tunnel by a 7 m long tube, was sent to a gas filled capillary for the production of the seed at 38 nm of 2 nJ, the 21st harmonic of the 800 nm Ti–Sa laser. The seed was sent in the accelerator. The first undulator was located 5 meters after the point of injection into the tunnel. A proof of the interaction and amplification of the seed, coupled to the electron bunch, was obtained on the first and second harmonics at 38 nm and 19 nm [581]. This is the shortest wavelength where harmonics generated in gas have been amplified in a single-pass FEL.

7.4.4 Seeded FEL facilities

FERMI, TRIESTE, ITALY

FERMI is the first seeded FEL user facility VUV/soft X-ray located at Trieste in Italy. It was launched

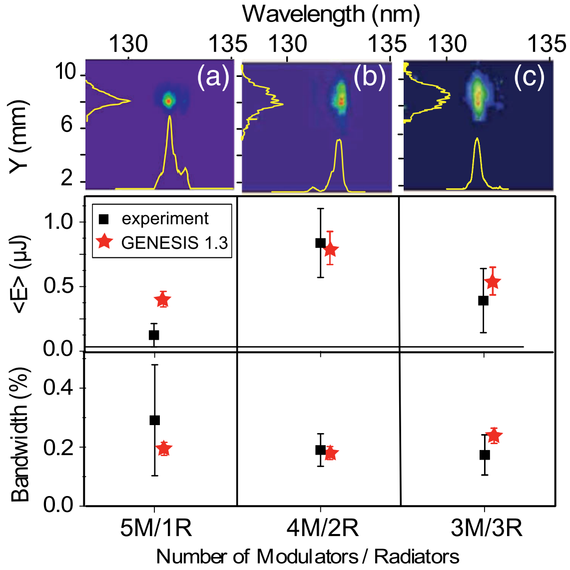


Fig. 87: HHG seeding at SPARC with 266 nm with (a) five modulators, one radiator, (b) four modulators, two radiators, (c) three modulators, three radiators. Data averaged over 100 shots with one standard deviation error bar and compared with GNESIS simulations. [580].

after the Italian initiative for FEL [582–585]. The electron bunches generated in a high-gradient photocathode gun is accelerated by a normal conducting linear accelerator up to a beam energy of 1.2 GeV (1π mm mrad emittance, 0.016% energy spread, 0.8 kA peak current) before reaching the two FEL branches in the HGHG cascade configuration, in order to provide a good longitudinal coherence. It relies on the two FEL branches, FEL 1 in the 100–20 nm via a single cascade harmonic generation, and FEL 2 in the 20–4 nm via a double cascade harmonic generation [586], as shown in Fig. 88. The seed laser is based on an optical parametric amplifier continuously tuneable in the range 230–260 nm, delivering pulses of few tens of microjoules [587, 588]. The modulators are planar undulators, and radiators are APPLE-II [589] type undulators for providing adjustable polarization. For FEL 1, the modulator is a 3.03 m long undulator of 160 mm period, providing a deflection parameter ranging between 3.9 to 4.9. The APPLE-II type radiators are 2.34 m long with 65 mm period, from a deflection parameter ranging between 2.4 and 4. For FEL 2, the modulator of the first stage has 30 $\times$ 100 mm periods, the three first stage radiators and second stage modulator have 44 $\times$ 55 mm periods in variable polarization, and the six second stage radiators have 69 $\times$ 35 mm periods in variable polarization.

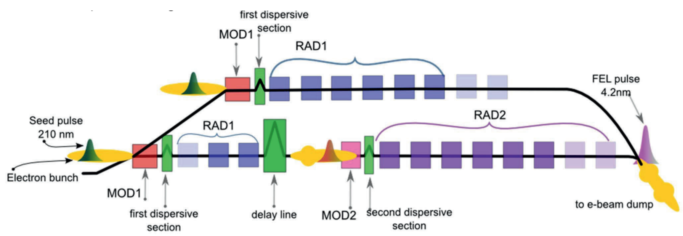


Fig. 88: FERMI HGHG FEL lines

FERMI lasing was achieved in December 2010 on FEL1 and in May 2012 on FEL2 [587,588,590]. Tuneability can be achieved on the injection source coupled to a gap change or by applying a chirp (frequency drift) both on the seed and on the electron bunch [591]. The combination of HGHG, fresh bunch technique, and harmonic cascade has recently enabled an up-frequency conversion by a factor of 192 [592]. Two-colour operation was achieved both with the pulse splitting technique [593, 594] or with a twin-pulse electron beam [595]. The polarization can be efficiently controlled thanks to the APPLE-II type undulators [596]. FERMI results constitute a major step in the community with the control of the temporal FEL distribution at short wavelength, and the flexibility in polarization.

There are different experimental stations for coherent diffraction imaging (DIPROI), absorption and elastic scattering from materials under extreme conditions (EIS-TIMEX), gas phase and cluster spectroscopy (LDM) with additional facilities for inelastic and transient grating spectroscopy (EIS-TIMER), and terahertz applications (TERAFERMI). Optical laser pulses are also available for pump-probe experiments (SLU).

DALIAN FEL, DALIAN, CHINA

The project was started in early 2012 within a collaboration between Dalian Institute of Chemical Physics (X. Yang), Shanghai Institute of Applied Source (Z. Zhao, D. Wang), from China Academy of Science. The Dalian facility covers 50–150 nm (8–24 eV) in both HGHG and SASE modes [597]. The Dalian FEL is sketched in Fig. 89.

Fig. 89: Dalian FEL sketch, from https://www.asianscientist.com/2017/01/topnews/brightest-vuv-free-electron-laser/

The FEL has been commissioned in January 2017. A flux of 1.4×10^{14} ph/pulse was achieved with undulator tapering [597]. Then, the performance of the Dalian FEL were achieved with pulses ranging between 100 fs and 1 ps and an energy reaching 1 mJ [598]. It be used to probe fuel combustion, biomolecules behaviour in gases, and reactions process at solid–gas interfaces.

7.4.5 Echo demonstration

NLCTA TEST EXPERIMENT, STANFORD, USA

Experimentally demonstrated so far in the UV experiment on the Next Linear Collider Test Accelerator at SLAC [599–601] with an up-frequency conversion up to the 75th harmonic and later on the 75th harmonic [602]. It constitutes a breakthrough in up-frequency conversion from a conceptual point of view, and in terms of compactness and pulse properties. Echo enables us to provide vortices [603,604].

SHANGHAI FEL, SHANGHAI, CHINA

A first multipurpose test experiment SDUV-FEL was set in the Shanghai Jiading campus for FEL principle studies, with an emphasis on seeding schemes. It uses an 148 MeV electron beam (0.2% energy spread, 4–6π mm mrad emittance, 100–300 pC charge, 2–8 ps duration), a 10 μJ seed at 1.16–1.58 μm, a modulator of 10 $\times$50 mm period (K = 2–3), a radiator of 80 $\times$ 40 mm period (K =0.9–2.5). Different features of harmonic generation have been studied: local energy spread measurements thanks to coherent harmonic generation [605], wide tuneability in the HGHG and cascaded HGHG configurations [606], phase-merging enhanced harmonic generation [607], phase space manipulation for seeding [608]. Polarization switching was tested [609]. Phase space linearization using corrugated chambers was demonstrated [610]. Echo was also achieved on the SDUV-FEL [611].

The SDUV FEL was a test facility in view of the development of the Shanghai X-ray FEL (SXFEL) in the main campus, a user facility in the soft X-ray (8.8 nm with 0.84 GeV with C-band accelerating structures, 0.5 nC charge, below 0.15% energy spread, 2π mm mrad) with cascaded HGHG or echo configurations to be extended to the hard X-ray (2 nm with 1.6 GeV with cascaded HGHG). The installation is completed and the SXFEL is presently under commissioning [612].

7.4.6 FEL self-seeding

Seeding with the FEL itself is also an alternative [613]. Indeed, self-seeding suits better the hard X-ray domain: a monochromator installed after the first undulator spectrally cleans the radiation before the last amplification in the final undulator. Recently, self-seeding with the spectral cleaning of the SASE radiation in a single crystal monochromator [614] appears to be very promising, as experimentally demonstrated at LCLS [615, 616] and at SACLA [617].

7.5 Applications of X-ray FELs

The recent advent of tuneable coherent X-ray FELs (XFELs) [418, 618] opened a new era for the investigation of matter [619]. “It is worthwhile to recount that the first five years of LCLS operation generated many unanticipated methods and discoveries. With many new next- generation x-ray FEL sources coming online in the next five years, the advancement of science will only continue to accelerate” [619]. They enable us to decrypt the structure of biomolecules and cells [620–622], to provide novel insight in the electronic structure of atoms and molecules [623–626], to observe non-equilibrium nuclear motion, disordered media, and distorted crystal lattices, thanks to recent progress of fs spectroscopy [627], and pump-probe techniques [628]. Detailed structural dynamics can be inferred from spectroscopic signatures [629]. XFELs can also reveal chemical reactions movies. With new imaging techniques [630, 631], they are exceptional tools for the investigation of ultrafast evolution of the electronic structure and provide a deeper insight in the extreme states of matter [632].

8 Conclusion and prospects

Among the various light sources such as synchrotron radiation [633], high-order harmonics generated in gas [567, 568], X-ray FELs are unique tuneable coherent light sources from far infrared to the X-ray domain.

Figure 90 shows the evolution of the FEL wavelength versus years. After the first lasing in the infrared in 1977, the second lasing in the visible in 1983, 2000 appeared to bring a transition where VUV is reached both in the oscillator and SASE configurations [634]. Then, thanks to the development of photoinjectors and more generally, to accelerator technology, the X-ray range was reached less than 10 years later, in 2009, 2011 with presently new facilities commissioned in 2016 and 2017, including the European XFEL being a high repetition rate one.

The FEL spatial coherence is usually quite good: in the resonator configuration, it results from the optical cavity for resonators and from the electron beam emittance on single pass systems, and possibly from the seed with optical guiding in the high gain regime. Temporal Coherence is usually good, the Fourier limit is reached in some cases (oscillators, seeding). Femtosecond pulses are possible (and there are various schemes proposed schemes for reaching 100 attosecond pulses). Polarization results from the undulator choice.

Major steps of FEL progress are recalled in Fig. 91.

Present developments are oriented in providing further advanced properties. The two-colour FEL

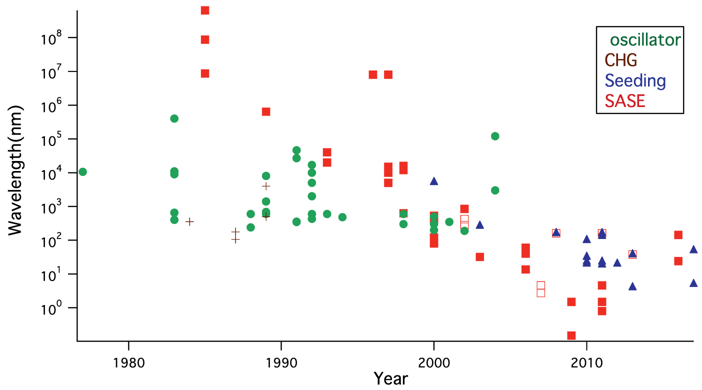


Fig. 90: Achieved FEL wavelengths versus year for various configurations (oscillators, coherent harmonic generation, SASE, seeding).

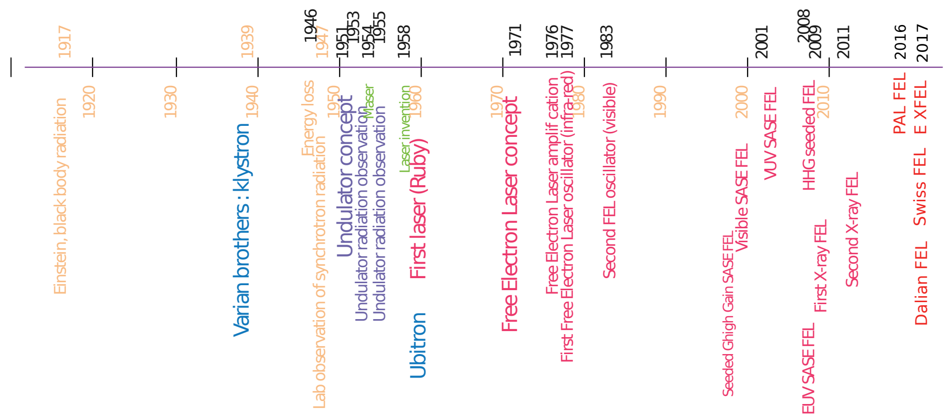


Fig. 91: Major FEL historical steps

concept can be applied to the X-ray domain in the SASE regime, either tuning the two series of undulators at different wavelengths [635–637], the delay being adjusted by a chicane, or by using twin bunches at different energies [638], enabling also operation in the self-seeded case. In the external seeding case, one can take advantage of the pulse splitting effect [438] combined with chirp [593, 594], or apply a double seeding [595, 639]. Several strategies are investigated in the quest towards in attosecond pulses and high peak power.

Another evolution trend is the search for compactness. Besides seeding and up-frequency conversion, one considers implementing the FEL using a compact accelerator or undulator. In a Laser Wakefield Accelerator (LWFA) [640], an intense laser pulse drives plasma density wakes to produce, by charge separation, strong longitudinal electric fields, with accelerating gradient than can reach a 100 GV/m [641, 642]. Electrons have to be set at a proper phase with respect to the wake, to be efficiently accelerated. LWFA can nowadays produce electron beams in the few hundreds of MeV to 1 GeV range with a typical current of a few kiloamperes with reasonable beam characteristics (relative energy spread of the order of 1%, and a normalized emittance of the order of 1π mm mrad. This new accelerating concept could thus be qualified by a FEL application [643, 644]. LWFA based undulator radiation has been observed, even at short wavelengths [645–648]. The present LWFA electron beam properties are not directly suited for enabling FEL amplification, and electron beam manipulation is required: the handling of the divergence with strong permanent magnet quadrupoles, the reduction of the slice energy spread by a demixing chicane [649, 650] where advantage can be taken from the introduced correlation between the energy and the position to focus the slices can be focused in synchronization with the optical wave advance, in the so-called supermatching scheme [651], or in using a transverse gradient undulator [652] coming back to the old FEL times where large energy spread had to be managed [166]. Several experiments are under way.

Fifty years after the laser discovery [653] and more than 30 years after the first FEL, the emergence of several mJ X-ray lasers for users in the Angstrom range constitutes a major breakthrough. Higher availability of X-ray pulses with stable energy, synchronized to an external pump laser, for jitter-free optical pump/resonant X-ray probe experiments will enable us to step further. Besides, exploration of future compact FELs has started. Present X-ray FELs enable us to pave the way towards unrevealed properties of matter and dynamical processes.

Acknowledgements

I would like to dedicate this work to John Madey (1943–2016), the inventor of the FEL, who passed away during summer 2016 and to P. Elleaume (1956–2011), Y. Petroff, M. Billardon, J. M. Ortéga, M. Velghe, who welcomed me in this field. Acknowledgments go also to my former and present students and post-docs, and to Sabine and Amin for careful checks of this report.

References

[1] C.A. Brau, *Free-Electron Lasers* (Academic Press, Boston, 1990).

[2] C.A. Brau, *Science* **239** (1988) 1115–1122, https://doi.org/10.1126/science.239.4844.1115.

[3] C. Pellegrini, *Eur. Phys. J. H*. **37** (2012) 659–708, https://doi.org/10.1140/epjh/e2012-20064-5.

[4] A.L. Schawlow, C.H. Townes, *Phys. Rev*. **112** (1958) 1940, https://doi.org/10.1103/physrev.112.1940.

[5] T.H. Maiman, *Nature* **187** (1960) 493–494, https://doi.org/10.1038/187493a0.

[6] J.M.J. Madey, *J. Appl. Phys*. **42** (1971) 1906–1913, https://doi.org/10.1063/1.1660466.

[7] A. Einstein, *Phys. Z.* **18** (1917) 121.

[8] J.H. van Vleck, *Phys. Rev.* **24** (1924) 330–346, https://doi.org/10.1103/physrev.24.330.

[9] J.H. van Vleck and D.L. Huber, *Rev. Mod. Phys.* **49** (1977) 939–959, https://doi.org/10.1103/revmodphys.49.939.

[10] Y. Petroff, Les rayons X: de l'astrophysique à la nanophysique (1998).

[11] U.E. Kruse, L. Marshall and J.R. Platt, *Astrophys. J.* **124** (1956) 601.

[12] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York Ld, 1962), Vol. 3.

[13] J. Larmor, *London, Edinburgh, Dublin Philos. Mag. J. Sci*. **44** (1897) 503–512, https://doi.org/10.1080/14786444108520780.

[14] A. Liénard, *Éclairage Électrique*. **16** (1898) 320–334.

[15] G.A. Schott, *Ann. Phys*. **24** (1907) 635–660.

[16] G.A. Schott, *Electromagnetic Radiation* (Cambridge University Press, Cambridge, 1912).

[17] D. Iwanenko and I. Pomeranchuk, *Phys. Rev*. **65** (1944) 343, https://doi.org/10.1103/physrev.65.343.

[18] M.O. Oliphant, The acceleration of particles to very high energies. *Classified memo submitted to DSIR University of Birmingham Archive* (1943).

[19] E.M. McMillan, *Phys. Rev*. **68** (1945) 143, https://doi.org/10.1103/physrev.68.143.

[20] V. Veksler, *J. Phys. USSR*. **9** (1945) 153–158.

[21] F.K. Goward and D.E. Barnes, *Nature* **158** (1946) 413, https://doi.org/10.1038/158413a0.

[22] J. Schwinger, *Phys. Rev*. **70** (1946) 798–799, https://doi.org/10.1103/physrev.70.41.

[23] J. Schwinger, *Phys. Rev*. **75** (1949) 1912, https://doi.org/10.1103/physrev.75.1912.

[24] J.P. Blewett, *Phys. Rev*. **69** (1946) 87, https://doi.org/10.1103/physrev.69.87.

[25] J.P. Blewett, *J. Synchrotron Radiat*. **5** (1998) 135–139, https://doi.org/10.1107/s0909049597043306.

[26] F.R. Elder, A.M. Gurewitsch, R.V. Langmuir and H.C. Pollock, *Phys. Rev*. **71** (1947) 829, https://doi.org/10.1103/physrev.71.829.5.

[27] V.L. Ginzburg, *Bull. Acad. Sci. USSR*. **11** (1947) 165–182.

[28] H. Motz, *J. Appl. Phys*. **22** (1951) 527–535, https://doi.org/10.1063/1.1700002.

[29] H. Motz and M. Nakamura, *Ann. Phys*. **7** (1959) 84–131, https://doi.org/10.1016/0003-4916(59)90049-1.

[30] H. Motz, W, Thon and R.N. Whitehurst, *J. Appl. Phys*. **24** (1953) 826–833, https://doi.org/10.1063/1.1721389.

[31] R. Combes and T. Frelot, *C R Hebd. Scéance Acad. Sci. Paris*. **241** (1955) 1559–1560.

[32] K. Halbach, *Le J. Phys. Colloq*. **44(C1)** (1983) 211–216, https://doi.org/10.1051/jphyscol:1983120.

[33] P.L. Csonka, *Part Accel*. **8** (1978) 225–234.

[34] T. Nakazato, M. Oyamada, N. Niimura, S. Urasawa, O. Konno, *et al*., *Phys. Rev. Lett*. **63** (1989) 1245, https://doi.org/10.1103/physrevlett.63.1245.

[35] R.H. Varian and S.F. Varian, *J. Appl. Phys*. **10** (1939) 321–327, https://doi.org/10.1063/1.1707311.

[36] J.M. Ortéga, Généralités sur le LEL, Notes de cours.

[37] K. Landecker, *Phys. Rev*. **86** (1952) 852, https://doi.org/10.1103/physrev.86.852.

[38] R.M. Phillips, *Nucl. Instrum. Methods Phys. Res. A*. **272** (1988) 1–9, https://doi.org/10.1016/j.nima.2016.03.040.

[39] R.M. Phillips, *IRE Trans. Electron Devices, ED-7*. **7** (1960) 231–241.

[40] C.E. Enderby and R.M. Phillips, *Proc. IEEE*. **53** (1965) 1648–1648.

[41] R.B. Palmer, *J. Appl. Phys*. **43** (1972) 3014–3023, https://doi.org/10.1063/1.1661650.

[42] R.H. Pantell, G. Soncini and H.E. Puthoff, *IEEE J. Quantum Electron*. **4** (1968) 905–907, https://doi.org/10.1109/jqe.1968.1074989.

[43] J.P. Gordon, H.J. Zeiger and C.H. Townes, NH_3*Phys. Rev*. **95** (1954) 282, https://doi.org/10.1103/physrev.95.282.

[44] R.G. Gould, The laser, light amplification by stimulated emission of radiation. *The Ann Arbor Conference on Optical Pumping* (The University of Michigan, 1959), vol. 15, p. 128.

[45] H. Kogelnik, T. Li, *Appl. Opt*. **5** (1966) 1550–1567, https://doi.org/10.1364/ao.5.001550.

[46] A. Yariv. *Quantum Electronics* (John Wiley & Sons, New York, 1975), p. 123.

[47] T.H. Maiman, R.H. Hoskins, I.J. D'Haenens, C.K. Asawa and V. Evtuhov, *Phys. Rev*. **123** (1961) 1151, https://doi.org/10.1103/physrev.123.1151.

[48] C. Fabre, *Images Phys* **30** (2010) 3–9.

[49] A. Javan, W.R. Bennett and D.R. Herriott, *Phys. Rev. Lett*. **6** (1961) 106, https://doi.org/10.1103/physrevlett.6.106.

[50] R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys and R.O. Carlson. *Phys. Rev. Lett*. **9** (1962) 366, https://doi.org/10.1103/physrevlett.9.366.

[51] R.J. Keyes and T.M. Quist, *Proc. IRE*. **1822** (1962) 16 IEEE.

[52] E. Schrödinger, *Ann. Phys*. **387** (1927) 257–264, https://doi.org/10.1002/andp.19273870210.

[53] P.L. Kapitza and P.A.M. Dirac, *Math. Proc. Cambridge Philos. Soc*. **29** (1933) 297–300, https://doi.org/10.1017/s0305004100011105.

[54] M.V. Fedorov, *Sov. J. Exp. Theor. Phys*. **24** (1967) 529.

[55] J.M.J. Madey, *Phys. Rev. Special Topics-Accelerators Beams*. **17** (2015) 074901, https://doi.org/10.1103/physrevstab.17.074901.

[56] J.M.J. Madey, Free Electron Lasers and the path to understanding. *Nobel Symposium on Free Electron Laser Research* (2015), pp. 14–18.

[57] H. Dreicer, *Phys. Fluids*. **7** (1964) 735–753, https://doi.org/10.1063/1.1711276.

[58] D. Marcuse, *Bell Syst Tech. J*. **41** (1962) 1557–1571, https://doi.org/10.1002/j.1538-7305.1962.tb03994.x.

[59] R.Q. Twiss, *Aust. J. Phys*. **11** (1958) 564–579, https://doi.org/10.1071/ph580564.

[60] J.M.J. Madey, E.T. Scharlemann and W.M. Fairbank, *IEEE Trans. Nucl. Sci.* **20** (1973) 980–980, https://doi.org/10.1109/tns.1973.4327304.

[61] V.P. Sukhatme and P.A. Wolff, *J. Appl. Phys*. **44** (1973) 2331–2334, https://doi.org/10.1063/1.1662560.

[62] N.M. Kroll and W.A. McMullin, *Phys. Rev. A*. **17** (1978) 300, https://doi.org/10.1103/physreva.17.300.

[63] P. Sprangle and V.L. Granatstein, *Appl. Phys. Lett*. **25** (1974) 377–379.

[64] P. Sprangle, V.L. Granatstein and L. Baker, *Phys. Rev. A*. **12** (1975) 1697.

[65] H. Haus, *IEEE J. Quantum Electron*. **17** (1981) 1427–1435, https://doi.org/10.1109/jqe.1981.1071289.

[66] A. Fruchtman, *Phys. Rev. A*. **37** (1988) 4252, https://doi.org/10.1103/physreva.37.4252.

[67] J. Fajans, J.S. Wurtele, G. Bekefi, D.S. Knowles and K. Xu, *Phys. Rev. Lett*. **57** (1986) 579, https://doi.org/10.1103/physrevlett.57.579.

[68] M.E. Conde and G. Bekefi, *Phys. Rev. Lett*. **67** (1991) 3082, https://doi.org/10.1103/physrevlett.67.3082.

[69] T.C. Marshall, *Free Electron Lasers* (Macmillan, New York, 1985).

[70] F.A. Hopf, P. Meystre, M.O. Scully and W.H. Louisell, *Phys. Rev. Lett*. **37** (1976) 1215, https://doi.org/10.1103/physrevlett.37.1215.

[71] F.A. Hopf, P. Meystre, M.O. Scully and W.H. Louisell, *Phys. Rev. Lett*. **37** (1976) 1342, https://doi.org/10.1103/physrevlett.37.1342.

[72] H. Al Abawi, F.A. Hopf, P. Meystre, *Phys. Rev. A*. **16** (1977) 666, https://doi.org/10.1103/physreva.16.666.

[73] P. Sarangle, *J. Plasma Phys*. **11** (1974) 299–309, https://doi.org/10.1017/s002237780002465x.

[74] T. Kwan, J.M. Dawson and A.T. Lin, *Plasma Phys* **20** (1977) 486–506,

[75] W.B. Colson, *Phys. Lett. A*. **59** (1976) 187–190, https://doi.org/10.1016/0375-9601(76)90561-2.

[76] W.B. Colson, *Phys. Lett. A*. **64** (1977) 190–192, https://doi.org/10.1016/0375-9601(77)90712-5.

[77] W.B. Coslon, Ph.D. thesis, Stanford University, 1977.

[78] W.B. Colson, Free electron laser theory, a dissertation submitted to the department of physics and the committee on graduate studies of stanford university. Technical report (Stanford University, 1977).

[79] W.H. Louisell, J.F. Lam, D.A. Copeland and W.B. Colson, *Phys. Rev. A*. **19** (1979) 288,

https://doi.org/10.1103/physreva.19.288.

[80] W.B. Colson and S.K. Ride, *Phys. Lett. A*. **76** (1980) 379–382, https://doi.org/10.1016/0375-9601(80)90737-9.

[81] W.B. Colson, *IEEE J. Quantum Electron*. **17** (1981) 1417–1427, https://doi.org/10.1109/jqe.1981.1071273.

[82] W.B. Colson, *Phys. Rev. A*. **24** (1981) 639, https://doi.org/10.1103/physreva.24.639.

[83] R. Coisson and F. De Martini. *Free-Electron Generators of Coherent Radiat*. **9** (1982) 939–1003.

[84] W.B. Colson, C. Pellegrini and A. Renieri, *Laser Handbook, Volume 6: Free Electron Lasers* (North Holland, 1990).

[85] W.B. Colson, *Nucl. Instrum. Methods Phys. Res. A*. **237** (1985) 1–9, https://doi.org/10.1016/0168-9002(85)90323-7.

[86] Z. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, now National Institute of Standards and Technology (NIST) Applied Mathematics Series - 55, Issued June 1964, Seventh printing, (1968).

[87] J.M.J. Madey, *Il Nuovo Cim. B Series 11*. **50** (1979) 64–88, https://doi.org/10.1007/bf02737619.

[88] W.B. Colson and P. Elleaume, *Appl. Phys. B*. **29** (1982) 101–109, https://doi.org/10.1007/bf00692771.

[89] D. Nutarelli, D. Garzella, M.E. Couprie, M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 64–69.

[90] A. Bambini and A. Renieri, *Coherence Spectrosc. Mod. Phys* **37** (1978) 361–379, https://doi.org/10.1007/978-1-4613-2871-1_18.

[91] A. Bambini, A. Renieri and S. Stenholm, *Phys. Rev. A*. **19** (1979) 2013, https://doi.org/10.1103/physreva.19.2013.

[92] A. Bambini, R Bonifacio and S. Stenholm, *Opt. Commun*. **32** (1980) 306–308, https://doi.org/10.1016/0030-4018(80)90130-3.

[93] G. Dattoli, A. Marino, A. Renieri and F. Romanelli, *IEEE J. Quantum Electron*. **17** (1981) 1371–1387, https://doi.org/10.1109/jqe.1981.1071268.

[94] G. Dattoli and R. Mignani, *J. Math. Phys*. **26** (1985) 3200–3203, https://doi.org/10.1063/1.526649.

[95] R. Bonifacio and M.O. Scully, *Opt. Commun*. **32** (1980) 291–294, https://doi.org/10.1016/0030-4018(80)90127-3.

[96] G. Dattoli, T. Hermsen, A. Renieri, A. Torre and J.C. Gallardo, *Phys. Rev. A*. **37** (1988) 4326, https://doi.org/10.1103/physreva.37.4326.

[97] G. Dattoli, T. Hermsen, L. Mezi, A. Renieri and A. Torre, *Phys. Rev. A*. **37** (1988) 4334, https://doi.org/10.1103/physreva.37.4334.

[98] G. Dattoli, *J. Appl. Phys*. **84** (1998) 2393–2398, https://doi.org/10.1063/1.368365.

[99] G. Dattoli, A. Renieri and A. Torre, *Lectures on the Free Electron Laser Theory and Related Topics* (World Scientific, Singapore, 1993), Vol. 637.

[100] F. Ciocci, G. Dattoli, A. Torre and A. Renieri, in *Series on Synchrotron Radiation Techniques and Applications*, Eds. D.H. Bilderback, K.O. Hodgson, M.P. Kiskinov and R. Rossi (World Scientific, Singapore, 2000), Vol. 6.

[101] F.V. Bunkin and V.D. Novikov, *Vestin. Akad. Nauk SSSR*. **8** (1981) 111–113.

[102] N.A. Vinokurov and A.N. Skrinsky, Power limit for an optical klystron installed on storage ring. *Preprint INP*, (1977) pp. 77–59.

[103] N.A. Vinokurov and A.N. Skrinsky, Generatornii klystron optitcheskovo diapazona na ultrarelativistiskii elektropnax. klystron generating in the optical range with relativistic electrons, preprint

INP 77–59. Technical report, Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk (1977).

[104] N.A. Vinokurov and A.N. Skrinsky, O predelnoi mochnosti optitchekovo klystrona, ustanovlennovo na elektronnii nakonitel. maximum power out of an optical klystron installed on the electron storage ring, preprint INP 77-67. Technical report, Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk (1977).

[105] N.A. Vinokurov and A.N. Skrinsky, O predelnoi mochnosti optitchekovo klystrona, ustanovlennovo na lineinii uskoritel elektronob. maximum power of an optical klystron installed on a linear accelerator, preprint INP 78-88. Technical report, Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk (1977).

[106] R. Coisson, *Part. Accel*. **11** (1981) 245–253.

[107] P. Elleaume, *Le J. Phys. Colloques*. **44(C1)** (1983) C1–333, https://doi.org/10.1051/jphyscol:1983127.

[108] R.R. Freeman and B.M. Kincaid, in *Laser Techniques in the Extreme Ultraviolet*, (AIP Publishing, 1984), Vol. 119, pp. 278–292.

[109] P. Elleaume, *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 220–227, https://doi.org/10.1016/0168-9002(86)90884-3.

[110] V.N. Litvinenko, J.M.J. Madey and N.A. Vinokurov, UV-VUV FEL Program at Duke Storage Ring with OK-4 optical klystron. In *Particle Accelerator Conference* (IEEE, 1993), pp. 1442–1444.

[111] M.E. Couprie, M. Belakhovsky, B. Gilquin, D. Garzella, M. Jablonka, F. Méot, P. Monot, A. Mosnier, L. Nahon and A. Rousse, *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 557–561, https://doi.org/10.1016/j.nima.2004.04.101.

[112] Y.K. Wu, N.A. Vinokurov, S. Mikhailov, J. Li and V. Popov, *Phys. Rev. Lett*. **96** (2006) 224801, https://doi.org/10.1103/physrevlett.96.224801.

[113] V.N. Baier and A.I. Milstein. To the theory of a free-electron laser, preprint INP 77-67. Technical report (Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk, 1977).

[114] V.N. Baier and A.I. Milstein, *Phys. Lett. A*. **65** (1978) 319–322, https://doi.org/10.1016/0375-9601(78)90716-8.

[115] V.N. Baier and A.I. Milstein, O generatsii kogerentnovo izlutcheniya blizko tsiklotronnovo pezonansa, generation of coherent radiation close to the cyclotron resonance, preprint INP 78-61. Technical report, Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk (1978).

[116] A.M. Kondratenko, *Sou Phys. Dokl. Part. Accelerators*. **24** (1979) 989.

[117] A.M. Kondratenko and E.L. Saldin, Generation of coherent radiation by a relativistic-electron beam in an undulator, preprint INP 79-48. Technical report (Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk, 1978).

[118] A.M. Kondratenko and E.L. Saldin, Generation of coherent radiation by a relativistic-electron beam in an undulator. In *Particle Accelerator Conference*, vol. 10 (1986), pp. 207–216, Gordon and Breach, Science Publishers, Inc..

[119] A.M. Kondratenko and E.L. Saldin, On the linear theory of free electron lasers with Fabry-Perot cavities, Preprint INP 80-172. Technical report (Institute of Nuclear Physics of the Siberian Division of the USSR Academy of Sciences, Novosibirsk, 1980).

[120] N.M. Kroll, Ph.L. Morton and M.N. Rosenbluth, in *Free-Electron Generators of Coherent Radiation*, Eds. SF Jacobs, HS Pilloff, M. Sargent, MO Scully and R. Spitzer, Vol. 1 (1980), pp. 89–112, Physics of Quantum Electronics, (Addison-Wesley, Reading, MA).

[121] W.B. Colson and R.A. Freedman, *Opt. Commun*. **46** (1983) 37–42, https://doi.org/10.1016/0030-

4018(83)90026-3.
[122] W.B. Colson, *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 168–175, https://doi.org/10.1016/0168-9002(86)90878-8.
[123] J.E. Sollid, D.W. Feldman and R.W. Warren, *Nucl. Instrum. Methods Phys. Res. A*. **285** (1989) 153–157, https://doi.org/10.1016/0168-9002(89)90443-9.
[124] Z. Huang and K.-J. Kim, *Phys. Rev. Special Topics-Accelerators and Beams*. **10** (2007) 034801.
[125] A.T. Lin and J.M. Dawson, *Phys. Rev. Lett*. **42** (1979) 1670, https://doi.org/10.1103/physrevlett.42.1670.
[126] P. Sprangle, C.-M. Tang and W.M. Manheimer, *Phys. Rev. Lett*. **43** (1979) 1932, https://doi.org/10.1103/physrevlett.43.1932.
[127] N.A. Vinokurov, O.A. Shevchenko and V.G. Tcheskidov, *Phys. Rev. Special Topics-Accelerators and Beams*. **14** (2011) 040701.
[128] P. Bosco and W.B. Colson, *Phys. Rev. A*. **28** (1983) 319, https://doi.org/10.1103/physreva.28.319.
[129] A. Renieri, *Il Nuovo Cim. B (1971–1996).* **53** (1979) 160–178, https://doi.org/10.1007/bf02739308.
[130] A. Renieri, *IEEE Trans. Nucl. Sci*. **26** (1979) 3827–3829, https://doi.org/10.1109/tns.1979.4330622.
[131] G. Dattoli and A. Renieri, *Il Nuovo Cim. B Series 11*. **59** (1980) 1–39, https://doi.org/10.1007/bf02739044.
[132] A. Renieri, Turbulence and bunch lengthening in electron-positron storage rings. *Frascati, Report. No. LNF-76* (1976), p. 11.
[133] G. Dattoli and A. Renieri, *Lett. Nuovo Cim*. **24** (1979) 121–128, https://doi.org/10.1007/bf02725605.
[134] G. Dattoli and A. Renieri, *Il Nuovo Cim. B*. **61** (1981) 153–180, https://doi.org/10.1007/bf02721321.
[135] G. Dattoli, A. Marino and A. Renieri. *Opt. Commun*. **35** (1980) 407–412, https://doi.org/10.1016/0030-4018(80)90062-0.
[136] W.B. Colson and J.L. Richardson, *Phys. Rev. Lett*. **50** (1983) 1050, https://doi.org/10.1103/physrevlett.50.1050.
[137] P. Elleaume and D.A.G. Deacon, *Appl. Phys. B*. **33** (1984) 9–16, https://doi.org/10.1007/bf00690020.
[138] L.R. Elias and J.C. Gallardo, *Phys. Rev. A*. **24** (1981) 3276, https://doi.org/10.1103/physreva.24.3276.[138] W.B. Colson and R.A. Freedman, *Phys. Rev. A*. **27** (1983) 1399, https://doi.org/10.1103/physreva.27.1399.
[139] W.B. Colson and R.A. Freedman, *Phys. Rev. A*. **27** (1983) 1399, https://doi.org/10.1103/physreva.27.1399.
[140] R. Bonifacio, *Opt. Commun*. **32** (1980) 440–442, https://doi.org/10.1016/0030-4018(80)90279-5.
[141] G. Dattoli, A. Renieri, F. Romanelli and R. Bonifacio, *Opt. Commun*. **34** (1980) 240–244, https://doi.org/10.1016/0030-4018(80)90024-3.
[142] R.J. Glauber, *Phys. Rev*. **130** (1963) 2529, https://doi.org/10.1103/physrev.130.2529.
[143] R.J. Glauber, *Phys. Rev*. **131** (1963) 2766, https://doi.org/10.1103/physrev.131.2766.
[144] G. Dattoli and A. Renieri, *Opt. Commun*. **39** (1981) 328–330, https://doi.org/10.1016/0030-4018(81)90102-4.
[145] G. Dattoli and A. Renieri, *Nucl. Instrum. Methods Phys. Res*. **208** (1983) 193–197, https://doi.org/10.1016/0167-5087(83)91124-9.
[146] I. Boscolo and V. Stagno, *Il Nuovo Cim. B*. **58** (1980) 267–285,

https://doi.org/10.1007/bf02874012.
[147] G. Penn, M. Reinsch and J.S. Wurtele, *Phys. Rev. Spec. Top. Accel. Beams.* **9** (2006) 060702, https://doi.org/10.1103/physrevstab.9.060702.
[148] G. Dattoli, P.L. Ottaviani and S. Pagnutti, *J. Appl. Phys*. **101** (2007) 024914, https://doi.org/10.1063/1.2424398.
[149] J.M. Ortéga, *Synchrotron Radiat. News.* **2** (1989) 18–23, https://doi.org/10.1080/08940888908261211.
[150] G. Dattoli, *Lett. Nuovo Cim*. **27** (1980) 247–251, https://doi.org/10.1007/bf02750352.
[151] A. Gover and P. Sprangle, *IEEE J. Quantum Electron*. **17** (1981) 1196–1215, https://doi.org/10.1109/jqe.1981.1071257.
[152] M. Orszag and R. Ramirez, *J. Opt. Soc. Am. B.* **3** (1986) 895–900, https://doi.org/10.1364/josab.3.000895.
[153] M. Orszag, *Phys. Rev. A*. **36** (1987) 189, https://doi.org/10.1103/physreva.36.189.
[154] R. Bonifacio, N. Piovella and G.R.M. Robb, *Nucl. Instrum. Methods Phys. Res. A*. **543** (2005) 645–652, https://doi.org/10.1016/j.nima.2005.01.324.
[155] R. Bonifacio, N. Piovella, G.R.M. Robb and A. Schiavi, *Phys. Rev. Spec. Top. Accel Beams*. **9** (2006) 090701, https://doi.org/10.1103/physrevstab.9.090701.
[156] N. Piovella, M.M. Cola, L. Volpe, A. Schiavi and R. Bonifacio, *Phys. Rev. Lett*. **100** (2008) 044801, https://doi.org/10.1103/physrevlett.100.044801.
[157] R. Bonifacio and F. Casagrande, *Opt. Commun*. **50** (1984) 251–255, https://doi.org/10.1016/0030-4018(84)90327-4.
[158] R. Bonifacio, N. Piovella, G.R.M. Robb and M.M. Cola, *Opt. Commun*. **252** (2005) 381–396, https://doi.org/10.1016/j.optcom.2005.04.037.
[159] J. Rossbach, E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *Nucl. Instrum. Methods Phys. Res. A*. **374** (1996) 401–407, https://doi.org/10.1016/0168-9002(96)00189-1.
[160] M.S. McAshan, K. Mittag, H.A. Schwettman, L.R. Suelzle and J.P. Turneaure, *Appl. Phys. Lett*. **22** (1973) 605–607, https://doi.org/10.1063/1.1654525.
[161] L.R. Elias, W.M. Fairbank, J.M.J. Madey, H.A. Schwettman and T.I. Smith, *Phys. Rev. Lett*. **36** (1976) 717, https://doi.org/10.1103/physrevlett.36.717.
[162] D.A.G. Deacon, L.R. Elias, J.M.J. Madey, G.J. Ramian, H.A. Schwettman and T.I. Smith, *Phys. Rev. Lett*. **38** (1977) 892–894, https://doi.org/10.1103/physrevlett.38.892.
[163] J.N. Eckstein, J.M.J. Madey, K. Robinson and T.I. Smith. Internal report HEPL-894. Technical Report, Stanford University (1981).
[164] S. Benson, D.A.G. Deacon, J.N. Eckstein, J.M.J. Madey, K. Robinson, T.I. Smith and R. Taberμ, *Phys. Rev. Lett*. **48** (1982) 235, https://doi.org/10.1103/physrevlett.48.235.
[165] D.A.G. Deacon and J.M.J. Madey, *Phys. Rev. Lett*. **44** (1980) 449, https://doi.org/10.1103/physrevlett.44.449.
[166] T.I. Smith, J.M.J. Madey, L.R. Elias and D.A.G. Deacon, *J. Appl. Phys*. **50** (1979) 4580–4583, https://doi.org/10.1063/1.326564.
[167] J.M.J. Madey, *J. Phys. Colloq*. **44(C1)** (1983) C1–169, https://doi.org/10.1051/jphyscol:1983116.
[168] A.S. Artamonov, N.A. Vinokurov, P.D. Voblyi, E.S. Gluskin, G.A. Kornyukhin, V.A. Kochubei, G.N. Kulipanov, V.N. Litvinenko, N.A. Mezentsev and A.N. Skrinsky, *Nucl. Instrum. Methods*. **177** (1980) 247–252, https://doi.org/10.1016/0029-554x(80)90557-1.
[169] R. Barbini, G. Vignola, S. Trillo, R. Boni, S. De Simone, S. Faini, S. Guiducci, M. Preger, M. Serio, B. Spataro, *et al.*, *J. Phys. Colloq*. **44(C1)** (1983) C1–1, https://doi.org/10.1051/jphyscol:1983101.

[170] S. Krinsky, A. Luccio, C. Pellegrini, A. Van Steenbergen and L.H. Yu, *J. Phys. Colloq*. **44(C1)** (1983) C1–113, https://doi.org/10.1051/jphyscol:1983107.

[171] L.R. Elias, *Phys. Rev. Lett*. **42** (1979) 977, https://doi.org/10.1103/physrevlett.42.977.

[172] L.R Elias and G. Ramian, *IEEE Trans. Nucl. Sci*. **32** (1985) 1732–1735.

[173] L.R. Elias and G. Ramian. Design of the UCSB FEL electron beam system. Technical Report, DTIC Document (1981).

[174] T.I. Smith and J.M.J. Madey, *Appl. Phys. B*. **27** (1982) 195–199, https://doi.org/10.1007/bf00697441.

[175] M. Billardon, P. Elleaume, J.M. Ortéga, C. Bazin, M. Bergher, M. Velghe, Y. Petroff, D.A.G. Deacon, K.E. Robinson and J.M.J. Madey, *Phys. Rev. Lett*. **51** (1983) 1652–1655, https://doi.org/10.1103/physrevlett.51.1652.

[176] M. Billardon, P. Elleaume, J.M. Ortéga, C. Bazin, M. Bergher, Y. Petroff, M. Velghe, D.A.G. Deacon and J.M.J. Madey. First operation of a storage ring free-electron laser. In *Free-electron generators of coherent radiation* (International Society for Optics and Photonics, 1984), pp. 269–274.

[177] J.A. Edighoffer, G.R. Neil, C.E. Hess, T.I. Smith, S.W. Fornaca and H.A. Schwettman, *Phys. Rev. Lett*. **52** (1984) 344, https://doi.org/10.1103/physrevlett.52.344.

[178] R.W. Warren, B.E. Newnam, W.E. Stein, J.G. Winston, R.L. Sheffield, M.T. Lynch, J.C. Goldstein, M.C. Whitehead, O.R. Norris and G. Luedemann. First operation of the Los Alamos free-electron laser oscillator. Technical Report, Los Alamos National Lab (1983).

[179] D.B. McDermott, T.C. Marshall, S.P. Schlesinger, R.K. Parker and V.L. Granatstein, *Phys. Rev. Lett*. **41** (1978) 1368, https://doi.org/10.1103/physrevlett.41.1368.

[180] M. Bazin, Y. Farge, M. Lemonnier, J. Perot and Y. Petroff, *Nucl. Instrum. Methods*. **172** (1980) 61–65, https://doi.org/10.1016/0029-554x(80)90608-4.

[181] C. Bazin, M. Billardon, D. Deacon, Y. Farge, J.M. Ortéga, J. Pérot, Y. Petroff and M. Velghe, *J. Phys. Lett*. **41** (1980) 547–550, https://doi.org/10.1051/jphyslet:019800041023054700.

[182] D.A.G. Deacon, K.E. Robinson, J.M.J. Madey, C. Bazin, M. Billardon, P. Elleaume, Y. Farge, J.M. Ortéga, Y. Petroff and M.F. Velghe, *Opt. Commun*. **40** (1982) 373–378, https://doi.org/10.1016/0030-4018(82)90366-2.

[183] J.M. Ortéga, C. Bazin, D.A.G. Deacon, C. Depautex and P. Elleaume, *Nucl. Instrum. Methods Phys. Res*. **206** (1983) 281–288, https://doi.org/10.1016/0167-5087(83)91269-3.

[184] J.M. Ortéga, M. Billardon, G. Jezequel, P. Thiry and Y. Petroff, *J. Phys*. **45** (1984) 1883–1888, https://doi.org/10.1051/jphys:0198400450120188300.

[185] M. Billardon, D.A.G. Deacon, P. Elleaume, J.M. Ortéga, K.E. Robinson, C. Bazin, M. Bergher, J.M.J. Madey, Y. Petroff and M. Velghe, *J. Phys. Colloq*. **44(C1)** (1983) C1–29, https://doi.org/10.1051/jphyscol:1983103.

[186] P. Elleaume, M. Velghe, M. Billardon and J.M. Ortéga, *Appl. Opt*. **24** (1985) 2762–2770, https://doi.org/10.1364/ao.24.002762.

[187] D.A.G. Deacon, M. Billardon, P. Elleaume, J.M. Ortéga, K.E. Robinson, C. Bazin, M. Bergher, M. Velghe, J.M.J. Madey and Y. Petroff, *Appl. Phys. B*. **34** (1984) 207–219, https://doi.org/10.1007/bf00697637.

[188] P. Elleaume, J.M. Ortéga, M. Billardon, C. Bazin, M. Bergher, M. Velghe, Y. Petroff, D.A.G. Deacon, K.E. Robinson and J.M.J. Madey, *J. Phys*. **45** (1984) 989–996, https://doi.org/10.1051/jphys:01984004506098900.

[189] M. Billardon, J.M. Ortéga, P. Elleaume, C. Bazin, M. Bergher, M. Velghe, D. Deacon and Y. Petroff, *IEEE J. Quantum Electron*. **21** (1985) 805–823, https://doi.org/10.1109/jqe.1985.1072733.

[190] P. Elleaume, *J. Phys*. **45** (1984) 997–1001, https://doi.org/10.1051/jphys:01984004506099700.

[191] K.E. Robinson, J.M.J. Madey, M.F. Velghe and D.A.G. Deacon, *IEEE Trans. Nucl. Sci*. **30** (1983) 3088–3090, https://doi.org/10.1109/tns.1983.4336578.

[192] M.E. Couprie and P. Elleaume, *Nucl. Instrum. Methods Phys. Res. A*. **259** (1987) 77–82, https://doi.org/10.1016/0168-9002(87)90432-3.

[193] B. Girard, Y. Lapierre, J.M. Ortéga, C. Bazin, M. Billardon, P. Elleaume, M. Bergher, M. Velghe and Y. Petroff, *Phys. Rev. Lett*. **53** (1984) 2405–2408, https://doi.org/10.1103/physrevlett.53.2405.

[194] R. Prazeres, J.M. Ortéga, C. Bazin, M. Bergher, M. Billardon, M.E. Couprie, H. Fang, M. Velghe and Y. Petroff, *Europhys. Lett.* **4** (1987) 817, https://doi.org/10.1209/0295-5075/4/7/010.

[195] R. Prazeres, J.M. Ortéga, C. Bazin, M. Bergher, M. Billardon, M.E. Couprie, M. Velghe and Y. Petroff, *Nucl. Instrum. Methods Phys. Res. A*. **272** (1988) 68–72, https://doi.org/10.1016/0168-9002(88)90197-0.

[196] K. Halbach, *Nucl. Instrum. Methods*. **169** (1980) 1–10, https://doi.org/10.1016/0029-554x(80)90094-4.

[197] R. Warren, B. Newnam, J. Winston, W. Stein, L. Young and C. Brau, *IEEE J. Quantum Electron*. **19** (1983) 391–401, https://doi.org/10.1109/jqe.1983.1071854.

[198] R.W. Warren, B.E. Newnam, W.E. Stein, J.G. Winston, R.L. Sheffield, M.T. Lynch, J.C. Goldstein, M.C. Whitehead, O.R. Norris, G. Luedemann, T.O. Gibson and C.M. Humphry. First operation of the Los Alamos free-electron laser oscillator. In *Sixth International Conference on Lasers and Applications, Lasers, San Franscisco, 12–16 December 1983*, vol. 425 (Society for Optical and Quantum Electronics, 1983), p. 042016.

[199] B.E. Newnam, R.W. Warren, K. Boyer, J.C. Goldstein, M.C. Whitehead and C.A. Brau. Optical gain results of the Los Alamos free-electron laser amplifier experiment. In *Free-electron generators of coherent radiation* (International Society for Optics and Photonics, 1984), pp. 118–129.

[200] B.E. Newnam, R.W. Warren, R.L. Sheffield, J.C. Goldstein and C.A. Brau, *Nucl. Instrum. Methods Phys. Res. A*. **237** (1985) 187–198, https://doi.org/10.1016/0168-9002(85)90347-x.

[201] K. Boyer, C.A. Brau, B.E. Newnam, W.E. Stein, R.W. Warren, J.G. Winston and L.M. Young, *J. Phys*. **C1** (1983) 109–111, https://doi.org/10.2172/6346334.

[202] D.W. Feldman, H. Takeda, R.W. Warren, J.E. Sollid, W.E. Stein, W.J. Johnson, A.H. Lumpkin and R.B. Feldman, *Nucl. Instrum. Methods Phys. Res. A*. **285** (1989) 11–16, https://doi.org/10.1016/0168-9002(89)90418-x.

[203] L.R. Elias and J.C. Gallardo, *Appl. Phys. B*. **31** (1983) 229–233, https://doi.org/10.1007/bf00690793.

[204] L.R. Elias, J. Hu and G. Ramian, *Nucl. Instrum. Methods Phys. Res. A*. **237** (1985) 203–206, https://doi.org/10.1016/0168-9002(85)90349-3.

[205] L.R. Elias, G. Ramian, J. Hu and A. Amir, *Phys. Rev. Lett*. **57** (1986) 424, https://doi.org/10.1103/physrevlett.57.424.

[206] W.B. Colson, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 532–535, https://doi.org/10.1016/0168-9002(94)01547-3.

[207] W.B. Colson, *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 6–8, https://doi.org/10.1016/s0168-9002(97)00418-x.

[208] W.B. Colson, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 397–400, https://doi.org/10.1016/s0168-9002(01)01687-4.

[209] D.H.C.C. Levy *et al*., *Free Electron Lasers and Other Advanced Sources of Light: Scientific Research Opportunities* (National Academy Press, Washington, 1994).

[210] M.E. Couprie, *Ann. Phys*. **19(C1)** (1994) C1–127, https://doi.org/10.1051/anphys/1994028.

[211] M.E. Couprie, *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 13–17,

https://doi.org/10.1016/s0168-9002(97)00420-8.

[212] I.B. Drobyazko, G.N. Kulipanov, V.N. Litvinenko, I.V. Pinayev, V.M. Popik, I.G. Silvestrov, A.N. Skrinsky, A.S. Sokolov and N.A. Vinokurov. Lasing in visible and ultraviolet regions in an optical klystron installed on the VEPP-3 storage ring. *Preprint INP* (1989) 89–118.

[213] I.B. Drobyazko, G.N. Kulipanov, V.N. Litvinenko, I.V. Pinayev, V.M. Popik, I.G. Silvestrov, A.N. Skrinsky, A.S. Sokolov and N.A. Vinokurov. Lasing in visible and ultraviolet regions in optical klystron installed on the VEPP-3 storage ring. In *1989 Intl Congress on Optical Science and Engineering* (International Society for Optics and Photonics, 1989), pp. 2–10.

[214] N.A. Vinokurov, I.B. Drobyazko, G.N. Kulipanov, V.N. Litvinenko and I.V. Pinayev, *Rev. Sci. Instrum.* **60** (1989) 1435–1438, https://doi.org/10.1063/1.1141084.

[215] G.N. Kulipanov, V.N. Litvinenko, I.V. Pinaev, V.M. Popik, A.N. Skrinsky, A.S. Sokolov and N.A. Vinokurov, *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 1–3, https://doi.org/10.1016/0168-9002(90)91179-f.

[216] G.N. Kulipanov, V.N. Litvinenko, A.S. Sokolov and N.A. Vinokurov, On mutual coherency of spontaneous radaition from two undulators separated by achromatic bend. *Preprint INP* (1989) 89–84.

[217] N.G. Gavrilov, G.N. Kulipanov, V.N. Litvinenko, I.V. Pinayev, V.M. Popik, I.G. Silvestrov, A.N. Skrinsky, A.S. Sokolov, N.A. Vinokurov and P.D. Vobly, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 63–65, https://doi.org/10.1016/0168-9002(91)90821-7.

[218] N.G. Gavrilov, L.G. Isayeva, G.N. Kulipanov, V.N. Litvinenko, S.F. Mikhailov, V.M. Popik, I.G. Silvestrov, A.S. Sokolov, E.M. Trakhtenberg, N.A. Vinokurov, *et al.*, *Nucl. Instrum. Methods Phys. Res. A*. **282** (1989) 422–423, https://doi.org/10.1016/0168-9002(89)90016-8.

[219] V.V. Anashin, M.M. Brovin, N.A. Vinokurov, P.D. Vobly, A.A. Didenko, I.B. Drobyazko, A.S. Kalinin, G.N. Kulipanov, V.N. Litvinenko, L.A. Mironenko, S.F. Mikhailov, S.I. Mishnev, V.M. Popik, Y.A. Pupkov, I.G. Silvestrov, A.N. Skrinsky, A.S. Sokolov and E.M. Trakhtenberg, VEPP-3 dedicated straight section for OK operation. *Preprint INP* (1989) 89–126.

[220] G.N. Kulipanov, V.N. Litvinenko, A.S. Sokolov and N.A. Vinokurov, Metod izmerenie koeffitsentov otrajeniia blizkix k edinitse, method of measurement of coefficient of reflectivity close to unity. *Preprint INP* (1979) 79–24.

[221] G.N. Kulipanov, I.V. Pinayev, V.M. Popik, T.V. Shaftan, A.S. Sokolov and N.A. Vimokurov, *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 98–102, https://doi.org/10.1016/0168-9002(93)90022-a.

[222] V.N. Litvinenko, M.E. Couprie, N.G. Gavrilov, G.N. Kulipanov, I.V. Pinaev, V.M. Popik, A.N. Skrinsky and N.A. Vinokurov, *IEEE J. Quantum Electron*. **27** (1991) 2560–2565, https://doi.org/10.1109/3.104133.

[223] M.-E. Couprie, N.G. Gavrilov, G.N. Kulipanov, V.N. Litvinenko, I.V. Pinaev, V.M. Popik, A.N. Skrinsky and N.A. Vinokurov, *Nucl. Instrum. Methods Phys. Res. A*. **308** (1991) 39–44, https://doi.org/10.1016/0168-9002(91)90583-c.

[224] V.M. Popik and N.A. Vinokurov, The simple model of the supermodes in FEL with an intracvity etalon. *Preprint INP* (1993) 93–72.

[225] V.N. Litvinenko and N.A. Vinokurov, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 66–71, https://doi.org/10.1016/0168-9002(91)90822-8.

[226] M.E. Couprie, C. Bazin, M. Billardon and M. Velghe, *Nucl. Instrum. Methods Phys. Res. A*. **285** (1989) 31–42, https://doi.org/10.1016/0168-9002(89)90421-x.

[227] M.E. Couprie, M. Billardon, M. Velghe, C. Bazin, J.M. Ortéga, R. Prazeres and Y. Petroff, *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 13–19, https://doi.org/10.1016/0168-9002(90)91182-b.

[228] M.E. Couprie, M. Velghe, R. Prazeres, D. Jaroszynski and M. Billardon, *Phys. Rev. A*. **44** (1991) 1301, https://doi.org/10.1103/physreva.44.1301.

[229] M.E. Couprie, M. Billardon, M. Velghe, C. Bazin, M. Bergher, H. Fang, J.M. Ortéga, Y. Petroff and R. Prazeres, *Nucl. Instrum. Methods Phys. Res. A*. **272** (1988) 166–173, https://doi.org/10.1016/0168-9002(88)90217-3.

[230] M. Billardon, M.E. Couprie, J.M. Ortéga and M. Velghe, *Appl. Opt*. **30** (1991) 344–351, https://doi.org/10.1364/ao.30.000344.

[231] M.E. Couprie, D. Garzella and M. Billardon, *Europhys. Lett*. **21** (1993) 909, https://doi.org/10.1209/0295-5075/21/9/006.

[232] D. Nutarelli, D. Garzella, E. Renault, L. Nahon and M.E. Couprie, *Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 143–148, https://doi.org/10.1016/s0168-9002(00)00051-6.

[233] M.E. Couprie, M. Velghe, R. Prazeres and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 53–57, https://doi.org/10.1016/0168-9002(91)90819-c.

[234] M.E. Couprie, T. Hara, D. Gontier, P. Troussel, D. Garzella, A. Delboulbé and M. Billardon, *Phys. Rev. E*. **53** (1996) 1871–1889, https://doi.org/10.1103/physreve.53.1871.

[235] T. Hara, M.E. Couprie and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 67–70, https://doi.org/10.1016/0168-9002(95)01425-x.

[236] M.E. Couprie, D. Garzella, T. Hara, J.H. Codar and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 374–377, https://doi.org/10.1016/0168-9002(94)01576-7.

[237] M.E. Couprie, V. Litvinienko, D. Garzella, A. Delboulbé, M. Velghe and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 37–41, https://doi.org/10.1016/0168-9002(93)90009-7.

[238] M. Billardon, D. Garzella and M.E. Couprie, *Phys. Rev. Lett*. **69** (1992) 2368, https://doi.org/10.1103/physrevlett.69.2368.

[239] M.E. Couprie, M. Velghe, D. Jaroszynski and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 58–62, https://doi.org/10.1016/0168-9002(91)90820-g.

[240] G. De Ninno, M.E. Couprie, D. Nutarelli, D. Garzella, E. Renault and M. Billardon, *Phys. Rev. E*. **64** (2001) 026502, https://doi.org/10.1103/physreve.64.026502.

[241] R. Bartolini, G. Dattoli, L. Mezi, A. Renieri, M. Migliorati, M.E. Couprie, G. De Ninno and R. Roux, *Phys. Rev. Lett*. **87** (2001) 134801, https://doi.org/10.1103/physrevlett.87.134801.

[242] S. Bielawski, C. Bruni, G.L. Orlandi, D. Garzella and M.E. Couprie, *Phys. Rev. E*. **69** (2004) 045502, https://doi.org/10.1103/physreve.69.045502.

[243] S. Bielawski, C. Szwaj, C. Bruni, D. Garzella, G.L. Orlandi and M.E. Couprie, *Phys. Rev. Lett*. **95** (2005) 034801, https://doi.org/10.1103/physrevlett.95.034801.

[244] K. Yamada, T. Yamazaki, S. Sugiyama, T. Tomimasu, H. Ohgaki, T. Noguchi, T. Mikado, M. Chiwaki and R. Suzuki, *Nucl. Instrum. Methods Phys. Res. A*. **318** (1992) IN1–37, https://doi.org/10.1016/0168-9002(92)91019-6.

[245] T. Yamazaki, K. Yamada, S. Sugiyama, H. Ohgaki, N. Sei, T. Mikado, T. Noguchi, M. Chiwaki, R. Suzuki, M. Kawai, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 27–33, https://doi.org/10.1016/0168-9002(93)90008-6.

[246] H. Hama, J. Yamazaki and G. Isoyama, *Nucl. Instrum. Methods Phys. Res. A*. **341** (194) 12–16, https://doi.org/10.1016/0168-9002(94)90307-7.

[247] H. Hama, *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 57–61, https://doi.org/10.1016/0168-9002(95)01428-4.

[248] H. Hama, K. Kimura, J. Yamazaki, S. Takano, T. Kinoshita and M.-E. Couprie, *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 32–38, https://doi.org/10.1016/0168-9002(96)00103-9.

[249] H. Hama, J. Yamazaki, T. Kinoshita, K. Kimura and G. Isoyama, *Nucl. Instrum. Methods Phys.*

Res. A. **358** (1995) 365–368, https://doi.org/10.1016/0168-9002(94)01513-9.

[250] S. Koda, M. Hosaka, J. Yamazaki, M. Katoh and H. Hama, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 211–216, https://doi.org/10.1016/s0168-9002(01)01574-1.

[251] V.N. Litvinenko, B. Burnham, S.H. Park, Y. Wu, R. Cataldo, M. Emamian, J. Faircloth, S. Goetz, N. Hower, J.M.J. Madey, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **407** (1998) 8–15, https://doi.org/10.1016/b978-0-444-82978-8.50008-7.

[252] V.N. Litvinenko, S.H. Park, I.V. Pinayev, Y. Wu, M. Emamian, N. Hower, P. Morcombe, O. Oakeley, G. Swift and P. Wang, *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 151–158, https://doi.org/10.1016/s0168-9002(99)00096-0.

[253] V.N. Litvinenko, S.H. Park, I.V. Pinayev and Y. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 195–204, https://doi.org/10.1016/s0168-9002(01)01559-5.

[254] V.N. Litvinenko, B. Burnham, J.M.J. Madey and Y. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 369–373, https://doi.org/10.1016/0168-9002(94)01575-9.

[255] V.N. Litvinenko, B. Burnham, J.M.J. Madey and Y. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 334–337, https://doi.org/10.1016/0168-9002(94)01290-3.

[256] G. De Ninno, E. Allaria, M. Coreno, S. Chowdhury, F. Curbis, M.B. Danailov, B. Diviacco, M. Ferianis, E. Karantzoulis, E.C. Longhi, *et al*., *Phys. Rev. Lett*. **100** (2008) 104801, https://doi.org/10.1103/physrevlett.100.104801.

[257] V.N. Litvinenko, S.H. Park, I.V. Pinayev and Y. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 240–246, https://doi.org/10.1016/s0168-9002(01)01569-8.

[258] V.N. Litvinenko, S.H. Park, I.V. Pinayev and Y. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 65–73, https://doi.org/10.1016/s0168-9002(01)01536-4.

[259] V.N. Litvinenko, S.H. Park, I.V. Pinayev, Y. Wu, A. Lumpkin and B. Yang, *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 234–239, https://doi.org/10.1016/s0168-9002(01)01570-4.

[260] The DELTA Group Berges, D. Nölle, D. Garzella, A. Geisler, L. Gianessi, M. Hirsch, H. Quick, M. Ridder, T. Schmidt and K. Wille, *Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 128–133, https://doi.org/10.1016/s0168-9002(00)00044-9.

[261] R.P. Walker, B. Diviacco, C. Fava, A. Gambitta, M. Marsi, F. Mazzolini, M.E. Couprie, L. Nahon, D. Nutarelli, E. Renault, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 179–184, https://doi.org/10.1016/s0168-9002(99)00100-x.

[262] R.P. Walker, J.A. Clarke, M.E. Couprie, G. Dattoli, M. Eriksson, D. Garzella, L. Giannessi, M. Marsi, L. Nahon, D. Nölle, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 20–27, https://doi.org/10.1016/s0168-9002(01)01529-7.

[263] M. Marsi, M. Trovo, R.P. Walker, L. Giannessi, G. Dattoli, A. Gatto, N. Kaiser, S. Günster, D. Ristau, M.E. Couprie, D. Garzella, J.A. Clarke and M.W. Poole,190 nm *Appl. Phys. Lett*. **80** (2002) 2851–2853, https://doi.org/10.1063/1.1471934.

[264] M. Trovo, J.A. Clarke, M.E. Couprie, G. Dattoli, D. Garzella, A. Gatto, L. Giannessi, S. Günster, N. Kaiser, M. Marsi, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 157–161.

[265] D. Garzella, M.E. Couprie, T. Hara, L. Nahon, M. Brazuna, A. Delboulbé and M. Billardon, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 387–391, https://doi.org/10.1016/0168-9002(94)01594-5.

[266] A. Gatto, J. Heber, N. Kaiser, D. Ristau, S. Günster, J. Kohlhaas, M. Marsi, M. Trovo and R.P. Walker, *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 357–362, https://doi.org/10.1016/s0168-9002(02)00343-1.

[267] A. Gatto, T. Feigl, N. Kaiser, D. Garzella, G. De Ninno, M.E. Couprie, M. Marsi, M. Trovo, R. Walker, M. Grewe, *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 172–176, https://doi.org/10.1016/s0168-9002(02)00306-6.

[268] A. Gatto, R. Thielsch, N. Kaiser, M. Hirsch, D. Garzella, D. Nutarelli, G. De Ninno, E. Renault, M.-E. Couprie, P. Torchio, *et al.*, Toward resistant UV mirrors at 200 nm for free electron lasers: manufacture, characterizations, and degradation tests. In *International Symposium on Optical Science and Technology* (International Society for Optics and Photonics, 2000), pp. 261–275.

[269] G. De Ninno, M. Trovó, M. Danailov, M. Marsi, E. Karantzoulis, B. Diviacco, R.P. Walker, R. Bartolini, G. Dattoli, L. Giannessi, *et al.*, *Nucl. Instrum. Methods Phys. Res. A*. **507** (2003) 274–280, https://doi.org/10.1016/s0168-9002(03)00928-8.

[270] M.E. Couprie, D. Nutarelli and M. Billardon, *Nucl. Instrum. Methods Phys. Res. B*, **144** (1998) 66–74, https://doi.org/10.1016/s0168-583x(98)00297-3.

[271] S.V. Benson and J.M.J. Madey, *Phys. Rev. A*. **39** (1989) 1579, https://doi.org/10.1103/physreva.39.1579.

[272] D.J. Bamford and D.A.G. Deacon, *Phys. Rev. Lett*. **62** (1989) 1106, https://doi.org/10.1103/physrevlett.62.1106.

[273] E.B. Szarmes, S.V. Benson and J.M.J. Madey, *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 755–761, https://doi.org/10.1016/0168-9002(90)91302-r.

[274] E.B. Szarmes, A.D. Madden and J.M.J. Madey, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 220–223, https://doi.org/10.1016/0168-9002(94)01514-7.

[275] A. Bhowmik, M.S. Curtin, W.A. McMullin, S.V. Benson, J.M.J. Madey, B.A. Richman and L. Vintro, *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 20–24, https://doi.org/10.1016/0168-9002(90)91183-c.

[276] C.A. Brau, *Nucl. Instrum. Methods Phys. Res. A*. **318** (1992) 38–41, https://doi.org/10.1016/0168-9002(92)91020-a.

[277] C.A Brau and M.H. Mendenhall, *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) ABS4–ABS6, https://doi.org/10.1016/0168-9002(93)90165-e.

[278] P.W. Van Amersfoort, R.J. Bakker, J.B. Bekkers, R.W.B. Best, R. Van Buuren, P.F.M. Delmee, B. Faatz, C.A.J. Van der Geer, D.A. Jaroszynski, P. Manintveld, *et al*. *Nucl. Instrum. Methods Phys. Res. A*. **318** (1992) 42–46.

[279] D. Oepts, R.J. Bakker, D.A. Jaroszynski, A.F.G. van der Meer and P.W. van Amersfoort. *Phys. Rev. Lett*. **68** (1992) 3543, https://doi.org/10.1103/physrevlett.68.3543.

[280] D.A. Jaroszynski, R.J. Bakker, A.F.G. Van der Meer, D. Oepts and P.W. van Amersfoort. Experimental observation of limit-cycle oscillations in a short-pulse free-electron laser. *Phys. Rev. Lett*. **70** (1993) 3412, https://doi.org/10.1103/physrevlett.70.3412.

[281] D. Oepts, A.F.G. Van der Meer, R.J. Bakker and P.W. Van Amersfoort. *Phys. Rev. Lett*. **70** (1993) 3255, https://doi.org/10.1103/physrevlett.70.3255.

[282] M.E. Couprie, P. Bourgeois, P.J. Carlos, J.L. Fallou, P. Garganne, A. Leprêtre, J. Martin and J.M. Ortéga. *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 710–713, https://doi.org/10.1016/0168-9002(91)90962-p.

[283] R. Prazeres, J.M. Berset, F. Glotin, D. Jaroszynski and J.M. Ortéga. *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 15–19, https://doi.org/10.1016/0168-9002(93)90006-4.

[284] R. Prazeres, F. Glotin and J.M. Ortéga. *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 83–87, https://doi.org/10.1016/b978-0-444-51727-2.50025-0.

[285] R. Prazeres, F. Glotin, C. Insa, D.A. Jaroszynski and J.M. Ortéga. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics*. **3** (1998) 87–93, https://doi.org/10.1007/s100530050151.

[286] D.A. Jaroszynski, R. Prazeres, F. Glotin, J.M. Ortéga, D. Oepts, A.F.G. van der Meer, G.M.H. Knippels and P.W. van Amersfoort. *Phys. Rev. Lett*. **74** (1995) 2224, https://doi.org/10.1103/physrevlett.74.2224.

[287] F. Glotin, R. Chaput, D. Jaroszynski, R. Prazeres and J.-M. Ortéga. *Phys. Rev. Lett.* **71** (1993) 2587, https://doi.org/10.1103/physrevlett.71.2587.

[288] S. Joly, P. Balleyguier, M.A. Beuve, A. Binet, J.P. Bleses, R. Bois, C. Bonetti, J.P. de Brion, C. Couillaud, R. Dei-Cas, *et al. Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 199–203, https://doi.org/10.1016/0168-9002(93)90042-g.

[289] D. Iracane, V. Fontenay, P. Guimbal, S. Joly, S. Striby and D. Touati. *Phys. Rev. Lett.* **72** (1994) 3985, https://doi.org/10.1103/physrevlett.72.3985.

[290] P. Michel, F. Gabriel, E. Grosse, P. Evtuschenko, T. Dekorsy, M. Krenz, M. Helm, U. Lehnert, W. Seidel, R. Wünsch, *et al*. First lasing of the ELBE mid-IR FEL. In *Proceedings of the 2004 FEL Conference* (2004), Jacow, pp. 8–13.

[291] T. Tomimasu, E. Oshita, S. Okuma, K. Wakita, T. Takii, A. Koga, S. Nishihara, A. Nagai, H. Tongu, K. Wakisaka, *et al. μNucl. Instrum. Methods Phys. Res. A*. **383** (1996) 337–341, https://doi.org/10.1016/s0168-9002(96)00827-3.

[292] T. Tomimasu, K. Saeki, Y. Miyauchi, E. Oshita, S. Okuma, K. Wakita, A. Kobayashi, T. Suzuki, A. Zako, S. Nishihara, *et al. Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 626–631, https://doi.org/10.1016/0168-9002(96)00097-6.

[293] A. Zako, Y. Kanazawa, Y. Konishi, S. Yamaguchi, A. Nagai and T. Tomimasu.*Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 136–140, https://doi.org/10.1016/s0168-9002(99)00093-5.

[294] T.J. Orzechowski, B. Anderson, W.M. Fawley, D. Prosnitz, E.T. Scharlemann, S. Yarema, D. Hopkins, A.C. Paul, A.M. Sessler and J. Wurtele. *Phys. Rev. Lett.* **54** (1985) 889, https://doi.org/10.1103/physrevlett.54.889.

[295] H. Bottollier-Curtet, J. Gardelle, J. Bardy, C. Bonnafond, A. Devin, G. Germain, J. Labrouche, J. Launspach, P. Le Taillandier and J. de Mascureau. *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 197–199, https://doi.org/10.1016/0168-9002(91)90848-k.

[296] F. Ciocci, G. Dattoli, A. Doria, G.P. Gallerano, G. Schettini and A. Torre. *Phys. Rev. A*. **36** (1987) 207, https://doi.org/10.1103/physreva.36.207.

[297] F. Ciocci, A. Doria, G.P. Gallerano, I. Giabbai, M.F. Kimmitt, G. Messina, A. Renieri and J.E. Walsh. *Phys. Rev. Lett.* **66** (1991) 699, https://doi.org/10.1103/physrevlett.66.699.

[298] V.I. Alexeev, A.V. Alieva, K.A. Belovintsev, E.G. Bessonov, A.V. Serov and P.A. Cherenkov. Observations on emission generation in the free electron laser based on the microtron, Preprint FIAN 87-339. Technical Report, P. N. Institute Lebedev Physical Institute of the Academy of Science of USSR, Moscow (1987).

[299] V.I. Alexeev, E.V. Alieva, K.A. Belovintsev, E.G. Bessonov, A.V. Serov and P.A. Cherenkov. *Nucl. Instrum. Methods Phys. Res. A*. **282** (1989) 436–438, https://doi.org/10.1016/0168-9002(89)90019-3.

[300] Y.U. Jeong, B.C. Lee, S.K. Kim, S.O. Cho, B.H. Cha, J. Lee, G. M. Kazakevitch, P. D. Vobly, N.G. Gavrilov, V.V. Kubarev, *et al. Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 47–50, https://doi.org/10.1016/s0168-9002(01)01533-9.

[301] M. Tigner. *Il Nuovo Cim.* **37** (1965) 1228–1231, https://doi.org/10.1007/bf02773204.

[302] G. R Neil and L. Merminga. *Reviews of Modern Physics*. **74** (2002) 685, https://doi.org/10.1103/revmodphys.74.685.

[303] G.R. Neil, S. Benson, G. Biallas, C.L. Bohn, D. Douglas, H.F. Dylla, R. Evans, J. Fugitt, J. Gubeli, R. Hill, *et al. Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 192–196, https://doi.org/10.1016/s0168-9002(00)00064-4.

[304] R.W. Thomson, L.R. Short, R.D. McGinnis, W.B. Colson, M.D. Shinn, J.F. Gubeli, K.C. Jordan, R.A. Hill, G.H. Biallas, R.L. Walker, *et al. Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 625–629, https://doi.org/10.1016/s0168-9002(01)01580-7.

[305] G.R. Neil, C.L. Bohn, S.V. Benson, G. Biallas, D. Douglas, H.F. Dylla, R. Evans, J. Fugitt, A. Grippo, J. Gubeli, *et al*. *Phys. Rev. Lett*. **84** (2000) 662, https://doi.org/10.1103/physrevlett.84.662.

[306] A. Christodoulou, D. Lampiris, K. Polykandriotis, W.B. Colson, P.P. Crooker, S. Benson, J. Gubeli and G.R. Neil. *Phys. Rev. E*. **66** (2002) 056502, https://doi.org/10.1103/physreve.66.056502.

[307] C. Behre, S. Benson, G. Biallas, J. Boyce, C. Curtis, D. Douglas, H.F. Dylla, L. Dillon-Townes, R. Evans, A. Grippo, *et al*. *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 19–22, https://doi.org/10.1016/b978-0-444-51727-2.50012-2.

[308] G.R. Neil, C. Behre, S.V. Benson, M. Bevins, G. Biallas, J. Boyce, J. Coleman, L.A. Dillon-Townes, D. Douglas, H.F. Dylla, *et al*. *Nucl. Instrum. Methods Phys. Res. A*. **557** (2006) 9–15, https://doi.org/10.1016/j.nima.2005.10.047.

[309] E.J. Minehara, M. Sugimoto, M. Sawamura, R. Nagai, N. Kikuzawa, T. Yamanouchi and N. Nishimori. *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 9–11, https://doi.org/10.1016/s0168-9002(99)00050-9.

[310] R. Hajima, T. Shizuma, M. Sawamura, R. Nagai, N. Nishimori, N. Kikuzawa and E.J. Minehara. *Nucl. Instrum. Methods Phys. Res. A*. **507** (2003) 115–119, https://doi.org/10.1016/s0168-9002(03)00849-0.

[311] R. Hajima, N. Nishimori, R. Nagai and E.J. Minehara. *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 270–275, https://doi.org/10.1016/s0168-9002(01)01621-7.

[312] N. Nishimori, R. Hajima, R. Nagai and E.J. Minehara. *Phys. Rev. Lett*. **86** (2001) 5707, https://doi.org/10.1103/physrevlett.86.5707.

[313] N. Nishimori, R. Hajima, R. Nagai and E.J. Minehara. *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 266–269, https://doi.org/10.1016/s0168-9002(01)01638-2.

[314] N.G. Gavrilov, E.U. Gorniker, G.N. Kulipanov, I.V. Kupsov, I.V. Pinaev, I.K. Sedlyarov, A.N. Skrinsky, A.S. Sokolov, V.G. Veshcherevich, N.A. Vinokurov and P.D. Vobly. Project of a race-track microtron-recuperator for free electron laser, preprint INP 90-82. *Preprint INP* (1990) pp. 90–82.

[315] E.A. Antokhin, R.R. Akberdin, V.S. Arbuzov, M.A. Bokov, V.P. Bolotin, D.B. Burenkov, A.A. Bushuev, V.F. Veremeenko, N.A. Vinokurov, P.D. Vobly, *et al*. *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 15–18.

[316] B.C. Lee, S.K. Kim, Y.U. Jeong, S.O. Cho, B.H. Cha and J. Lee. *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 28–31, https://doi.org/10.1016/0168-9002(95)01502-7.

[317] B.C. Lee, Y.U. Jeong, S.H. Park, Y.G. Lim and S. Miginsky. *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 106–109, https://doi.org/10.1016/j.nima.2004.04.028.

[318] K.-J. Kim, Y. Shvyd'ko and S. Reiche. *Phys. Rev. Lett*. **100** (2008) 244802, https://doi.org/10.1103/physrevlett.100.244802.

[319] K.-J. Kim and Y.V. Shvyd'ko. *Phys. Rev. Special Topics-Accelerators and Beams*. **12** (2009) 030703, https://doi.org/10.1103/physrevstab.12.030703.

[320] R.R. Lindberg, K.-J Kim, Y. Shvyd'ko and W.M. Fawley. *Phys. Rev. Special Topics-Accelerators and Beams*. **14** (2011) 010701, https://doi.org/10.1103/physrevstab.14.010701.

[321] R. Prazeres, P. Guyot-Sionnest, J.M. Ortéga, D. Jaroszynski, M. Billardon, M.E. Couprie, M. Velghe and Y. Petroff. *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 72–76.

[322] R. Prazeres, P. Guyot-Sionnest, J.M. Ortéga, D. Jaroszynski, M. Billardon, M.E. Couprie, M. Velghe and Y. Petroff. *IEEE journal of Quantum Electronics*. **27** (1991) 1061–1068, https://doi.org/10.1109/3.83341.

[323] V.N. Litvinenko. *Nucl. Instrum. Methods Phys. Res. A*. **507** (2003) 265–273, https://doi.org/10.1016/s0168-9002(03)00927-6.

[324] G. D. Ninno, E. Allaria, M. Coreno, F. Curbis, M.B. Danailov, E. Karantzoulis, A. Lo-

catelli, T.O. Mentes, M.A. Nino, C. Spezzani, *et al*. *Phys. Rev. Lett*. **101** (2008) 053902, https://doi.org/10.1103/physrevlett.101.053902.

[325] M. Labat, M. Hosaka, A. Mochihashi, M. Shimada, M. Katoh, G. Lambert, T. Hara, Y. Takashima and M.E. Couprie. *Eur. Phys. J. D*. **44** (2007) 187, https://doi.org/10.1140/epjd/e2007-00177-6.

[326] M. Labat, M. Hosaka, M. Shimada, N. Yamamoto, M. Katoh and M.E. Couprie. *Phys. Rev. Lett*. **102** (2009) 014801, https://doi.org/10.1103/physrevlett.102.014801.

[327] M Labat, M Hosaka, M Shimada, M Katoh, and ME Couprie. Optimization of a seeded free-electron laser with helical undulators. *Phys. Rev. Lett*. **101** (2008) 164803.

[328] S. Thorin, M. Brandin, S. Werin, K. Goldammer and J. Bahrdt. *Phys. Rev. Special Topics-Accelerators and Beams*. **10** (2007) 110701, https://doi.org/10.1103/physrevstab.10.110701.

[329] N. Čutic', F. Lindau, S. Thorin, S. Werin, J. Bahrdt, W. Eberhardt, K. Holldack, C. Erny, A.L'Huillier and E. Mansten. *Phys. Rev. Special Topics-Accelerators and Beams*. **14** (2011) 030706, https://doi.org/10.1103/physrevstab.14.030706.

[330] R.W. Warren, L.C. Haynes, D.W. Feldman, W.E. Stein and S.J. Gitomer. *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 84–88.

[331] S. Benson and M. Shinn. In *Proceedings of FEL 2006, BESSY, Berlin, Germany*, Jacow, (2006) pp. 323–326.

[332] A.-M. Fauchet, J. Feinstein, A. Gover and R. Pantell. *IEEE Journal of Quantum Electronics*. **20** (1984) 1332–1341, https://doi.org/10.1109/jqe.1984.1072332.

[333] R.H. Pantell, J. Feinstein, A.S. Fisher, T.L. Deloney, M.B. Reid and W.M. Grossman. *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 312–315, https://doi.org/10.1016/0168-9002(86)90901-0.

[334] R.H. Pantell and M. Özcan. *Phys. Fluids B*. **2** (1990) 1311–1312, https://doi.org/10.1063/1.859547.

[335] J.A. Edighoffer, W.D. Kimura, R.H. Pantell, M.A. Piestrup and D.Y. Wang. *Phys. Rev. A*. **23** (1981) 1848, https://doi.org/10.1103/physreva.23.1848.

[336] J. Feinstein, A.S. Fisher, M.B. Reid, A. Ho, M. Özcan, H.D. Dulman and R.H. Pantell. *Phys. Rev. Lett*. **60** (1988) 18, https://doi.org/10.1103/physrevlett.60.18.

[337] R.H. Pantell, A.S. Fisher, J. Feinstein, A.H. Ho, M. Özcan, H.D. Dulman and M.B. Reid. *JOSA B*. **6** (1989) 1008–1014, https://doi.org/10.1364/josab.6.001008.

[338] M.B. Reid, A.S. Fisher, J. Feinstein, A.H. Ho, M. Özcan, H.D. Dulman, Y.J. Lee and R.H. Pantell. *Phys. Rev. Lett*. **62** (1989) 249, https://doi.org/10.1103/physrevlett.62.249.

[339] S.J. Smith and E.M. Purcell. *Phys. Rev* **92** (1953) 1069, https://doi.org/10.1103/physrev.92.1069.

[340] M. Piestrup. *IEEE Journal of Quantum Electronics*. **24** (1988) 591–597, https://doi.org/10.1109/3.167.

[341] G. Dattoli, A. Doria, G.P. Gallerano, A. Renieri, G. Schettini and A. Torre. *Il Nuovo Cim B*. **101** (1988) 79–84, https://doi.org/10.1007/bf02828071.

[342] J.E. Walsh. Free electron laser utilizing grating coupling, November 16 1993. US Patent 5,263,043.

[343] J. Walsh, K. Woods and S. Yeager. *Nucl. Instrum. Methods Phys. Res. A*. **341** (1994) 277–279, https://doi.org/10.1016/0168-9002(94)90364-6.

[344] J.E. Walsh. *Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 214–221, https://doi.org/10.1016/s0168-9002(00)00069-3.

[345] H.L. Andrews and C.A. Brau. *Phys. Rev. Special Topics-Accelerators and Beams*. **7** (2004) 070701, https://doi.org/10.1103/physrevstab.7.070701.

[346] G. Doucas, J.H. Mulvey, M. Omori, J. Walsh and M.F. Kimmitt. *Phys. Rev. Lett*. **69** (1992) 1761, https://doi.org/10.1103/physrevlett.69.1761.

[347] A. Bakhtyari, J.E. Walsh and J.H. Brownell. *Phys. Rev. E*. **65** (2002) 066503, https://doi.org/10.1103/physreve.65.066503.

[348] J. Urata, M. Goldstein, M.F. Kimmitt, A. Naumov, C. Platt and J.E. Walsh. *Phys. Rev. Lett*. **80** (1998) 516, https://doi.org/10.1103/physrevlett.80.516.

[349] T.I. Smith and H.A. Schwettman. *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 812–821, https://doi.org/10.1016/0168-9002(91)90982-v.

[350] M.-E. Couprie and J.-M. Ortéga. *Analusis*. **28** (2000) 725–736.

[351] T.I. Smith, J.C. Frisch, R. Rohatgi, H.A. Schwettman and R.L. Swent. *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 33–36, https://doi.org/10.1016/0168-9002(90)91185-e.

[352] D. Zimdars, A. Tokmakoff, S. Chen, S.R. Greenfield, M.D. Fayer, T.I. Smith and H.A. Schwettman. *Phys. Rev. Lett*. **70** (1993) 2718, https://doi.org/10.1103/physrevlett.70.2718.

[353] S.R. Greenfield, Y.S. Bai and M.D. Fayer. *Chemical Physics Letters*. **170** (1990) 133–138, https://doi.org/10.1016/0009-2614(90)87104-y.

[354] T.I. Smith, Y.S. Bai, S.R. Greenfield, J.C. Frisch, R.L. Swent, M.D. Fayer and H.A. Schwettman. *JOSA B*. **8** (1991) 1652–1656, https://doi.org/10.1364/josab.8.001652.

[355] W.S. Fann, S.V. Benson, J.M.J. Madey, S. Etemad, G.L. Baker, L. Rothberg, M. Roberson and R.H. Austin. *Nucl. Instrum. Methods Phys. Res. A***296** (1990) 804–808, https://doi.org/10.1109/fel.1989.716112.

[356] G.S. Edwards, R.H. Austin, F.E. Carroll, M.L. Copeland, M.E. Couprie, W.E. Gabella, R.F. Haglund, B.A. Hooper, M.S. Hutson, E.D. Jansen, *et al*. *Rev. Sci. Instrum*. **74** (2003) 3207–3245.

[357] C. Crépin, M. Broquier, H. Dubost, D. L'hermite, A. Tramer, J.P. Galaup, J.M. Berset and J.M. Ortéga. *Laser Chemistry*. **19** (1999) 65–69, https://doi.org/10.1155/1999/95804.

[358] J. Lemaire, P. Boissel, M. Heninger, G. Mauclaire, G. Bellec, H. Mestdagh, A. Simon, S. Le Caer, J.M. Ortega, F. Glotin, *et al*. *Phys. Rev. Lett*. **89** (2002) 273002, https://doi.org/10.1103/physrevlett.89.273002.

[359] A. Dazzi, R. Prazeres, F. Glotin and J.M. Ortéga. *Optics Letters*. **30** (2005) 2388–2390, https://doi.org/10.1364/ol.30.002388.

[360] D. Oepts, A.F.G. Van der Meer and P.W. Van Amersfoort. *Infrared Physics & Technology*. **36** (1995) 297–308, https://doi.org/10.1016/1350-4495(94)00074-u.

[361] T. Tomimasu, T. Takii, T. Suzuki, E. Nishimura, S. Ogino, A. Nagai and M. Yasumoto. *Nucl. Instrum. Methods Phys. Res. A*. **407** (1998) 494–499, https://doi.org/10.1016/s0168-9002(98)00075-8.

[362] U. Lehnert, P. Michel, W. Seidel, G. Staats, J. Teichert and R. Wünsch. First experiences with the FIR-FEL at ELBE. *Proceeding of FEL, Novosibirsk*, Jacow, 2007.

[363] W. Seidel, E. Cizmar, O. Drachenko, M. Helm, M. Justus, U. Lehnert, P. Michel, M. Ozerov, H. Schneider, R. Schurig, *et al*. Three years of cw-operation at FELBE–experiences and applications. In *Proceedings 30th International Free Electron Laser Conference FEL 2008*, Jacow, Citeseer, 2008.

[364] A. Xie, L. van der Meer, W. Hoff and R.H. Austin. *Phys. Rev. Lett*. **84** (2000) 5435, https://doi.org/10.1103/physrevlett.84.5435.

[365] Z. Zhu and M.J. Kelley. *Applied Surface Science*. **252** (2006) 6619–6623, https://doi.org/10.1016/j.apsusc.2006.02.088.

[366] G.R. Neil and Jefferson Lab FEL Team. *Nucl. Instrum. Methods Phys. Res. B*. **144** (1998) 40–49, https://doi.org/10.1016/s0168-583x(98)00295-x.

[367] B.A. Knyazev, G.N. Kulipanov and N.A. Vinokurov. *Measurement Science and Technology*. **21** (2010) 054017, https://doi.org/10.1088/0957-0233/21/5/054017.

[368] Y.U. Jeong, G.M. Kazakevitch, H.J. Cha, S.H. Park and B.C. Lee. *Nucl. Instrum. Methods Phys.*

Res. A. **575** (2007) 58–62, https://doi.org/10.1016/j.nima.2007.01.024.

[369] L. Nahon, E. Renault, M.E. Couprie, F. Mérola, P. Dumas, M. Marsi, A. Taleb-Ibrahimi, D. Nutarelli, R. Roux and M. Billardon. *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 489–496, https://doi.org/10.1016/s0168-9002(99)00153-9.

[370] M.E. Couprie, F. Merola, P. Tauc, D. Garzella, A. Delboulbé, T. Hara and M. Billardon. *Rev. Sci. Instrum*. **65** (1994) 1485–1495, https://doi.org/10.1063/1.1144880.

[371] L. Nahon, E. Renault, M.-E. Couprie, D. Nutarelli, D. Garzella, M. Billardon, G.L. Carr, G.P. Williams and P. Dumas. Two-color experiments combining the UV storage ring free-electron laser and the SA5 IR beamline at Super-ACO. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation* (International Society for Optics and Photonics, 1999), pp. 145–154.

[372] M. Marsi, M.E. Couprie, L. Nahon, D. Garzella, T. Hara, R. Bakker, M. Billardon, A. Delboulbe, G. Indlekofer and A. Taleb-Ibrahimi. *Appl. Phys. Lett*. **70** (1997) 895–897, https://doi.org/10.1063/1.118307.

[373] M. Marsi, R. Belkhou, C. Grupp, G. Panaccione, A. Taleb-Ibrahimi, L. Nahon, D. Garzella, D. Nutarelli, E. Renault, R. Roux, *et al*. $SiO_2/SiPhys.$ *Rev. B*. **61** (2000) R5070, https://doi.org/10.1103/physrevb.61.r5070.

[374] H. Ade, W. Yang, S.L. English, J. Hartman, R.F. Davis, R.J. Nemanich, V.N. Litvinenko, I.V. Pinayev, Y. Wu and J.M.J. Madey. *Surface Review and Letters*. **5** (1998) 1257–1268, https://doi.org/10.1142/s0218625x98001596.

[375] M. Hosaka, S. Koda, M. Katoh, J. Yamazaki, K. Hayashi, K. Takashima, T. Gejo and H. Hama. *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002):146–151, https://doi.org/10.1016/s0168-9002(02)00301-7.

[376] T. Gejo, E. Shigemasa, E. Nakamura, M. Hosaka, S. Koda, A. Mochihashi, M. Katoh, J.-I. Yamazaki, K. Hayashi, Y. Takashima, *et al*. *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 627–631, https://doi.org/10.1016/j.nima.2004.04.116.

[377] M. Marsi, A. Locatelli, M. Trovo, R.P. Walker, M.E. Couprie, D. Garzella, L. Nahon, D. Nutarelli, E. Renault, A. Gatto, *et al*. *Surf. Rev. Lett*. **9** (2002) 599–607, https://doi.org/10.1142/s0218625x02001768.

[378] F. Glotin, J.-M. Ortéga, R. Prazeres, G. Devanz and O. Marcouillé. *Phys. Rev. Lett*. **77** (1996) 3130, https://doi.org/10.1103/physrevlett.77.3130.

[379] M. Hosaka, H. Hama, K. Kimura, J. Yamazaki and T. Kinoshita. *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 525–529.

[380] M.E. Couprie, D. Nutarelli, R. Roux, B. Visentin, L. Nahon, R. Bakker, A. Delboulbé and M. Billardon. *Journal of Physics B*. **32** (1999) 5657, https://doi.org/10.1088/0953-4075/32/24/303.

[381] D. Nutarelli, M.E. Couprie, L. Nahon, R. Bakker, A. Delboulbé, R. Roux, B. Visentin and M. Billardon. *Nucl. Instrum. Methods Phys. Res. A*. **407** (1998) 459–463, https://doi.org/10.1016/s0168-9002(98)00068-0.

[382] V.N. Litvinenko, B. Burnham, M. Emamian, N. Hower, J.M.J. Madey, P. Morcombe, P.G. O'Shea, S.H. Park, R. Sachtschale, K.D. Straub, *et al*. *Phys. Rev. Lett*. **78** (1997) 4569.

[383] S.H. Park, V.N. Litvinenko, W. Tornow and C. Montgomery. *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 425–431, https://doi.org/10.1016/s0168-9002(01)01635-7.

[384] V.N. Litvinenko and J.M.J. Madey. High-power inverse Compton γ-ray source at the Duke storage ring. In *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation* (International Society for Optics and Photonics, 1995), pp. 55–77.

[385] H.R. Weller and M.W. Ahmed. *Modern Phys. Lett. A*. **18** (2003) 1569–1590, https://doi.org/10.1142/s0217732303011216.

[386] J.M.J. Madey and C. Pellegrini. Free electron generation of extreme ultraviolet coherent radiation-1983. Technical Report, American Institute of Physics, New York (1984).

[387] P. Sprangle and R.A. Smith. *Phys. Rev. A*. **21** (1980) 293.

[388] N.M. Kroll, P.L. Morton and M.N. Rosenbluth. *Phys. Quantum Electron.* **7** 147.

[389] W.A. McMullin and G. Bekefi. *Phys. Rev. A*. **25** (1982) 1826, https://doi.org/10.1103/physreva.25.1826.

[390] E. Jerby and A. Gover. *IEEE J. Quantum Electron*. **21** (1985) 1041–1058, https://doi.org/10.1109/jqe.1985.1072748.

[391] P. Sprangle, C.M. Tang and C.W. Roberson. *Nucl. Instrum. Methods Phys. Res. A*. **239** (1985) 1–18.

[392] W.B. Colson, J.C. Gallardo and P.M. Bosco. *Phys. Rev. A*. **34** (1986) 4875, https://doi.org/10.1103/physreva.34.4875.

[393] Y.S. Derbenev, A.M. Kondratenko and E.L. Saldin. *Nucl. Instrum. Methods Phys. Res*. **193** (1982) 415–421, https://doi.org/10.1016/0029-554x(82)90233-6.

[394] R. Bonifacio, F. Casagrande and L.A. Lugiato. *Opt. Commun*. **36** (1981) 159–163, https://doi.org/10.1016/0030-4018(81)90160-7.

[395] R. Bonifacio, F. Casagrande and G. Casati. *Opt. Commun*. **40** (1982) 219–223, https://doi.org/10.1016/0030-4018(82)90265-6.

[396] R. Bonifacio, C. Pellegrini and L.M. Narducci. *Opt. Commun*. **50** (1984) 373–378, https://doi.org/10.1016/0030-4018(84)90105-6.

[397] W.B. Colson. *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 82–85.

[398] C. Pellegrini. *Nucl. Instrum. Methods Phys. Res. A*. **475** (2001) 1–12, https://doi.org/10.1016/s0168-9002(01)01527-3.

[399] J.-M. Wang and L.-H. Yu. *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 484–489, https://doi.org/10.1016/0168-9002(86)90928-9.

[400] J.B. Murphy, C. Pellegrini and R. Bonifacio. *Opt. Commun*. **53** (1985) 197–202, https://doi.org/10.1016/0030-4018(85)90331-1.

[401] R. Bonifacio, F. Casagrande and C. Pellegrini. *Opt. Commun*. **61** (1987) 55–60, https://doi.org/10.1016/0030-4018(87)90124-6.

[402] R. Bonifacio, F. Casagrande, G. Cerchioni, L. de S. Souza, P. Pierini and N. Piovella. *J. Quantum Electron*. **17** (1990) 1417–1427.

[403] G. Dattoli and P.L. Ottaviani. *Opt. Commun*. **204** (2002) 283–297, https://doi.org/10.1016/s0030-4018(02)01201-4.

[404] G. Dattoli, L. Giannessi, P.L. Ottaviani and C. Ronsivalle. *J. Appl. Phys*. **95** (2004) 3206–3210, https://doi.org/10.1063/1.1645979.

[405] G. Dattoli and P.L. Ottaviani. *Opt. Commun*. **254** (2005) 330–339, https://doi.org/10.1016/j.optcom.2005.05.056.

[406] Y. Ding, A. Brachmann, F.-J. Decker, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, A. Miahnahri, H.-D. Nuhn, D. Ratner, J. Turner, J. Welch, W. White and J. Wu. *Phys. Rev. Lett*. **102** (2009) 254801–254805, https://doi.org/10.1103/physrevlett.102.254801.

[407] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov. *Opt. Commun*. **148** (1998) 383–403, https://doi.org/10.1016/s0030-4018(97)00670-6.

[408] R. Bonifacio, L. De Salvo, P. Pierini, N. Piovella and C. Pellegrini. *Phys. Rev. Lett*. **73** (1994) 70, https://doi.org/10.1103/physrevlett.73.70.

[409] E.T. Scharlemann, A.M. Sessler and J.S. Wurtele. *Phys. Rev. Lett*. **54** (1985) 1925.

[410] G.T. Moore. *Opt. Commun*. **52** (1984) 46–51, https://doi.org/10.1016/0030-4018(84)90071-3.

[411] G.T. Moore. *Opt. Commun*. **54** (1985) 121–125, https://doi.org/10.1016/0030-4018(85)90014-8.

[412] E.T. Scharlemann, A.M. Sessler and J.S. Wurtele, *Nucl. Instrum. Methods Phys. Res. A*. **239** (1985) 29–35, https://doi.org/10.1016/0168-9002(85)90694-1

[413] A. Fruchtman, *Phys. Rev. A*. **37** (1988) 2989, https://doi.org/10.1103/physreva.37.2989

[414] K.-J. Kim. *Phys. Rev. Lett*. **57** (1986) 1871–1874, https://doi.org/10.1103/physrevlett.57.1871

[415] Kwang-Je Kim, *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 396–403, https://doi.org/10.1016/0168-9002(86)90916-2

[416] M. Xie, *Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 59–66, https://doi.org/10.1016/s0168-9002(00)00114-5

[417] Ming Xie, Design optimization for an X-ray free electron laser driven by SLAC linac. In *Particle Accelerator Conference, 1995., Proc. of the 1995*, vol. 1 (IEEE, 1995), pp. 183–185.

[418] C. Pellegrini, A. Marinelli and S. Reiche. *Rev. Mod. Phys*. **88** (2016) 015006, https://doi.org/10.1103/revmodphys.88.015006

[419] W.M. Fawley, A user manual for GINGER and its post-processor XPLOTGIN. *Lawrence Berkeley National Laboratory* (2002).

[420] S. Reiche, *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 243–248, https://doi.org/10.1016/s0168-9002(99)00114-x

[421] L. Giannessi, Overview of Perseo, a system for simulating FEL dynamics in Mathcad. In *Proc. of the Free-Electron Laser Conference* Jacow (2006).

[422] S.G. Biedron, Y.C. Chae, R.J. Dejus, B. Faatz, H.P. Freund, S.V. Milton, H.-D. Nuhn and S. Reiche, *Nucl. Instrum. Methods Phys. Res. A*. **445** (2000) 110–115, https://doi.org/10.1016/s0168-9002(00)00124-8

[423] L.T. Campbell and B.W.J. McNeil, *Phys of Plasmas*. **19** (2012) 093119, https://doi.org/10.1063/1.4752743

[424] T.M. Tran and J.S. Wurtele, *Comput. Phys. Commun*. **54** (1989) 263–272, https://doi.org/10.1016/0010-4655(89)90090-8

[425] Z. Huang and K.-J. Kim, *Phys. Rev. E*. **62** (2000) 7295, https://doi.org/10.1103/physreve.62.7295

[426] G. Dattoli, P.L. Ottaviani and S. Pagnutti, *J. Appl. Phys*. **97** (2005) 113102, https://doi.org/10.1063/1.1886890

[427] G. Dattoli, L Giannessi, P.L. Ottaviani, H.P. Freund, S.G. Biedron and S. Milton, *Nucl. Instrum. Methods Phys. Res. A*. **495** (2002) 48–57, https://doi.org/10.1016/s0168-9002(02)01281-0

[428] I. Ben-Zvi, L.F. Di Mauro, S. Krinsky, M.G. White and L.H. Yu, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 181–186, https://doi.org/10.1016/0168-9002(91)90845-h

[429] L.-H. Yu, M. Babzien, I. Ben-Zvi, L.F. DiMauro, A. Doyuran, W. Graves, E. Johnson, S. Krinsky, R. Malone, I. Pogorelsky, J. Skaritka, G. Rakowsky, L. Solomon, X.J. Wang, M. Woodle, V. Yakimenko, S.G. Biedron, J.N. Galayda, E. Gluskin, J. Jagger, V. Sajaev and I. Vasserman, *Science* **289** (2000) 932–934.

[430] L.H. Yu, *Phys. Rev. A*. **44** (1991) 5178, https://doi.org/10.1103/physreva.44.5178

[431] L.H. Yu and J. Wu, *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 493–498, https://doi.org/10.1016/s0168-9002(02)00368-6

[432] E. Allaria and G. De Ninno, *Phys. Rev. Lett*. **99** (2007) 014801, https://doi.org/10.1103/physrevlett.99.014801

[433] I. Ben-Zvi, K.M. Yang and L.H. Yu, *Nucl. Instrum. Methods Phys. Res. A*. **318** (1992) 726–729, https://doi.org/10.1016/0168-9002(92)91147-2

[434] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *Opt. Commun*. **202** (2002) 169–187,

https://doi.org/10.1016/s0030-4018(02)01091-x
[435] A. Gover and E. Dyunin, *Phys. Rev. Lett*. **102** (2009) 154801, https://doi.org/10.1103/physrevlett.102.154801
[436] L. Giannessi, P. Musumeci and S. Spampinati, *J. Appl. Phys*. **98** (2005) 043110, https://doi.org/10.1063/1.2010624
[437] L. Giannessi and P. Musumeci, *New J. Phys*. **8** (2006) 294, https://doi.org/10.1088/1367-2630/8/11/294
[438] M. Labat, N. Joly, S. Bielawski, C. Szwaj, C. Bruni and M.E. Couprie *Phys. Rev. Lett*. **103** (2009) 264801, https://doi.org/10.1103/physrevlett.103.264801
[439] G. Stupakov, *Phys. Rev. Lett*. **102** (2009) 074801–074804, https://doi.org/10.1103/physrevlett.102.074801
[440] M.E. Couprie, M. Labat, C. Evain, C. Szwaj, S. Bielawski, N. Hubert, C. Benabderrahmane, F. Briquez, L. Chapuis, F. Marteau *et al*., *J. Mod. Opt*. **63** (2016) 309–323, https://doi.org/10.1080/09500340.2015.1075617
[441] D. Xiang and G. Stupakov, *New J. Phys*. **13** (2011) 093028–093036, https://doi.org/10.1088/1367-2630/13/9/093028
[442] C. Evain, A. Loulergue, A. Nadji, J.M. Filhol, M.E. Couprie and A.A. Zholents, *New J. Phys*. **14** (2012) 023003, https://doi.org/10.1088/1367-2630/14/2/023003
[443] J.B. Murphy and C. Pellegrini, *JOSA B*. **2** (1985) 259–264, https://doi.org/10.1364/josab.2.000259
[444] J.B. Murphy and C. Pellegrini, *Nucl. Instrum. Methods Phys. Res. A*. **237** (1985) 159–167, https://doi.org/10.1016/0168-9002(85)90344-4
[445] M. Cornacchia, J. Bisognano, S. Chattopadhyay, A. Garren, K. Halbach, A. Jackson, K.J. Kim, H. Lancaster, J. Peterson, M.S. Zisman *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 57–63, https://doi.org/10.1016/0168-9002(86)90860-0
[446] C. Travier, *Nucl. Instrum. Methods Phys. Res. A*. **304** (1991) 285–296, https://doi.org/10.1016/0168-9002(91)90870-v
[447] C. Travier, *Nucl. Instrum. Methods Phys. Res. A*. **340** (1994) 26–39, https://doi.org/10.1016/0168-9002(94)91278-5
[448] M.Curtin, G. Bennett, R. Burke, A. Bhowmik, P. Metty, S. Benson and J.M.J. Madey, *Nucl. Instrum. Methods Phys. Res. A*. **296** (1990) 127–133, https://doi.org/10.1016/0168-9002(90)91199-l
[449] J.S. Fraser, R.L. Sheffield and E.R. Gray, *Nucl. Instrum. Methods Phys. Res. A*. **250** (1986) 71–76, https://doi.org/10.1016/0168-9002(86)90862-4
[450] R.L. Sheffield, E.R. Gray and J.S. Fraser, *Nucl. Instrum. Methods Phys. Res. A*. **272** (1988) 222–226, https://doi.org/10.1016/0168-9002(88)90227-6
[451] P.G. O'Shea, S.C. Bender, B.E. Carlsten, J.W. Early, D.W. Feldman, R.B. Feldman, W.J.D. Johnson, A.H. Lumpkin, R.L. Sheffield, R.W. Springer *et al*. Performance of the photoinjector accelerator for the Los Alamos free-electron laser. In *Proc. IEEE Part. Accel. Conf., San Francisco, CA* (1991).
[452] D.W. Feldman, S.C. Bender, B.E. Carlsten, J. Early, R.B. Feldman, W.J.D. Johnson, A.H Lumpkin, P.G. O'Shea, W.E. Stein, R.L. Sheffield *et al*., *IEEE J. Quantum Electron*. **27** (1991) 2636–2643, https://doi.org/10.1109/3.104143
[453] B.E. Carlsten, *Nucl. Instrum. Methods Phys. Res. A*. **285** (1989) 313–319, https://doi.org/10.1016/0168-9002(89)90472-5
[454] K.T. McDonald, *IEEE Trans. Electron Devices*. **35** (1988) 2052–2059, https://doi.org/10.1109/16.7427
[455] S.C. Hartman, N. Barov, C. Pellegrini, S. Park, J. Rosenzweig, G. Travish, R. Zhang, C. Clayton, P. Davis, M. Everett *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **340** (1994) 219–230,

https://doi.org/10.1016/0168-9002(94)91305-6

[456] D.T. Palmer, X.J. Wang, R.H. Miller, M. Babzien, I. Ben-Zvi, C. Pellegrini, J. Sheehan, J. Skaritka, H. Winick, M. Woodle *et al.*, Emittance studies of the BNL/SLAC/UCLA 1.6 cell photocathode RF gun. In *Particle Accelerator Conference, 1997. Proc. of the 1997*, vol. 3 (IEEE, 1997), 2687–2689.

[457] C. Travier, M. Bernard, P. Dufresne, G. Michaud, M. Omeich, M. Roch and J. Kodier, *Nucl. Instrum. Methods Phys. Res. B*. **89** (1994) 27–32, https://doi.org/10.1016/0168-583x(94)95139-x

[458] C. Pellegrini, *Nucl. Instrum. Methods Phys. Res. A*. **272** (1988) 364–367, https://doi.org/10.1016/0168-9002(88)90252-5

[459] J. C. Gallardo. 1 ÅIn Proc. of the Workshop: Prospects for a 1 Å Free Electron Laser, Sag Harbor, New York, Center for Accelerator Physics, April 22-27, Ed, J. C. Gallardo (1990), pp. iii. Center for Accelerator Physics, Brookhaven National Laboratory, BNL–52273, DE91 007631.

[460] L.H. Yu. 1 ÅIn Proc. of the Workshop: Prospects for a 1 Å Free Electron Laser, Sag Harbor, New York, April 22-27, Ed, J. C. Gallardo (1990), pp. 13–30. Center for Accelerator Physics, Brookhaven National Laboratory, BNL–52273, DE91 007631.

[461] C. Pellegrini, SASE and Development of an X-Ray FEL. In Proc. of the Workshop: Prospects for a 1 Å Free Electron Laser, Sag Harbor, New York, April 22–27, Ed. J. C. Gallardo (1990), pp. 3–12. Center for Accelerator Physics, Brookhaven National Laboratory, BNL–52273, DE91 007631.

[462] J.C. Goldstein, C.J. Elliot and M.J. Schmitt, Three-Dimensional Simulations of the Generation of One-Angstrom Radiation by a Self-Amplified Spontaneous Emission Free-Electron Laser. In *Proc. of the Workshop: Prospects for a* 1 Å *Free Electron Laser, Sag Harbor, New York, April 22-27*, Ed. J. C. Gallardo (1990), pp. 13–30. Center for Accelerator Physics, Brookhaven National Laboratory, BNL–52273, DE91 007631.

[463] M. Cornacchia and H. Winick, Proc. of a Workshop on IV Generation Light Sources. Technical report, SSRL/SLAC Rep. 92/02 (1992).

[464] J.L. Laclare. Proc. of the 10th ICFA Beam Dynamics Workshop on 4th Generation Light Source. In *Contribution booklet,* Ed., J. L. Laclare (ESRF, 1996).

[465] T.J. Orzechowski, B.R. Anderson, J.C. Clark, W.M. Fawley, A.C. Paul, D. Prosnitz, E.T. Scharlemann, S.M. Yarema, D.B. Hopkins, A.M. Sessler and J.S. Wurtele, *Phys. Rev. Lett*. **57** (1986) 2172, https://doi.org/10.1103/physrevlett.57.2172

[466] D.A. Kirkpatrick, G. Bekefi, A.C. Dirienzo, H.P. Freund, and A.K. Ganguly, *Nucl. Instrum. Methods Phys. Res. A*. **285** (1989) 43–46, https://doi.org/10.1016/0168-9002(89)90422-1

[467] J. Gardelle, J. Labrouche and J.L. Rullier, *Phys. Rev. Lett*. **76** (1996) 4532, https://doi.org/10.1103/physrevlett.76.4532

[468] J. Gardelle, T. Lefevre, G. Marchese, J.L. Rullier and J.T. Donohue, *Phys. Rev. Lett*. **79** (1997) 3905, https://doi.org/10.1103/physrevlett.79.3905

[469] T. Lefevre, J. Gardelle, G. Marchese, J.L. Rullier and J.T. Donohue, *Phys. Rev. Lett*. **82** (1999) 323, https://doi.org/10.1103/physrevlett.82.323

[470] S. Okuda, J. Ohkuma, N. Kimura, Y. Honda, T. Okada, S. Takamuku, T. Yamamoto and K. Tsumori, *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 76–78, https://doi.org/10.1016/0168-9002(93)90017-c

[471] D. Bocek, P. Kung, H.-C. Lihn, C. Settakorn and H. Wiedemann, *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 13–16, https://doi.org/10.1016/0168-9002(95)01361-x

[472] R. Prazeres, J.M. Ortéga, F. Glotin, D.A. Jaroszynski and O. Marcouillé, *Phys. Rev. Lett*. **78** (1997) 2124, https://doi.org/10.1103/physrevlett.78.2124

[473] M. Babzien, I. Ben-Zvi, P. Catravas, J.-M. Fang, T.C. Marshall, X.J. Wang, J.S. Wurtele, V. Yaki-

menko and L.H. Yu, *Phys. Rev. E*. **57** (1998) 6093–6100, https://doi.org/10.1103/physreve.57.6093

[474] M. Hogan, C. Pellegrini, J. Rosenzweig, G. Travish, A. Varfolomeev, S. Anderson, K. Bishofberger, P. Frigola, A. Murokh, N. Osmanov *et al*., *Phys. Rev. Lett*. **80** (1998) 289, https://doi.org/10.1103/physrevlett.80.289

[475] C.M. Fortgang, *Nucl. Instrum. Methods Phys. Res. A*. **393** (1997) 385–388, https://doi.org/10.1016/s0168-9002(97)00519-6

[476] D.C. Nguyen, R.L. Sheffield, C.M. Fortgang, J.C. Goldstein, J.M. Kinross-Wright and N.A. Ebrahim, *Phys. Rev. Lett*. **81** (1998) 810, https://doi.org/10.1103/physrevlett.81.810

[477] M.J. Hogan, C. Pellegrini, J. Rosenzweig, S. Anderson, P. Frigola, A. Tremaine, C. Fortgang, D.C. Nguyen, R.L. Sheffield, J. Kinross-Wright *et al*., *Phys. Rev. Lett*. **81** (1998) 4867, https://doi.org/10.1103/physrevlett.81.4867.

[478] A. Tremaine, J.B. Rosenzweig, S. Anderson, P. Frigola, M. Hogan, A. Murokh, C. Pellegrini, D.C. Nguyen and R.L. Sheffield, *Phys. Rev. Lett*. **81** (1998) 5816, https://doi.org/10.1103/physrevlett.81.5816.

[479] S.V. Milton, E. Gluskin, S.G. Biedron, R.J. Dejus, P.K. Den Hartog, J.N. Galayda, K.-J. Kim, J.W. Lewellen, E.R. Moog, V. Sajaev *et al*., *Phys. Rev. Lett*. **85** (2000) 988, https://doi.org/10.1103/physrevlett.85.988.

[480] S.V. Milton, E. Gluskin, N.D. Arnold, C. Benson, W. Berg, S.G. Biedron, M. Borland, Y.-C. Chae, R.J. Dejus, P.K. Den Hartog, B. Deriy, M. Erdmann, Y.I. Eidelman, M.W. Hahne, Z. Huang, K.-J. Kim, J.W. Lewellen, Y. Li, A.H. Lumpkin, O. Makarov, E.R. Moog, A. Massiri, V. Sajaev, R. Soliday, B.J. Tieman, E.M. Trakhtenberg, G. Travish, I.B. Vasserman, N.A. Vinokurov, X.J. Wang, G. Wiemerslage and B.X. Yang, *Science*. **292** (2001) 2037–2041, https://doi.org/10.1126/science.1059955.

[481] A. Tremaine, X.J. Wang, M. Babzien, I. Ben-Zvi, M. Cornacchia, H.-D. Nuhn, R. Malone, A. Murokh, C. Pellegrini, S. Reiche *et al*., *Rev. Lett*. **88** (2002) 204801, https://doi.org/10.1103/physrevlett.88.204801.

[482] A. Murokh, R. Agustsson, M. Babzien, I. Ben-Zvi, L. Bertolini, K. van Bibber, R. Carr, M. Cornacchia, P. Frigola, J. Hill *et al*., *Phys. Rev. E*. **67** (2003) 066501, https://doi.org/10.1103/physreve.67.066501.

[483] J.R. Schneider, *J. Phys. B* **43** (2010) 194001, https://doi.org/10.1088/0953-4075/43/19/194001.

[484] D.A. Edwards *et al*., TESLA test facility linac-design report. *DESY Print March* (1995) 95–01.

[485] J. Rossbach, *Nucl. Instrum. Methods Phys. Res. A*. **375** (1996) 269–273, https://doi.org/10.1016/0168-9002(95)01332-6

[486] J. Andruszkow, B. Aune, V. Ayvazyan, N. Baboi, R. Bakker, V. Balakin, D. Barni, A. Bazhan, M. Bernard, A. Bosotti, J.C. Bourdon, W. Brefeld, R. Brinkmann, S. Buhler, J.-P. Carneiro, M. Castellano, P. Castro, L. Catani, S. Chel, Y. Cho, S. Choroba, E.R. Colby, W. Decking, P. Den Hartog, M. Desmons, M. Dohlus, D. Edwards, H.T. Edwards, B. Faatz, J. Feldhaus, M. Ferrario, M.J. Fitch, K. Flöttmann, M. Fouaidy, A. Gamp, T. Garvey, C. Gerth, M. Geitz, E. Gluskin, V. Gretchko, U. Hahn, W.H. Hartung, D. Hubert, M. Hüning, R. Ischebek, M. Jablonka, J.M. Joly, M. Juillard, T. Junquera, P. Jurkiewicz, A. Kabel, J. Kahl, H. Kaiser, T. Kamps, V.V. Katelev, J.L. Kirchgessner, M. Körfer, L. Kravchuk, G. Kreps, J. Krzywinski, T. Lokajczyk, R. Lange, B. Leblond, M. Leenen, J. Lesrel, M. Liepe, A. Liero, T. Limberg, R. Lorenz, L.H. Hua, L.F. Hai, C. Magne, M. Maslov, G. Materlik, A. Matheisen, J. Menzel, P. Michelato, W.-D. Möller, A. Mosnier, U.-C. Müller, O. Napoly, A. Novokhatski, M. Omeich, H.S. Padamsee, C. Pagani, F. Peters, B. Petersen, P. Pierini, J. Pflüger, P. Piot, B.P. Ngoc, L. Plucinski, D. Proch, K. Rehlich, S. Reiche, D. Reschke, I. Reyzl, J. Rosenzweig, J. Rossbach, S. Roth, E.L. Saldin, W. Sandner, Z. Sanok, H. Schlarb, G. Schmidt, P. Schmüser, J.R. Schneider, E.A. Schneidmiller, H.-J. Schreiber, S. Schreiber, P. Schütt, J. Sekutowicz, L. Serafini, D. Sertore, S. Setzer, S. Simrock,

B. Sonntag, B. Sparr, F. Stephan, V.A. Sytchev, S. Tazzari, F. Tazzioli, M. Tigner, M. Timm, M. Tonutti, E. Trakhtenberg, R. Treusch, D. Trines, V. Verzilov, T. Vielitz, V. Vogel, G.v. Walter, R. Wanzenberg, T. Weiland, H. Weise, J. Weisend, M. Wendt, M. Werner, M.M. White, I. Will, S. Wolff, M.V. Yurkov, K. Zapfe, P. Zhogolev and F. Zhou, *Phys. Rev. Lett.* **85** (2000) 3825–3829, https://doi.org/10.1103/physrevlett.85.3825

[487] V. Ayvazyan, N. Baboi, I. Bohnet, R. Brinkmann, M. Castellano, P. Castro, L. Catani, S. Choroba, A. Cianchi, M. Dohlus *et al.*, *Eur. Phys. J. D-Atomic, Molecular, Optical and Plasma Physics*. **20** (2002) 149–156, https://doi.org/10.1140/epjd/e2002-00121-4

[488] V. Ayvazyan, N. Baboi, I. Bohnet, R. Brinkmann, M. Castellano, P. Castro, L. Catani, S. Choroba, A. Cianchi, M. Dohlus, H.T. Edwards, B. Faatz, A.A. Fateev, J. Feldhaus, K. Flöttmann, A. Gamp, T. Garvey, H. Genz, Ch. Gerth, V. Gretchko, B. Grigoryan, U. Hahn, C. Hessler, K. Honkavaara, M. Hüning, R. Ischebeck, M. Jablonka, T. Kamps, M. Körfer, M. Krassilnikov, J. Krzywinski, M. Liepe, A. Liero, T. Limberg, H. Loos, M. Luong, C. Magne, J. Menzel, P. Michelato, M. Minty, U.-C. Müller, D. Nölle, A. Novokhatski, C. Pagani, F. Peters, J. Pflüger, P. Piot, L. Plucinski, K. Rehlich, I. Reyzl, A. Richter, J. Rossbach, E.L. Saldin, W. Sandner, H. Schlarb, G. Schmidt, P. Schmüser, J.R. Schneider, E.A. Schneidmiller, H.-J. Schreiber, S. Schreiber, D. Sertore, S. Setzer, S. Simrock, R. Sobierajski, B. Sonntag, B. Steeg, F. Stephan, K.P. Sytchev, K. Tiedtke, M. Tonutti, R. Treusch, D. Trines, D. Türke, V. Verzilov, R. Wanzenberg, T. Weiland, H. Weise, M. Wendt, I. Will, S. Wolff, K. Wittenburg, M.V. Yurkov and K. Zapfe, *Phys. Rev. Lett.* **88** (2002) 104802, https://doi.org/10.1103/physrevlett.88.104802

[489] H. Wabnitz, L. Bittner, A.R.B. De Castro, R. Döhrmann, P. Gürtler, T. Laarmann, W. Laasch, J. Schulz, A. Swiderski, K. von Haeften *et al.*, *Nature* **420** (2002) 482–485, https://doi.org/10.1038/nature01197

[490] V. Ayvazyan, N. Baboi, J. Bähr, V. Balandin, B. Beutner, A. Brandt, I. Bohnet, A. Bolzmann, R. Brinkmann, O.I. Brovko *et al.*, *Eur. Phys. J. D-Atomic, Molecular, Optical and Plasma Physics*. **37** (2005) 297–303, https://doi.org/10.1140/epjd/e2005-00308-1

[491] W. Ackermann, G. Asova, V. Ayvazyan, A. Azima, N. Baboi, J. Bähr, V. Balandin, B. Beutner, A. Brandt, A. Bolzmann, R. Brinkmann, O.I. Brovko, M. Castellano, P. Castro, L. Catani, E. Chiadroni, S. Choroba, A. Cianchi, J.T. Costello, D. Cubaynes, J. Dardis, W. Decking, H. Delsim-Hashemi, A. Delserieys, G. Di Pirro, M. Dohlus, S. Düsterer, A. Eckhardt, H.T. Edwards, B. Faatz, J. Feldhaus, K. Flättmann, J. Frisch, L. Frählich, T. Garvey, U. Gensch, Ch. Gerth, M. Gärler, N. Golubeva, H.-J. Grabosch, M. Grecki, O. Grimm, K. Hacker, U. Hahn, J.H. Han, K. Honkavaara, T. Hott, M. Häning, Y. Ivanisenko, E. Jaeschke, W. Jalmuzna, T. Jezynski, R. Kammering, V. Katalev, K. Kavanagh, E.T. Kennedy, S. Khodyachykh, K. Klose, V. Kocharyan, M. Kärfer, M. Kollewe, W. Koprek, S. Korepanov, D. Kostin, M. Krassilnikov, G. Kube, M. Kuhlmann, C.L.S. Lewis, L. Lilje, T. Limberg, D. Lipka, F. Löhl, H. Luna, M. Luong, M. Martins, M. Meyer, P. Michelato, V. Miltchev, W.D. Müller, L. Monaco, W.F.O. Müller, O. Napieralski, O. Napoly, P. Nicolosi, D. Nölle, T. Nuäez, A. Oppelt, C. Pagani, R. Paparella, N. Pchalek, J. Pedregosa-Gutierrez, B. Petersen, B. Petrosyan, G. Petrosyan, L. Petrosyan, J. Pflüger, E. Plünjes, L. Poletto, K. Pozniak, E. Prat, D. Proch, P. Pucyk, P. Radcliffe, H. Redlin, K. Rehlich, M. Richter, M. Roehrs, J. Roensch, R. Romaniuk, M. Ross, J. Rossbach, V. Rybnikov, M. Sachwitz, E.L. Saldin, W. Sandner, H. Schlarb, B. Schmidt, M. Schmitz, P. Schmüser, J.R. Schneider, E.A. Schneidmiller, S. Schnepp, S. Schreiber, M. Seidel, D. Sertore, A.V. Shabunov, C. Simon, S. Simrock, E. Sombrowski, A.A. Sorokin, P. Spanknebel, R. Spesyvtsev, L. Staykov, B. Steffen, F. Stephan, F. Stulle, H. Thom, K. Tiedtke, M. Tischer, S. Toleikis, R. Treusch, D. Trines, I. Tsakov, E. Vogel, T. Weiland, H. Weise, M. Wellhöfer, M. Wendt, I. Will, A. Winter, K. Wittenburg, W. Wurth, P. Yeates, M.V. Yurkov, I. Zagorodnov and K. Zapfe, *Nat Photonics*. **1** (2007) 336–342, https://doi.org/10.1038/nphoton.2007.76

[492] S. Schreiber, B. Faatz and K. Honkavaara. In *Proc. EPAC* (2008).

[493] S. Schreiber *et al*. In *Proc. of FEL* Jacow (2011).

[494] B. Faatz, E. Plönjes, S. Ackermann, A. Agababyan, V. Asgekar, V. Ayvazyan, S. Baark, N. Baboi, V. Balandin, N. von Bargen *et al*., *New J. Phys*. **18** (2016) 062002, https://doi.org/10.1088/1367-2630/18/6/062002

[495] H. Kitamura, *J. Synchrotron Radiat*. **7** (2000) 121–130, https://doi.org/10.1107/s0909049500002983

[496] H. Kitamura, T. Bizen, T. Hara, X. Marechal, T. Seike and T. Tanaka, *Nucl. Instrum. Methods Phys. Res. A*. **467** (2001) 110–113, https://doi.org/10.1016/s0168-9002(01)00242-x

[497] T. Shintake, H. Matsumoto, T. Ishikawa and H. Kitamura, SPring-8 compact SASE source (SCSS). In *International Symposium on Optical Science and Technology*, pages 12–23. International Society for Optics and Photonics, 2001.

[498] K. Togawa, H. Baba, K. Onoe, T. Inagaki, T. Shintake and H. Matsumoto, CeB_6*Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 312–315, https://doi.org/10.1016/j.nima.2004.04.077

[499] K. Togawa, T. Shintake, T. Inagaki, K. Onoe, T. Tanaka, H. Baba and H. Matsumoto, CeB_6*Phys. Rev. Special Topics-Accelerators and Beams*. **10** (2007) 020703, https://doi.org/10.1103/physrevstab.10.020703

[500] T. Shintake, H. Tanaka, T. Hara, T. Tanaka, K. Togawa, M. Yabashi, Y. Otake, Y. Asano, T. Bizen, T. Fukui, S. Goto, A. Higashiya, T. Hirono, N. Hosoda, T. Inagaki, S. Inoue, M. Ishii, Y. Kim, H. Kimura, M. Kitamura, T. Kobayashi, H. Maesaka, T. Masuda, S. Matsui, T. Matsushita, X. Maréchal, M. Nagasono, H. Ohashi, T. Ohata, T. Ohshima, K. Onoe, K. Shirasawa, T. Takagi, S. Takahashi, M. Takeuchi, K. Tamasaku, R. Tanaka, Y. Tanaka, T. Tanikawa, T. Togashi, S. Wu, A. Yamashita, K. Yanagida, C. Zhang, H. Kitamura and T. Ishikawa, *Nat. Photonics*. **2** (2008) 555–559, https://doi.org/10.1038/nphoton.2008.134

[501] T. Shintake, H. Tanaka, T. Hara, T. Tanaka, K. Togawa, Makina ÄÄ, Y. Otake, Y. Asano, T. Fukui, T. Hasegawa *et al*., *Phys. Rev. Special Topics-Accelerators and Beams*. **12** (2009) 070701, https://doi.org/10.1103/physrevstab.12.070701

[502] C. Pellegrini, J. Rosenzweig, H.-D. Nuhn, P. Pianetta, R. Tatchyn, H. Winick, K. Bane, P. Morton, T. Raubenheimer, J. Seeman *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **331** (1993) 223–227, https://doi.org/10.1016/0168-9002(93)90047-l

[503] T.O. Raubenheimer, *Nucl. Instrum. Methods Phys. Res. A*. **358** (1995) 40–43, https://doi.org/10.1016/0168-9002(94)01502-3

[504] J. Arthur, G. Materlik and H. Winick, Workshop on scientific applications of coherent X-rays. SLAC Report, 437 (1994).

[505] M. Cornacchia, Linac coherent light source (LCLS) design study report. Technical report, Stanford Linear Accelerator Center, Menlo Park, CA (US), 1998.

[506] B. Birgeneau and Z.X. Shen, Report of the Basic Energy Sciences Advisory Committee Panel on DOE synchrotron radiation sources and science. Technical report (1997).

[507] P.R. Bolton, J.E. Clendenin, D.H. Dowell, M. Ferrario, A.S. Fisher, S.M. Gierman, R.E. Kirby, P. Krejcik, C.G. Limborg, G.A. Mulhollan *et al*., *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 296–300, https://doi.org/10.1016/s0168-9002(02)00331-5

[508] C. Limborg, P. Bolton, D.H. Dowell, S. Gierman and J.F. Schmerge, *Nucl. Instrum. Methods Phys. Res. A*. **528** (2004) 350–354, https://doi.org/10.1016/j.nima.2004.04.066

[509] J. Arthur *et al*., Linac Coherent Light Source (LCLS) Conceptual Design Report. Technical report, SLAC-R-593, 2002.

[510] M. Borland, *Phys. Rev. Special Topics-Accelerators and Beams*. **11** (2008) 030701, https://doi.org/10.1103/physrevstab.11.030701

[511] E. Trakhtenberg, V. Tcheskidov, I. Vasserman, N. Vinokurov, M. Erdmann and J. Pfluger, *Nucl.*

Instrum. Methods Phys. Res. A. **543** (2005) 42–46, https://doi.org/10.1016/j.nima.2005.01.110

[512] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P.H. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H.-D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky and J. Galayda, *Nat. Photonics*. **4** (2010) 641–647, https://doi.org/10.1038/nphoton.2010.176

[513] Y. Ding, A. Brachmann, F.-J. Decker, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, Ph. Hering, Z. Huang *et al.*, *Phys. Rev. Lett*. **102** (2009) 254801, https://doi.org/10.1103/physrevlett.102.254801

[514] P. Emma, R. Carr and H.-D. Nuhn, *Nucl. Instrum. Methods Phys. Res. A*. **429** (1999) 407–413, https://doi.org/10.1016/s0168-9002(99)00117-5

[515] Z. Huang, A. Brachmann, F.-J. Decker, Y. Ding, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, R. Iverson, H. Loos, A. Miahnahri, H.-D. Nuhn, D. Ratner, G. Stupakov, J. Turner, J. Welch, W. White, J. Wu and D. Xiang, *Phys. Rev. Special Topics-Accelerators and Beams*. **13** (2010) 020703, https://doi.org/10.1103/physrevstab.13.020703

[516] T. Ishikawa, H. Aoyagi, T. Asaka, Y. Asano, N. Azumi, T. Bizen, H. Ego, K. Fukami, T. Fukui, Y. Furukawa, S. Goto, H. Hanaki, T. Hara, T. Hasegawa, T. Hatsui, A. Higashiya, T. Hirono, N. Hosoda, M. Ishii, T. Inagaki, Y. Inubushi, T. Itoga, Y. Joti, M. Kago, T. Kameshima, H. Kimura, Y. Kirihara, A. Kiyomichi, T. Kobayashi, C. Kondo, T. Kudo, H. Maesaka, X.M. Maréchal, T. Masuda, S. Matsubara, T. Matsumoto, T. Matsushita, S. Matsui, M. Nagasono, N. Nariyama, H. Ohashi, T. Ohata, T. Ohshima, S. Ono, Y. Otake, C. Saji, T. Sakurai, T. Sato, K. Sawada, T. Seike, K. Shirasawa, T. Sugimoto, S. Suzuki, S. Takahashi, H. Takebe, K. Takeshita, K. Tamasaku, H. Tanaka, R. Tanaka, T. Tanaka, T. Togashi, K. Togawa, A. Tokuhisa, H. Tomizawa, K. Tono, S. Wu, M. Yabashi, M. Yamaga, A. Yamashita, K. Yanagida, C. Zhang, T. Shintake, H. Kitamura and N. Kumagai, *Nat. Photonics*. **6** (2012) 540–544, https://doi.org/10.1038/nphoton.2012.141

[517] T. Tanaka, S. Goto, T. Hara, T. Hatsui, H. Ohashi, K. Togawa, M. Yabashi and H. Tanaka. *Phys. Rev. Special Topics-Accelerators and Beams*. **15** (2012) 110701, https://doi.org/10.1103/physrevstab.15.110701

[518] H. Tanaka, Present status of SACLA. In *Proc. of FEL2017, Santa Fe, NM, USA*, Jacow, (2017).

[519] J.S. Oh, C.W. Chung, D.E. Kim, E.S. Kim, H.S. Kang, I.S. Ko, K.B. Lee, S.H. Nam, S.J. Park, T.Y. Koo *et al.*, Compact X-ray free electron laser program at the PAL. In *Proc. of APAC 2004, Gyeongju, Korea*, Jacow, (2004), pp. 261–263.

[520] D.E. Kim, Y.G. Jung, H.S. Kang, I.S. Ko, H.G. Lee, S.B. Lee, W.W. Lee, D.G. Oh, K.H. Park, H.S. Suh and J. Pflueger, Development of PAL-XFEL undulator system. In *Proc. of IPAC 2016, Busan, Korea*, Jacow, (2016), pp. 4044–4046.

[521] J. Hong, S.-Ik, Moon, Y.-W. Parc, M. Cho, I.S. Ko, C. Kim, S.-J. Park, Y.-J. Park, S.-H. Kim and W. Namkung. *J. Korean Phys. Soc.* **58** (2011) 198–204.

[522] SEOUL, June 30 (Korea Bizwire), http://koreabizwire.com/korea-succeeds-in-producing-x-ray-free-electron-laser/59745.

[523] PALFEL, http://pal.postech.ac.kr/paleng/menu.pal?method=menuview&pagemode=paleng&top=7&sub=5&sub2=0&sub3=0.

[524] H. H. Braun for SwissFEL team, , Commissioning of SwissFEL. In *Proc. of IPAC 2017, Copenhagen, Denmark*, Jacow, (2017).

[525] R. Ganter, SwissFEL-Conceptual design report. Technical report, Paul Scherrer Institute (PSI), 2010.

[526] S. Reiche, Status of the SwissFEL facility at the Paul Scherrer Institute. In *Proc. FEL 2011, Shanghai, China*, Jacow, pages 223–226, 2011.

[527] T. Schmidt and S. Reiche, Undulators for the SwissFEL. In *Proc. of FEL2009, Liverpool, UK*, Jacow, pages 706–713, 2009.

[528] T. Schietinger, M. Pedrozzi, M. Aiba, V. Arsov, S. Bettoni, B. Beutner, M. Calvi, P. Craievich, M. Dehler, F. Frei *et al.*, *Phys. Rev. Accelerators and Beams*. **19** (2016) 100702, https://doi.org/10.1103/physrevaccelbeams.19.034402

[529] The Swiss FEL, https://www.psi.ch/swissfel.

[530] M. Altarelli, R. Brinkmann, M. Chergui, W. Decking, B. Dobson, S. Dusterer, G. Grubel, W. Graeff, H. Graafsma, J. Hajdu *et al.*, The European X-ray free electron laser, Technical Design Report, DESY. Technical report, 2006.

[531] The European X-Ray Laser Project XFEL, www.xfel.eu/.

[532] W. Decking, H. Weise on behalf of the European XFEL Accelerator Consortium and Commissioning Team, In *Proc. of IPAC2017, Copenhagen, Denmark*, Jacow, MOC03, 2017.

[533] H. Weise, W. Decking, on behalf of the European XFEL Accelerator Consortium and Commissioning Team, In *Proc. of FEK2017, Santa Fe, NM, USA*, Jacow, MOXAA1, 2017.

[534] T. Raubenheimer *et al.*, The LCLS-II, a new FEL facility at SLAC. In *Proc. 36th Int. Free-Electron Laser Conf., Basel*, Jacow, (2014).

[535] Z. Y. Zhu, Z. T. Zhao, D. Wang, Z. Liu, R. X. Li, L. X. Yin, Z. H. Yang for the the SCLF Team. In *Proc. of FEK2017, Santa Fe, NM, USA*, Jacow, MOP055, 2017.

[536] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *Opt. Commun*. **148** (1998) 383–403, https://doi.org/10.1016/s0030-4018(97)00670-6

[537] S. Reiche, P. Musumeci, C. Pellegrini and J.B. Rosenzweig, *Nucl. Instrum. Methods Phys. Res. A* **593** (2008) 45–48, https://doi.org/10.1016/j.nima.2008.04.061

[538] L. Giannessi, A. Bacci, M. Bellaveglia, F. Briquez, M. Castellano, E. Chiadroni, A. Cianchi, F. Ciocci, M.E. Couprie, L. Cultrera, G. Dattoli, D. Filippetto, M.D. Franco, G.D. Pirro, M. Ferrario, L. Ficcadenti, F. Frassetto, A. Gallo, G. Gatti, M. Labat, G. Marcus, M. Moreno, A. Mostacci, E. Pace, A. Petralia, V. Petrillo, L. Poletto, M. Quattromini, J.V. Rau, C. Ronsivalle, J. Rosenzweig, A.R. Rossi, V.R. Albertini, E. Sabia, M. Serluca, S. Spampinati, I. Spassovsky, B. Spataro, V. Surrenti, C. Vaccarezza and C. Vicario, *Phys. Rev. Lett.* **106** (2011) 144801, https://doi.org/10.1103/physrevlett.106.144801

[539] L. Giannessi, D. Alesini, P. Antici, A. Bacci, M. Bellaveglia, R. Boni, M. Boscolo, F. Briquez, M. Castellano, L. Catani *et al.*, *Phys. Rev. Special Topics-Accelerators and Beams*. **14** (2011) 060712, https://doi.org/10.1103/physrevstab.14.102803

[540] N.R. Thompson and B.W.J. McNeil, *Phys. Rev. Lett*. **100** (2008) 203901, https://doi.org/10.1103/physrevlett.100.203901

[541] P. Emma, K. Bane, M. Cornacchia, Z. Huang, H. Schlarb, G. Stupakov and D. Walz, *Phys. Rev. Lett*. **92** (2004) 074801, https://doi.org/10.1103/physrevlett.92.074801

[542] C.B. Schroeder, C. Pellegrini, S. Reiche, J. Arthur and P. Emma, *Nucl. Instrum. Methods Phys. Res. A*. **483** (2002) 89–93, https://doi.org/10.1016/s0168-9002(02)00291-7

[543] A.A. Zholents and W.M. Fawley, *Phys. Rev. Lett*. **92** (2004) 224801, https://doi.org/10.1103/physrevlett.92.224801

[544] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *Opt. Commun*. **239** (2004) 161–172, https://doi.org/10.1016/j.optcom.2004.05.024

[545] A.A. Zholents, *Phys. Rev. Special Topics-Accelerators and Beams*. **8** (2005) 040701, https://doi.org/10.1103/physrevstab.8.050704

[546] E.L Saldin, E.A Schneidmiller and M.V. Yurkov, *Phys. Rev. Special Topics-Accelerators and Beams*. **9** (2006) 050702, https://doi.org/10.1103/physrevstab.9.030702

[547] A.A. Zholents and M.S. Zolotorev, *New J. Phys*. **10** (2008) 025005, https://doi.org/10.1088/1367-

2630/10/2/025005

[548] T. Tanaka, *Phys. Rev. Lett.* **110** (2013) 084801.

[549] D.J. Dunning, B.W.J. McNeil and N.R. Thompson, *Phys. Rev. Lett.* **110** (2013) 104801, https://doi.org/10.1103/physrevlett.110.104801

[550] E. Prat and S. Reiche, *Phys. Rev. Lett.* **114** (2015) 244801, https://doi.org/10.1103/physrevlett.114.244801

[551] S. Huang, Y. Ding, Z. Huang and G. Marcus, *Phys. Rev. Accel. Beams.* **19** (2016) 080702, https://doi.org/10.1103/physrevaccelbeams.19.080702

[552] R. Bonifacio, L. De Salvo Souza, P. Pierini and N. Piovella, *Nucl. Instrum. Methods Phys. Res. A.* **296** (1990) 358–367, https://doi.org/10.1016/0168-9002(90)91234-3

[553] C. Bourassin-Bouchet and M.-E. Couprie, *Nat. Commun.* **6** (2015), https://doi.org/10.1038/ncomms7465[547] C. Bourassin-Bouchet and M-E. Couprie, *Nat. Commun.* **6** (2015) 6465, https://doi.org/10.1038/ncomms7465.

[554] A. Singer, I.A. Vartanyants, M. Kuhlmann, S. Duesterer, R. Treusch and J. Feldhaus, *Phys. Rev. Lett.* **101** (2008) 254801, https://doi.org/10.1103/physrevlett.101.254801.

[555] R. Bachelard, P. Mercère, M. Idir, M-E. Couprie, M. Labat, O. Chubar, Guillaume Lambert, P. Zeitoun, H. Kimura, H Ohashi, *et al.*, *Phys. Rev. Lett.* **106** (2011) 234801, https://doi.org/10.1103/physrevlett.106.234801.

[556] A.B. Temnykh, *Phys. Rev. ST Accel. Beams* **11** (2008) 120702–120711, https://doi.org/10.1103/physrevstab.11.120702.

[557] A.A. Lutman, J.P. MacArthur, M. Ilchen, A.O. Lindahl, J. Buck, R.N. Coffee, G.L. Dakovski, L. Dammann, Y. Ding, H.A. Dürr, *et al.*, *Nat. Photonics* **10** (2016) 468–472.

[558] L.-H. Yu and I. Ben-Zvi, High-gain harmonic generation of soft X-rays with the "fresh bunch" technique. *Nucl. Instrum. Methods Phys. Res. A* **393** (1997) 96–99, https://doi.org/10.1016/s0168-9002(97)00435-x.

[559] L.-H. Yu, L. DiMauro, A. Doyuran, W.S. Graves, E.D. Johnson, R. Heese, S. Krinsky, H. Loos, J.B. Murphy, G. Rakowsky, *et al.*, *Phys. Rev. Lett.* **91** (2003) 074801.

[560] A. Doyuran, M. Babzien, T. Shaftan, L.H. Yu, L.F. DiMauro, I. Ben-Zvi, S.G. Biedron, W. Graves, E. Johnson, S. Krinsky, *et al.*, *Phys. Rev. Lett.* **86** (2001) 5902.

[561] W. Li, R.R. Lucchese, A. Doyuran, Z. Wu, H. Loos, G.E. Hall and A.G. Suits, *Phys. Rev. Lett.* **92** (2004) 083002–083005, https://doi.org/10.1103/physrevlett.92.083002.

[562] T. Shaftan and L.H. Yu, *Phys. Rev. E.* **71** (2005) 046501, https://doi.org/10.1103/physreve.71.046501.

[563] T. Watanabe, X.J. Wang, J.B. Murphy, J. Rose, Y. Shen, T. Tsang, L. Giannessi, P. Musumeci and S. Reiche, *Phys. Rev. Lett.* **98** (2007) 034802, https://doi.org/10.1103/physrevlett.98.034802.

[564] M. Ferrario, D. Alesini, A. Bacci, M. Bellaveglia, R. Boni, M. Boscolo, M. Castellano, L. Catani, E. Chiadroni, S. Cialdi, *et al.*, *Phys. Rev. Lett.* **99** (2007) 234801, https://doi.org/10.1103/physrevlett.99.234801.

[565] L. Giannessi, M. Artioli, M. Bellaveglia, F. Briquez, E. Chiadroni, A. Cianchi, M. E. Couprie, G. Dattoli, E. Di Palma, G. Di Pirro, M. Ferrario, D. Filippetto, F. Frassetto, G. Gatti, M. Labat, G. Marcus, A. Mostacci, A. Petralia, V. Petrillo, L. Poletto, M. Quattromini, J. V. Rau, J. Rosenzweig, E. Sabia, M. Serluca, I. Spassovsky and V. Surrenti, *Phys. Rev. Lett.* **108** (2012) 164801, https://doi.org/10.1103/physrevlett.108.164801.

[566] L. Giannessi, M. Bellaveglia, E. Chiadroni, A. Cianchi, M.E. Couprie, M. Del Franco, G. Di Pirro, M. Ferrario, G. Gatti, M. Labat, *et al.*, *Phys. Rev. Lett.* **110** (2013) 044801, https://doi.org/10.1103/physrevlett.110.044801.

[567] Marie Emmanuelle Couprie, *J. Electron Spectros. Relat. Phenomena* **196** (2014) 3–13,

https://doi.org/10.1016/j.elspec.2013.12.007.

[568] M.E. Couprie, *Nucl. Instrum. Methods Phys. Res. B* **364** (2015) 4–15, https://doi.org/10.1016/j.nimb.2015.08.084.

[569] M. Ferray, A. l'Huillier, X.F. Li, L.A. Lompré, G. Mainfray and C. Manus, $1064\,nm$*J. Phys. B* **21** (1988) L31–L36, https://doi.org/10.1088/0953-4075/21/3/001.

[570] A. McPherson, G. Gibson, H. Jara, U. Johann, Ting S. Luk, I.A. McIntyre, Keith Boyer and Charles K. Rhodes, *J. Opt. Soc. Am. B* **4** (1987) 595–601, https://doi.org/10.1364/josab.4.000595.

[571] F. Canova and L. Poletto, *Optical Technologies for Extreme-Ultraviolet and Soft X-ray Coherent Sources*, Vol. 197 (Springer, 2015).

[572] D. Garzella, T. Hara, B. Carre, P. Salieres, T. Shintake, H. Kitamura and M.E. Couprie, *Nucl. Instrum. Methods Phys. Res. A* **528** (2004) 502–505, https://doi.org/10.1016/j.nima.2004.04.089.

[573] G. Lambert, T. Hara, D. Garzella, T. Tanikawa, M. Labat, B. Carre, H. Kitamura, T. Shintake, M. Bougeard, S. Inoue, Y. Tanaka, P. Salieres, H. Merdji, O. Chubar, O. Gobert, K. Tahara and M-E. Couprie, *Nat. Phys*. **4** (2008) 296–300, https://doi.org/10.1038/nphys889.

[574] T. Tanikawa, Guillaume Lambert, T. Hara, M. Labat, Y. Tanaka, M. Yabashi, O. Chubar and M.E. Couprie. *EPL (Europhys. Lett.)* **94** (2011) 34001, https://doi.org/10.1209/0295-5075/94/34001.

[575] G. Lambert, T. Hara, M. Labat, T. Tanikawa, Y. Tanaka, M. Yabashi, D. Garzella, B. Carré and M.E. Couprie, *EPL (Europhy. Lett.)* **88** (2009) 54002, https://doi.org/10.1209/0295-5075/88/54002.

[576] T. Togashi, E.J. Takahashi, K. Midorikawa, M. Aoyama, K. Yamakawa, T. Sato, A. Iwasaki, S. Owada, T. Okino, K. Yamanouchi, F. Kannari, A. Yagishita, H. Nakano, M-E. Couprie, K. Fukami, T. Hatsui, T. Hara, T. Kameshima, H. Kitamura, N. Kumagai, S. Matsubara, M. Nagasono, H. Ohashi, T. Ohshima, Y. Otake, T. Shintake, K. Tamasaku, H. Tanaka, T. Tanaka, K. Togawa, H. Tomizawa, T. Watanabe, M. Yabashi and T. Ishikawa, *Opt. Express* **19** (2011) 317–324.

[577] S. Matsubara, T. Togashi, E.J. Takahashi, K. Midorikawa, M. Aoyama, K. Yamakawa, T. Sato, A. Iwasaki, S. Owada, K. Yamanouchi, *et al*., in *Lasers and Electro-Optics Europe (CLEO EUROPE/IQEC), 2013 Conference on and International Quantum Electronics Conference* (IEEE, 2013), pp. 1–1.

[578] K. Ogawa, T. Sato, S. Owada, S. Matsubara, Y. Okayasu, T. Togashi, T. Watanabe, E.J. Takahashi, K. Midorikawa, M. Aoyama, *et al*., in *International Conference on Ultrafast Phenomena* (Optical Society of America, 2014), p. 07.

[579] E.J. Takahashi, T. Togashi, M. Aoyama, K. Yamakawa, T. Sato, A. Iwasaki, S. Owada, K. Yamanouchi, T. Hara, S. Matsubara, *et al*. In *Lasers and Electro-Optics (CLEO), 2013 Conference on* (IEEE, 2013), pp. 1–2.

[580] M. Labat, M. Bellaveglia, M. Bougeard, B. Carré, F. Ciocci, E. Chiadroni, A. Cianchi, M-E. Couprie, L. Cultrera, M. Del Franco, G. Di Pirro, A. Drago, M. Ferrario, D. Filippetto, F. Frassetto, A. Gallo, D. Garzella, G. Gatti, L. Giannessi, G. Lambert, A. Mostacci, A. Petralia, V. Petrillo, L. Poletto, M. Quattromini, J.V. Rau, C. Ronsivalle, E. Sabia, M. Serluca, I. Spassovsky, V. Surrenti, C. Vaccarezza and C. Vicario, *Phys. Rev. Lett*. **107** (2011) 224801–224805.

[581] S. Ackermann, A. Azima, S. Bajt, J. Bödewadt, F. Curbis, H. Dachraoui, H. Delsim-Hashemi, M. Drescher, S. Düsterer, B. Faatz, *et al*., *Phys. Rev. Lett*. **111** (2013) 114801, https://doi.org/10.1103/physrevlett.111.114801.

[582] C.J. Bocchetta, D. Bulfone, F. Cargnello, M. Danailov, G. D'Auria, B. Diviacco, M. Ferianis, A. Gambitta, E. Karantzoulis, G. Loda, *et al*., *Nucl. Instrum. Methods Phys. Res. A* **507** (2003) 484–488, https://doi.org/10.1016/s0168-9002(03)00901-x.

[583] S. Di Mitri, E. Allaria, L. Badano, C. Bontoiu, M. Cornacchia, P. Craievich, M. Danailov, G. De Ninno, B. Diviacco, O. Ferrando, *et al*., *Nucl. Instrum. Methods Phys. Res. A* **608** (2009) 19–27.

[584] S. Di Mitri, M. Cornacchia, S. Spampinati and S. Milton, *Phys. Rev. ST Accel. Beams* **13** (2010) 010702, https://doi.org/10.1103/physrevstab.13.010702.

[585] S. Di Mitri, M. Cornacchia, C. Scafuri and Magnus Sjöström, *Phys. Rev. ST Accel. and Beams* **15** (2012) 012802, https://doi.org/10.1103/physrevstab.15.012802.

[586] Enrico Allaria, Carlo Callegari, Daniele Cocco, William M. Fawley, Maya Kiskinova, Claudio Masciovecchio and Fulvio Parmigiani, *New J. Phys*. **12** (2010) 075002, https://doi.org/10.1088/1367-2630/12/7/075002.

[587] E. Allaria, Roberto Appio, L. Badano, W.A. Barletta, S. Bassanese, S.G. Biedron, A. Borga, E. Busetto, D. Castronovo, P. Cinquegrana, *et al*., *Nat Photonics* **6** (2012) 699–704.

[588] E. Allaria, Roberto Appio, L. Badano, W.A. Barletta, S. Bassanese, S.G. Biedron, A. Borga, E. Busetto, D. Castronovo, P. Cinquegrana, S. Cleva, D. Cocco, M. Cornacchia, P. Craievich, I. Cudin, G. D'Auria, M. Dal Forno, M.B. Danailov, R. De Monte, G. De Ninno, P. Delgiusto, A. Demidovich, S. Di Mitri, B. Diviacco, A. Fabris, R. Fabris, W. Fawley, M. Ferianis, E. Ferrari, S. Ferry, L. Froehlich, P. Furlan, G. Gaio, F. Gelmetti, L. Giannessi, M. Giannini, R. Gobessi, R. Ivanov, E. Karantzoulis, M. Lonza, A. Lutman, B. Mahieu, M. Milloch, S.V. Milton, M. Musardo, I. Nikolov, S. Noe, F. Parmigiani, G. Penco, M. Petronio, L. Pivetta, M. Predonzani, F. Rossi, L. Rumiz, A. Salom, C. Scafuri, C. Serpico, P. Sigalotti, S. Spampinati, C. Spezzani, M. Svandrlik, C. Svetina, S. Tazzari, M. Trovo, R. Umer, A. Vascotto, M. Veronese, R. Visintini, M. Zaccaria, D. Zangrando and M. Zangrando, *Nat Photonics* **6** (2012) 699–704.

[589] S. Sasaki, K. Miyata and T. Takada, *Jpn. J. Appl. Phys*. **31 (12B)** (1992) 1794–1796, https://doi.org/10.1143/jjap.31.l1794.

[590] E. Allaria, D. Castronovo, P. Cinquegrana, P. Craievich, M. Dal Forno, M.B. Danailov, G. D'Auria, A. Demidovich, G. De Ninno, S. Di Mitri, *et al*., *Nat Photonics* **7** (2013) 913–918.

[591] E. Allaria, A. Battistoni, F. Bencivenga, R. Borghes, C. Callegari, F. Capotondi, D. Castronovo, P. Cinquegrana, D. Cocco, M. Coreno, *et al*., *New J. Phys*. **14** (2012) 113009–113027.

[592] L. Giannessi, in *Proc. FEL Conference, Basel, Switzerland*, Jacow, (2014), pp. 9–13.

[593] G. De Ninno, B. Mahieu, E. Allaria, L. Giannessi and S. Spampinati, *Phys. Rev. Lett*. **110** (2013) 064801, https://doi.org/10.1103/physrevlett.110.064801.

[594] B. Mahieu, E. Allaria, D. Castronovo, M. B. Danailov, A. Demidovich, G. De Ninno, S. Di Mitri, W. M. Fawley, E. Ferrari, L. Fröhlich, *et al*., *Opt. Express* **21** (2013) 22728–22741, https://doi.org/10.1364/oe.21.022728.

[595] E. Allaria, F. Bencivenga, R. Borghes, F. Capotondi, D. Castronovo, P. Charalambous, P. Cinquegrana, M.B. Danailov, G. De Ninno, A. Demidovich, *et al*., Two-colour pump–probe experiments with a twin-pulse-seed extreme ultraviolet free-electron laser. *Nat. Commun*. **4** (2013) 2476, https://doi.org/10.1038/ncomms3476.

[596] Enrico Allaria, Bruno Diviacco, Carlo Callegari, Paola Finetti, Benot Mahieu, Jens Viefhaus, Marco Zangrando, Giovanni De Ninno, Guillaume Lambert, Eugenio Ferrari, *et al*., *Phys. Rev. X* **4** (2014) 041040.

[597] F.E.L. Dalian, (2017) *www.laserfocusworld.com/articles/2017/01/china-develops-world-s-brightest-vuv-free-electron-laser-research-facility.html*.

[598] W. Q. Zhang, D. X. Dai, G. L. Wang, G. R. Wu, X. M. Yang, S. Chen, C. Feng, D. Wang, M. Zhang, Z. T. Zhao , in *Proc. FEL Conference, Santa Fe, NM, USA*, Jacow, (2017), MOC04.

[599] D. Xiang, E. Colby, M. Dunning, S. Gilevich, C. Hast, K. Jobe, D. McCormick, J. Nelson, T.O. Raubenheimer, K. Soong, G. Stupakov, Z. Szalata, D. Walz, S. Weathersby and M. Woodley, *Phys. Rev. Lett*. **105** (2010) 114801–114804, https://doi.org/10.1103/physrevlett.105.114801.

[600] E. Hemsing, M. Dunning, C. Hast, T.O. Raubenheimer, S. Weathersby and D. Xiang, *Phys. Rev. ST Accel. Beams* **17** (2014) 070702, https://doi.org/10.1103/physrevstab.17.070702.

[601] E. Hemsing, D. Xiang, M. Dunning, S. Weathersby, C. Hast and T. Raubenheimer, *Phys. Rev. ST Accel. Beams* **17** (2014) 010703, https://doi.org/10.1103/physrevstab.17.010703.

[602] E. Hemsing, M. Dunning, B. Garcia, C. Hast, T. Raubenheimer, G. Stupakov and D. Xiang, *Nat. Photonics* **10** (2016) 512–515, https://doi.org/10.1038/nphoton.2016.101.

[603] E. Hemsing and A. Marinelli, *Phys. Rev. Lett.* **109** (2012) 224801, https://doi.org/10.1103/physrevlett.109.224801.

[604] E. Hemsing, A. Knyazik, M. Dunning, D. Xiang, A. Marinelli, C. Hast and J.B. Rosenzweig, *Nat. Phys.* **9** (2013) 549–553, https://doi.org/10.1038/nphys2712.

[605] C. Feng, T. Zhang, J. Chen, H. Deng, M. Zhang, X. Wang, B. Liu, T. Lan, D. Wang and Z. Zhao, *Phys. Rev. ST Accel. Beams* **14** (2011) 090701, https://doi.org/10.1103/physrevstab.14.090701.

[606] B. Liu, W.B. Li, J.H. Chen, Z.H. Chen, H.X. Deng, J.G. Ding, Y. Fan, G.P. Fang, C. Feng, L. Feng, *et al.*, *Phys. Rev. ST Accel. Beams* **16** (2013) 020704.

[607] C. Feng, H. Deng, D. Wang and Z. Zhao, *New J. Phys.* **16** (2014) 043021, https://doi.org/10.1088/1367-2630/16/4/043021.

[608] C. Feng, T. Zhang, H. Deng and Z. Zhao, *Phys. Rev. ST Accel. Beams* **17** (2014) 070701, https://doi.org/10.1103/physrevstab.17.070701.

[609] H. Deng, T. Zhang, L. Feng, C. Feng, B. Liu, X. Wang, T. Lan, G. Wang, W. Zhang, X. Liu, *et al.*, *Phys. Rev. ST Accel. Beams* **17** (2014) 020704, https://doi.org/10.1103/physrevstab.17.020704.

[610] H. Deng, M. Zhang, C. Feng, T. Zhang, X. Wang, T. Lan, L. Feng, W. Zhang, X. Liu, H. Yao, *et al.*, *Phys. Rev. Lett.* **113** (2014) 254802, https://doi.org/10.1103/physrevlett.113.254802.

[611] Z.T. Zhao, Dong Wang, J.H. Chen, Z.H. Chen, H.X. Deng, J.G. Ding, C. Feng, Q. Gu, M.M. Huang, T.H. Lan, Y.B. Leng, D.G. Li, G.Q. Lin, B. Liu, E. Prat, X.T. Wang, Z.S. Wang, K.R. Ye, L.Y. Yu, H.O. Zhang, J.Q. Zhang, Me Zhang, Mi Zhang, T. Zhang, S.P. Zhong and Q.G. Zhou, *Nat. Photonics* **6** (2012) 360–363.

[612] Z.T. Zhao, D. Wang, L.X. Yin, Q. Gu, G.P. Fang and B. Liu, *AAPS Bull.* **26** (2016) 12–24.

[613] J. Feldhaus, E.L. Saldin, J.R. Schneider, E.A. Schneidmiller and M.V. Yurkov, *Opt. Commun.* **140** (1997) 341–352, https://doi.org/10.1016/s0030-4018(97)00163-6.

[614] G. Geloni, V. Kocharyan and E. Saldin, A novel self-seeding scheme for hard X-ray FELs. *J. Mod. Opt.* **58** (2011) 1391–1403, https://doi.org/10.1080/09500340.2011.586473.

[615] J. Amann, W. Berg, V. Blank, F-J. Decker, Y. Ding, P. Emma, Y. Feng, J. Frisch, D. Fritz, J. Hastings, Z. Huang, J. Krzywinski, R. Lindberg, H. Loos, A. Lutman, H-D. Nuhn, D. Ratner, J. Rzepiela, D. Shu, Yu Shvyd'ko, S. Spampinati, S. Stoupin, S. Terentyev, E. Trakhtenberg, D. Walz, J. Welch, J. Wu, A. Zholents and D. Zhu, *Nat. Photonics* **6** (2012) 693–698.

[616] D. Ratner, R. Abela, J. Amann, C. Behrens, D. Bohler, G. Bouchard, C. Bostedt, M. Boyes, K. Chow, D. Cocco, *et al.*, *Phys. Rev. Lett.* **114** (2015) 054801.

[617] M. Yabashi and T. Tanaka, *Nat. Photonics* **6** (2012) 648–649, https://doi.org/10.1038/nphoton.2012.237.

[618] B. McNeil, *Nat. Photonics* **3** (2009) 375–377, https://doi.org/10.1038/nphoton.2009.110.

[619] C. Bostedt, S. Boutet, D.M. Fritz, Z. Huang, H.J. Lee, H.T. Lemke, A. Robert, W.F. Schlotter, J.J. Turner and G.J. Williams, *Rev. Mod. Phys.* **88** (2016) 015007.

[620] H.N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M.S. Hunter, J. Schulz, D.P. DePonte, U. Weierstall, R.B. Do Andak, F.R.N.C. Maia, A.V. Martin, I. Schlichting, L. Lomb, N. Coppola, R.L. Shoeman, S.W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M.J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hamke, C. Reich, D. Pietschner, L. Strader, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K-U. Kahnel, M. Messerschmidt, J-D. Bozek, S.P. Hau-Riege,

M. Frank, C.Y. Hampton, R.G. Sierra, D. Starodub, G.J. Williams, J. Hajdu, N. Timneanu, M.M. Seibert, J. Andreasson, A. Rocker, O. Jansson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.D. Schröter, F. Krasniqi, M. Bott, K.E. Schmidt, X. Wang, I. Grotjohann, J.M. Holton, T.R.M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J.C.H. Spence. *Nature* **470** (2011) 73–77.

[621] M.M. Seibert, T. Ekeberg, F.R.N.C. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odic', B. Iwan, A. Rocker, D. Westphal, M. Handantke, D.P. DePonte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T.A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J-M. Claverie, C. Bostedt, J.D. Bozek, S. Boutet, A.A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K.O. Hodgson, M.J. Bogan, C.Y. Hampton, R.G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J.C.H. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R.B. Doak, S. Marchesini, S.P. Hau-Riege, M. Frank, R.L. Shoeman, L. Lomb, S.W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Hamke, C. Reich, D. Pietschner, G. Weidenspointner, L. Strader, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.U. Kahnel, R. Andritschke, C-D. Schrater, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H.N. Chapman and J. Hajdu. *Nature* **470** (2011) 78–81.

[622] R.A. Kirian, T.A. White, J.M. Holton, H.N. Chapman, P. Fromme, A. Barty, L. Lomb, A. Aquila, F.R.N.C. Maia, A.V. Martin, R. Fromme, X. Wang, M.S. Hunter, K.E. Schmidt and JCH Spence, *Acta Crystallogr. A* **67** (2011) 131–140, https://doi.org/10.1107/s0108767310050981.

[623] L. Young, E.P. Kanter, B. Krässig, Y. Li, A.M. March, S.T. Pratt, R. Santra, S.H. Southworth, N. Rohringer, L.F. DiMauro, G. Doumy, C.A. Roedig, N. Berrah, L. Fang, M. Hoener, P.H. Bucksbaum, J.P. Cryan, S. Ghimire, J.M. Glownia, D.A. Reis, J.D. Bozek, C. Bostedt and M. Messerschmidt, *Nature* **466** (2010) 56–61, https://doi.org/10.1038/nature09177.

[624] N. Berrah, J. Bozek, J.T. Costello, S. Düsterer, L. Fang, J. Feldhaus, H. Fukuzawa, M. Hoener, Y.H. Jiang, P. Johnsson, E.T. Kennedy, M. Meyer, R. Moshammer, P. Radcliffe, M. Richter, A. Rouzae, A. Rudenko, A.A. Sorokin, K. Tiedtke, K. Ueda, J. Ullrich and M.J.J. Vrakking, *J. Mod. Opt*. **57** (2010) 1015–1040, https://doi.org/10.1080/09500340.2010.487946.

[625] G. Doumy, C. Roedig, S-K. Son, C.I. Blaga, A.D. DiChiara, R. Santra, N. Berrah, C. Bostedt, J.D. Bozek, P.H. Bucksbaum, J.P. Cryan, L. Fang, S. Ghimire, J.M. Glownia, M. Hoener, E.P. Kanter, B. Krassig, M. Kuebel, M. Messerschmidt, G.G. Paulus, D.A. Reis, N. Rohringer, L. Young, P. Agostini and L.F. DiMauro, *Phys. Rev. Lett*. **106** (2011) 083002–083006.

[626] M. Richter, S.V. Bobashev, A.A. Sorokin and K. Tiedtke, *J. Phys. B At. Mol. Opt. Phys*. **43** (2010) 194005–194012.

[627] A.H. Zewail, *J. Phys. Chem. A* **104** (2000) 5660–5694, https://doi.org/10.1021/jp001460h.

[628] M. Meyer, J.T. Costello, S. Düsterer, W.B. Li, and P. Radcliffe, *J. Phys. B At. Mol. Opt. Phys*. **43** (2010) 194006–194015.

[629] C.M. Günther, B. Pfau, R. Mitzner, B. Siemer, S. Roling, H. Zacharias, O. Kutz, I. Rudolph, D. Schondelmaier, R. Treusch, *et al.*, *Nat. Photonics* **5** (2011) 99–102.

[630] S. Roy, D. Parks, K.A. Seu, Run Su, J.J. Turner, Weilun Chao, E.H. Anderson, Stefano Cabrini and S.D. Kevan, *Nat. Photonics*. **5** (2011) 243–245, https://doi.org/10.1038/nphoton.2011.11.

[631] J.M. Glownia, J. Cryan, J. Andreasson, A. Belkacem, N. Berrah, C.I. Blaga, C. Bostedt, J. Bozek, L.F. DiMauro, L. Fang, J. Frisch, O. Gessner, M. Gahr, J. Hajdu, M.P. Hertlein, M. Hoener, G. Huang, O. Kornilov, J.P. Marangos, A.M. March, B.K. McFarland, H. Merdji, V.S. Petrovic, C. Raman, D. Ray, D.A. Reis, M. Trigo, J.L. White, W. White, R. Wilcox, L. Young, R.N. Coffee and P.H. Bucksbaum, *Opt. Express* **18** (2010) 17620–17630.

[632] E. Galtier, F.B. Rosmej, T. Dzelzainis, D. Riley, F.Y. Khattak, P. Heimann, R.W. Lee, A.J. Nelson, S.M. Vinko, T. Whitcher, J.S. Wark, T. Tschentscher, S. Toleikis, R.R. Faustlin, R. Sobierajski, M. Jurek, L. Juha, J. Chalupsky, V. Hajkova, M. Kozlova, J. Krzywinski and B. Nagler20 $fs92\,eV$, *Phys. Rev. Lett*. **106** (2011) 164801–164806, https://doi.org/10.1103/physrevlett.106.164801.

[633] M.-E. Couprie and J.-M. Filhol, *C. R. Phys.* **9** (2008) 487–506, https://doi.org/10.1016/j.crhy.2008.04.001.

[634] M. E. Couprie, *C. R. Acad. Sci. IV Phys*. **1** (2000) 329–345.

[635] A.A. Lutman, R. Coffee, Y. Ding, Z. Huang, J. Krzywinski, T. Maxwell, M. Messerschmidt and H-D. Nuhn, *Phys. Rev. Lett*. **110** (2013) 134801, https://doi.org/10.1103/physrevlett.110.134801.

[636] A. Marinelli, A.A. Lutman, J. Wu, Y. Ding, J. Krzywinski, H-D. Nuhn, Y. Feng, R.N. Coffee and C. Pellegrini, *Phys. Rev. Lett*. **111** (2013) 134801, https://doi.org/10.1103/physrevlett.111.134801.

[637] T. Hara, Y. Inubushi, T. Katayama, T. Sato, H. Tanaka, T. Tanaka, T. Togashi, K. Togawa, K. Tono, M. Yabashi, *et al*., *Nat. Commun*. **4** (2013) 2919, https://doi.org/10.1038/ncomms3919.

[638] A. Marinelli, D. Ratner, A.A. Lutman, J. Turner, J. Welch, F-J. Decker, H. Loos, C. Behrens, S. Gilevich, A.A. Miahnahri, *et al*., *Nat. Commun.* **6** (2015) 6369, https://doi.org/10.1038/ncomms7369.

[639] A. Petralia, M.P. Anania, M. Artioli, A. Bacci, M. Bellaveglia, M. Carpanese, E. Chiadroni, A. Cianchi, F. Ciocci, G. Dattoli, *et al*., *Phys. Rev. Lett*. **115** (2015) 014801, https://doi.org/10.1103/physrevlett.115.014801.

[640] T. Tajima and J.M. Dawson, *Phys. Rev. Lett*. **43** (1979) 267–270.

[641] E. Esarey, C.B. Schroeder and W.P. Leemans, *Rev. Mod. Phys*. **81** (2009) 1229, https://doi.org/10.1103/revmodphys.81.1229.

[642] V. Malka, J. Faure, Y.A. Gauduel, E. Lefebvre, A. Rousse and K.T. Phuoc, *Nat. Phys*. **4** (2008) 447–453, https://doi.org/10.1038/nphys966.

[643] K. Nakajima, *Nat. Phys*. **4** (2008) 92–93, https://doi.org/10.1038/nphys846.

[644] F. Grüner, S. Becker, U. Schramm, T. Eichner, M. Fuchs, R. Weingartner, D. Habs, J.M. Vehn, M. Geissler, M. Ferrario, *et al*., *Appl. Phys. B* **86** (2007) 431–435, https://doi.org/10.1007/s00340-006-2565-7.

[645] H-P. Schlenvoigt, K. Haupt, A. Debus, F. Budde, O. Jäckel, S. Pfotenhauer, H. Schwoerer, E. Rohwer, J.G. Gallacher, E. Brunetti, R.P. Shanks, S.M. Wiggins and D.A. Jaroszynski, *Nat. Phys*. **4** (2008) 130–133, https://doi.org/10.1038/nphys811.

[646] M. Fuchs, R. Weingartner, A. Popp, Z. Major, S. Becker, J. Osterhoff, I. Cortrie, B. Zeitler, R. Hörlein, G.D. Tsakiris, U. Schramm, T.P. Rowlands-Rees, S.M. Hooker, D. Habs, F. Krausz, S. Karsch and F. Grüner, *Nat. Phys*. **5** (2009) 826–829, https://doi.org/10.1038/nphys1404.

[647] M.P. Anania, D. Clark, S.B. van der Geer, M.J. de Loos, R. Isaac, A.J.W. Reitsma, G.H. Welsh, S.M. Wiggins and D.A. Jaroszynski. In *SPIE Europe Optics+ Optoelectronics* (International Society for Optics and Photonics, 2009), pp. 735916–735916.

[648] G. Lambert, S. Corde, K. Ta Phuoc, V. Malka, A. Ben Ismail, E. Benveniste, A. Specka, M. Labat, A. Loulergue, R. Bachelard and M-E. Couprie. In *Proc. FEL conf., Nara, Japan*, Jacow, (2012), p. 2.

[649] A.R. Maier, A. Meseck, S. Reiche, C.B. Schroeder, T. Seggebrock and F. Gruener, *Phys. Rev. X* **2** (2012) 031019, https://doi.org/10.1103/physrevx.2.031019.

[650] M.-E. Couprie, A. Loulergue, M. Labat, R. Lehe and Victor Malka, *J. Phys. B At. Mol. Opt. Phys*. **47** (2014) 234001, https://doi.org/10.1088/0953-4075/47/23/234001.

[651] A. Loulergue, M. Labat, C. Evain, C. Benabderrahmane, Victor Malka and M.E. Couprie, *New J. Phys*. **17** (2015) 023028, https://doi.org/10.1088/1367-2630/17/2/023028.

[652] Z. Huang, Y. Ding and C.B. Schroeder, *Phys. Rev. Lett*. **109** (2012) 204801–204805,

https://doi.org/10.1103/physrevlett.109.204801.

[653] C. Townes, *A Century of Nature, 21 Discoveries that Changes Science in the World*, The University of Chicago Press Books (2003).

Appendices

This appendix gives the list of FEL Prize winners.

Table .1: FEL Prize winners

Year	Town	Country	FEL	Prize winners
1988	Jerusalem	Israel	FEL 10	John Madey
1989	Naples	USA	FEL11	William Colson
1990	Paris	France	FEL12	Todd Smith and Luis Elias
1991	Santa Fe	USA	FEL13	Phillip Sprangle and Nikolai Vinokurov
1992	Kobe	Japan	FEL14	Robert Phillips
1993	The Hague	The Netherland	FEL15	Roger Warren
1994	Stanford	USA	FEL16	Alberto Renieri and Giuseppe Dattoli
1995	New York	USA	FEL17	Richard Pantell and George Bekefi
1996	Rome	Italy	FEL18	Charles Brau
1997	Beijing	China	FEL 19	Kwang-Je Kim
1998	Williamsburg	USA	FEL 20	John Walsh
1999	Hamburg	Germany	FEL21	Claudio Pellegrini
2000	Durham	USA	FEL 22	Stephen V. Benson, Eisuke J. Minehara and George I
2001	Darmstadt	Germany	FEL 23	Michel Billardon, Marie-Emmanuelle Couprie and Jean-Michel Ortega
2002	Argonne	USA	FEL24	H. Alan Schwettman and Alexander F.G. van der M
2003	Tsukuba	Japan	FEL25	Li-Hua Yu
2004	Trieste	Italy	FEL26	Hiroyuki Hama and Vladimir Livinenko
2005	Stanford	USA	FEL27	Avraham Gover
2006	Berlin	Germany	FEL28	Evgeni Saldin and Jorg Rosbach
2009	Liverpool	Great Britain	FEL29	Paul Emma and David Dowell
2010	Malmö	Sweden	FEL30	Sven Reiche
2011	Shanghai	China	FEL31	Tsumoru Shintake
2012	Nara	Japan	FEL32	John Galayda
2013	New York	USA	FEL33	Luca Giannessi and Young Uk Jeong
2014	Basel	Switzerland	FEL34	William Fawley and Zhirong Huang
2015	Daejeon	Korea	FEL35	Mikhail Yurkov and Evgeny Schneidmiller
2017	Santa-Fe	USA	FEL36	Bruce Carlsten, Dinh Nguyen and Richard Sheffi

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Lasers in FEL Facilities

M. Divall
PSI, Villigen, Switzerland

Abstract
This paper gives an overview about where and how conventional lasers are used in Free-Electron Laser (FEL) facilities, covering lasers utilized for timing system distribution, electron injection, beam treatment and diagnostics, as well as in experimental stations. It starts with a short introduction to the necessary laser terminology. The required laser parameters for each application are shown, especially highlighting the specifications, which are state-of-the-art. Examples from FEL facilities are listed. Finally, the future for the building of compact FELs is discussed.

Keywords
Lasers; optical timing distribution; photo-injector; pump-probe experiments.

1 Introduction to lasers in layman's terms

In this section I focus on the basic building blocks of the laser system. The aim is to give a pointer towards the choices made in the scope of a Free-Electron Laser (FEL). More details can be found about lasers in general in [1].

A laser is a device that emits light through a process of optical amplification based on the stimulated emission of electromagnetic radiation. The term 'laser' originated as an acronym for 'light amplification by stimulated emission of radiation' [2]. The main difference between a conventional laser and a free-electron laser is that, in the former, electron are bound into the structure of a material, a lasing medium; while in the FEL the 'lasing medium' consists of very-high-speed electrons moving freely through a magnetic structure, and hence they are free. In the former the lasing properties will be determined by the material and its energy levels and other physical properties, while in the FEL, by the properties of the incoming electrons and the magnetic field structure they encounter.

Einstein postulated that photons prefer to travel together in the same state [3]. If one has a large collection of atoms containing a great deal of excess energy or, so to say, in an excited state, they will randomly emit a photon. If a photon of the correct wavelength passes through this excited material, its presence will stimulate the atoms to release their photons early – and those photons will travel in the same direction with the identical frequency and phase as the original stray photon. A cascading effect will start where these identical photons move through the rest of the material. As a result, ever more photons will be emitted from their atoms to join them, and an amplification process will take place. The observation was intuitive, but proven to be correct. This is a bit like announcing an FEL School and hence exciting likeminded people to gather together in the same place at the same time. May great things come out of such events.

1.1 The choice of the laser material

The choice of the laser material is determined by many different parameters. We would like our laser to emit at a specific wavelength and therefore we look for materials with a given emission band. This will automatically bring an absorption band, where we would like to find a matching pump source, which will bring our material to an excited state in the most efficient way, as shown in Fig. 1.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.327

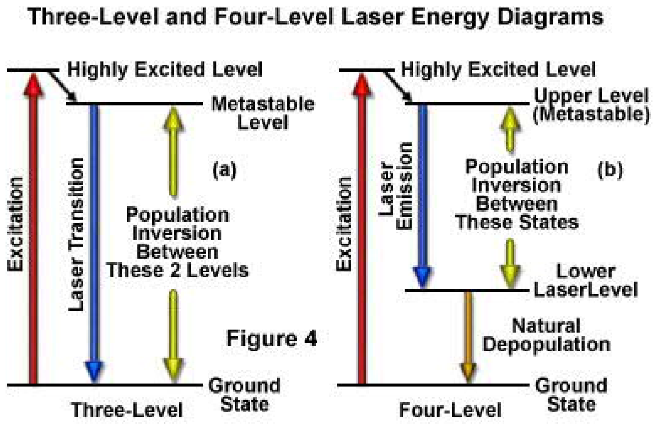


Fig. 1: The process of excitation and stimulated emission in (a) three-level systems; and (b) four-level systems [4]

The pulse shape and length and rise time are important for low emittance machines for FEL. When producing ultra-short laser pulses, below 100 fs, ultra-broadband emission is required, where the choice of materials is limited. Often, a second laser source is required for pumping. As in most cases a solid crystal material is used for such applications, we will limit ourselves to solid-state lasers. Some FELs, such as the European XFEL, generate a long train of electron bunches that are induced by the laser. Here, the fluorescence lifetime of the upper laser level will limit the length of the train, which can be produced with identical laser parameters. The stimulated emission cross-section and the gain properties of the material will determine the necessary amplification stages and hence the complexity of the laser system. The thermal properties of the crystal will determine the scalability of the system to high average powers as well as the quality of the beam.

As in FELs, synchronizability to the RF accelerating structures is required: the laser system starts with a laser cavity, called the oscillator. The repetition rate will limit the available materials as well as the pulse energy, often calling for several stages of amplification. The so-called saturation fluence determines how much energy can be extracted from a given volume of the material. Hence, the available crystal sizes and their thermal fracture limit will determine the maximum fluence that can be extracted efficiently from a single amplification stage. The difference between the pump and the emission band determines the proportion of the power converted into thermal losses, and will limit the average power or complicate the cooling system requirements.

Table 1 shows for the properties of the three most typically used materials for solid-state lasers,. highlighting the parameters where a material excels compared to the others [5].

Table 1: Examples of the different amplifier materials and their properties

	Crystals		
	Nd:YAG	**Yb:YAG**	**Ti:Saph**
Fluorescence lifetime [ms]	0.23	**0.96**	0.0032
Stimulated-emission cross-section [× 10−20 cm^{-1}]	20 to 30	2.1	30
Lasing wavelengths [nm]	1064	1030	660 to 1100
Absorption wavelengths [nm]	**808**	**941**	514 to 532
Fluorescence bands (FWHM) [nm]	0.67	to 10	440
Absorption bands (FWHM) [nm]	1.9	>10	200
Pumping quantum efficiency	0.76	**0.91**	0.55
Saturation fluence [J/cm^2]	0.67	9.2	0.9

For picosecond pulse generation the materials most used are neodymium-doped, such as Nd:YLF, $Nd:YVO_4$ and Nd:YAG. The advantage of these materials is that the upper laser level lifetime is relatively long, in the hundreds of microsecond range, which allows for efficient energy storage, as well as extraction of high repetition rate pulse trains, as opposed to single pulse generation. Furthermore, these materials can be pumped directly with diodes at ~800 nm. These diodes are available relatively cheaply, and also in stacks, which brings down the overall cost of the amplifiers.

For femtosecond pulse generation the most widely known material is titanium sapphire (Ti:Saph). The amplification bandwidth of this material allows for sub-10 fs pulse generation, which can be further amplified from the tens of millijoule level to sub-100 fs, and it is readily available on the market. The large bandwidth allows for manipulation of the pulse shape, which is often required for low emittance sources. The drawback is that it is not suitable for long pulse trains due to the short fluorescence lifetime. The absorption band (~500 nm) also falls outside the widely available pump diodes' range and therefore frequency-doubled Nd:YAG lasers are used to pump these systems, which increases cost and complexity.

Lasers based on ytterbium-doped materials (Yb:YAG, Yb:KGW, Yb:glass) provide a good compromise. They allow for sub-picosecond pulses and diode pumping, as well as for fibre-based oscillators and pre-amplifiers, making the system more robust and simple. Since a part for the system could be in fibre, the beam profile is also improved substantially.

The pumping source to achieve excitation of the upper laser level can vary, depending on the crystal geometry, the wavelength of absorption and the required repetition rate. Pumping by diodes is by far the most efficient and cost-effective solution. Diodes in the 800 nm to 950 nm range are readily available in diode-stack options, providing 1 kW/cm^2 mean power up to kilohertz repetition rates. The lifetime is estimated to be ~10^9 shots, which for a 100 Hz system would provide half a year of non-stop operation. When power is reduced to below nominal, the lifetime could be further extended. Flash-lamps are cheaper, but only provide a few weeks of operation at 100 Hz, and constant degradation often leads to other laser parameters changing between maintenance. When other lasers are used for pumping the main amplifier chain it can lead to very high energy broadband systems, but at the cost of inefficiency and complexity.

So the material is chosen after taking into consideration the pulse length, pulse structure, wavelength and maximum pulse energy required. In most cases a compromise needs to be taken between flexibility and robustness. As much of the system as possible should be acquired commercially, from industrial sources.

1.2 The architecture of the laser

The first element of a laser-chain is the oscillator. The oscillator has to provide the proper input wavelength for the rest of the amplifier chain, and the active material is often chosen to be the same as that of the amplifiers.

Figure 2 shows a schematic for an oscillator, containing a continuously pumped gain medium as well as a pulse-shortening and mode-locking device. What is unique to applications in FELs is that the laser pulses have to arrive synchronized at the RF accelerating cavities; the injector is synchronized to the electron bunches for the diagnostics, and to the FEL pulses at the experiments. This requires specific frequencies for the oscillator as well as an actively stabilized cavity to maintain the timing. The specific macro-pulse structure will also have an influence on the repetition rate.

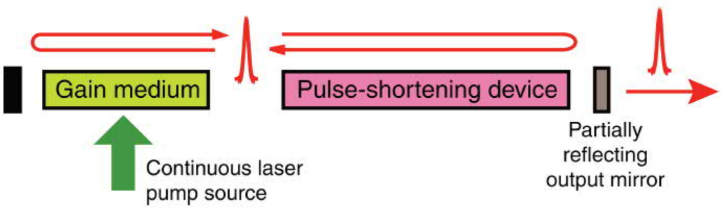
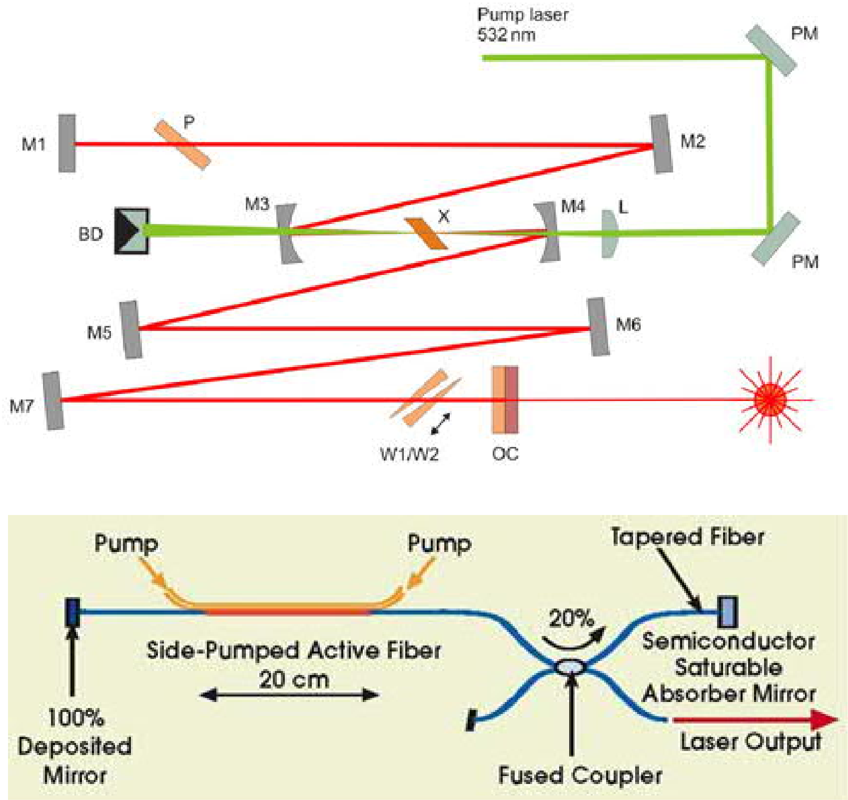


Fig. 2: Schematic of an oscillator (http://slideplayer.com/slide/5110070 slide 6)

The pulses that are not needed in the machine are discarded, using acousto- or electro-optical switches, and the pulses are then further amplified. Figure 3(a) shows an example for the cavity design of a Ti:Saph laser, pumped by an external laser source and complex architecture to maintain the broad bandwidth, providing pulses in the tens of femtoseconds range, commercially available. In contrast Fig. 3 (b) illustrates the simplicity of a diode pumped Yb fibre oscillator [6], providing unmatchable high power 10 ps pulses directly from the oscillator. For synchronization purposes at the 10s of fs level the pump-laser power noise, the mechanical stability, straight-light and electrical noise can all have an effect.



(a)

(b)

Fig. 3: Example of a (a) 4 fs Ti:Saph oscillator (http://www.iqo.uni-hannover.de/591.html); (b) ytterbium fibre laser.

As the non-linearities limit the energy that can be reached directly in an oscillator, the pulse energy below nanojoule level will require further amplification. The amplification can take place in different architectures. Figure 4(a) shows a multi-pass amplifier, most typically used for pre-amplification and the final amplifier stages. This arrangement makes it possible to amplify several pulses or pulse trains. The regenerative amplifier shown in Fig. 5(b) is recirculating a single pulse inside the amplifier, allowing for high gain and efficiency, but is only suitable for single pulse amplification.

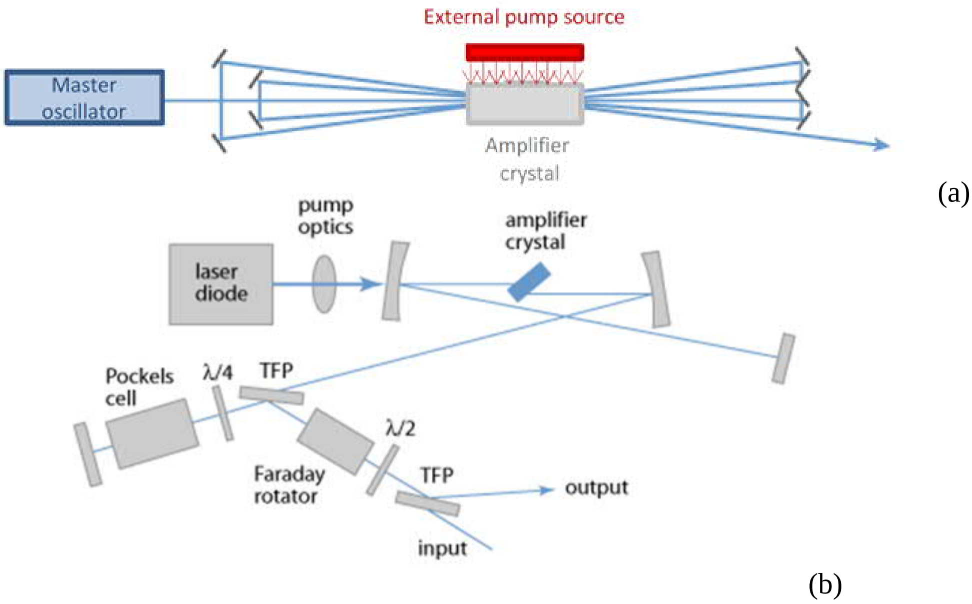


Fig. 4: (a) Multi-pass amplification; (b) regenerative amplifier

All of these amplifier types can be combined with the so-called chirped pulse amplification (CPA), shown in Fig. 5. To avoid damage in the amplifier due to the high peak intensity in a single short pulse, the laser pulse is stretched after the oscillator or pre-amplifier stage and only compressed again after the required energy is reached [7, 8].

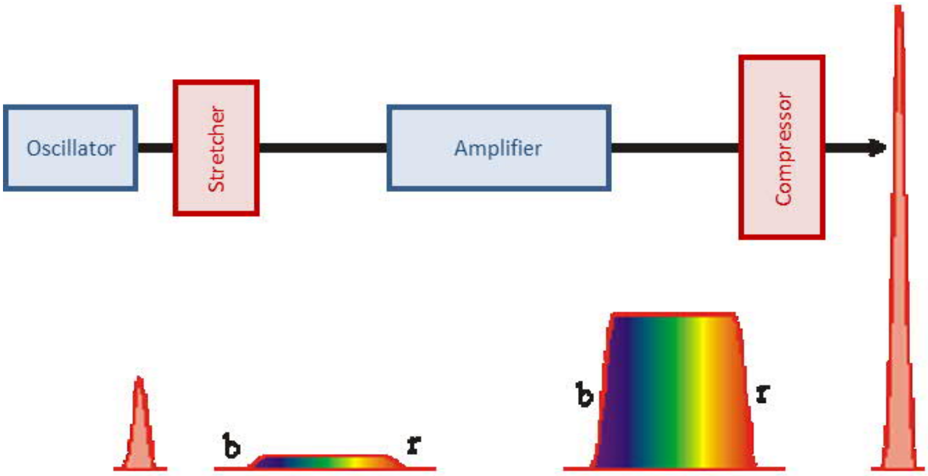


Fig. 5: Basic layout of a chirped pulse amplifier (CPA)

This is possible due to the high bandwidth of the laser, which allows for controlled spectral dispersion, or chirp. These types of amplifiers are able to reach Travelling Wave (TW) levels in a standard Ti:Saph system. The mechanical design of these systems is very important, as are the total amplification path and the alignment through the stretcher and the compressor. Another term needs to be clarified at this point. The layout of a so-called master oscillator power amplifier is shown in Fig. 6. There are an increasing number of FELs where the aim is to provide several pulses in a train, with the same properties. They should therefore be able to distribute FEL pulses between experiments, as well as sample processes that are on the timescale of the laser pulse train.

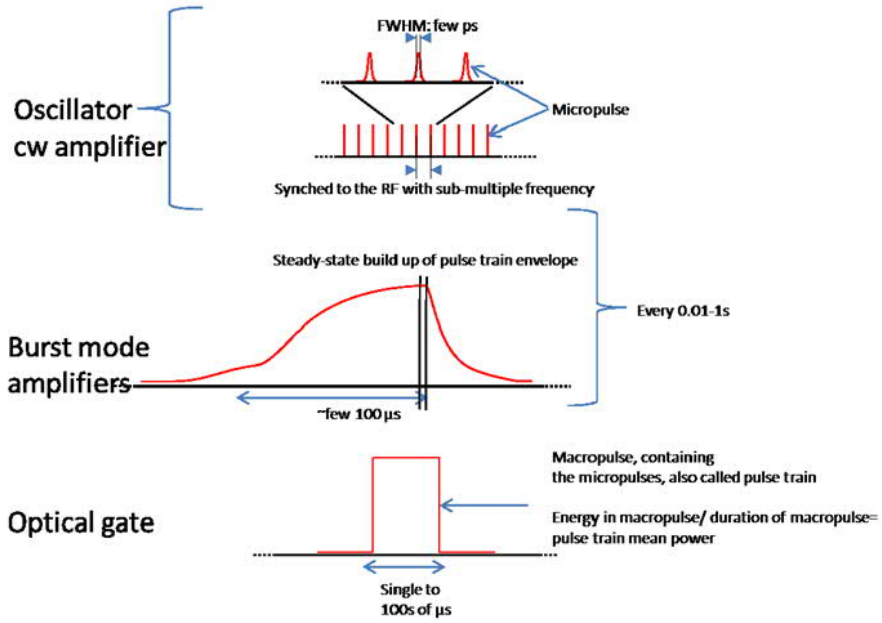


Fig. 6: Master oscillator power amplifiers for burst mode operation

In principle, oscillators allow for GHz pulse-train generation, but this often results in an extremely low energy single pulse at the front end of the system, making it difficult to maintain constant pulse characteristics throughout the amplification process. Master oscillator power amplifiers typically use a pulse train in the 100 MHz to 250 MHz repetition rate region. The pulse train is selected at the beginning of the amplification process. The amplifiers are operated in a burst mode and the train, containing pulses with identical parameters, is selected at the end. Thermal management of such systems can be difficult.

Last, but not least, we need to mention optical parametric amplifiers (OPA). These types of lasers are just becoming commercially available with the new European projects, such as ELI. Here the amplification takes place in a non-linear optical material, using a parametric process. The energy is instantaneously transferred from the shorter wavelength pump-light to the signal beam, using the non-linear properties of the material and taking advantage of the available high intensities. The transfer is instantaneous: there is no energy storage. To avoid damage and unwanted non-linear effects and to efficiently match and overlap with the length of the pump pulse, OPA can be combined with the abovementioned CPA technique, where pulses are stretched prior to amplification. These are called Optical Parametric Chirped Pulse Amplifiers (OPCPA). OPAs can cover a wide range of wavelengths from 300 nm to 4 µm. Most of the commercially available tunable wavelength sources are based on parametric oscillators or amplifiers. State-of-the-art OPCPA systems can generate 20 fs pulses with 1 J of energy and provide a focused intensity around 10^{20} W/cm^2, well exceeding the barrier for relativistic optics at 10^{18} W/cm^2 and giving the onset of non-linear effects and ionization.

The material is often chosen for its broad bandwidth or gain properties. Most efficient laser systems are working in the IR region for efficient pumping and energy storage, as well as for longer lifetimes. Applications often require light in the visible (experiments) or UV (injectors) ranges, which is achieved by converting the output wavelength, using non-linear crystals. Ultra-intense pulses can interact with non-linear or gaseous materials to extend their wavelength range from XUV to terahertz radiation. High order harmonics down to the XUV regime and with pulse durations inherently shorter than the drive pulse can be created, which opens up new application fields of attosecond science [10].The scope of this paper does not allow for detailed expansion on this subject, but recommended reading is found in [11, 12].

Finally for some applications one might require treatment of the pulse after amplification, including spatial and longitudinal shaping. These include the use of adaptive optics, broadening techniques in solid or gas, carrier-envelope stabilization for a few femtosecond pulses and programmable pulse trains. The specific techniques will be highlighted for applications with examples.

2 Lasers used in FELs, requirements, parameters and examples

FELs, based on LINACs and undulators, have been making their way to shorter wavelengths and a higher degree of coherence over the past decade, and at the 1 kV photon energy range they currently provide the brightest beams. This requires extremely high quality short electron bunches to be produced, with ultralow emittance.

Conventional lasers take their place all over the machine, as Fig. 7 illustrates. Fibre lasers and distribution systems are now routinely used in short-pulse FELs as the optical clock to provide the heart of the whole machine. These lasers can be locked to an external microwave crystal oscillator to achieve ultra-low reference noise and are distributed re-amplified and re-synchronized, achieving kilometre-long stabilized optical links [13, 14]. An overall stability of 10 fs can be routinely achieved in these systems. Photo-injection (PI) is used as the electron source, where electron bunches are produced through photo-emission from a cathode irradiated by a laser and are accelerated to a few MeV by the field present in the gun [15]. A laser heater is used to smooth out micro-bunching instabilities by inducing uncorrelated modulation to the bunch at a low energy [16, 17]. Lasers are an essential part of a FEL facility, not just by providing the primary electron source, but also by delivering ultra-short synchronized laser pulses for pump-probe measurements. As electron bunches are becoming shorter and shorter in the quest to produce XUV FEL pulses, conventional diagnostics devices are replaced by laser-based diagnostics, such as a laser wire to measure bunch cross-section [18], electro-optical bunch length and bunch arrival measurement [19–22]. These systems do not just provide superior accuracy in the femtosecond region, but are also non-invasive, by using laser fields for detection. Short wavelength FELs don't allow for amplification of the FEL pulse in cavities and therefore radiation is built up from self-amplification of spontaneous radiation from the undulators. As one seeds the undulators with a specific sub-harmonic wavelength, a higher gain and better beam quality can be reached in the same undulator length. Deep-UV and high-harmonic generation laser systems are used here for seeding the FEL [23–26].

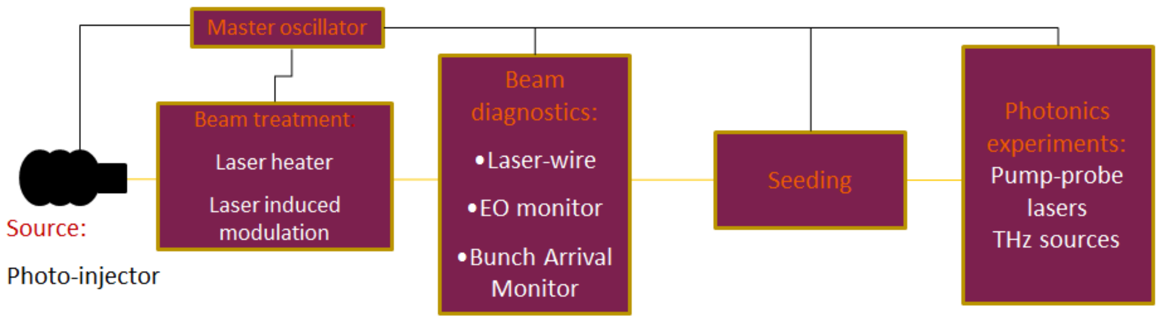


Fig. 7: Lasers used in FELs

As the stability and reliability of ultra-short pulse lasers is improved towards these applications, lasers are taking their own territory in accelerators and FELs by providing overall solutions for acceleration and beam production as well, aiming toward compact FELs.

All applications together require wavelengths from few a nanometres to terahertz, pulse energies from picojoules to multi-megajoules and pulse lengths from a single optical cycle to many picoseconds. Most of the time, additional flexibility is required, to cater for short- and long-pulse modes, and different pulse structures and wavelengths from the FEL. This calls for tunability in beam size, pulse length and sometimes also in the wavelength of the laser. It is not possible to cover all this within the scope of this paper, but in the following sections some examples will be brought to illustrate the main parameters, requirements and challenges for each application.

2.1 Timing of the lasers

One of the requirements general to all lasers used in an FEL facility is the timing of the laser to the reference and to the entire machine. Figure 8 shows the schematic of the timing distribution over SwissFEL with all the time-of-arrival measurement and correction points. An optical master oscillator, based on an ytterbium fibre laser from OneFive, is operating at the low dispersion wavelength at 1550 nm [27]. The 200 fs long pulses at 142.8 MHz, which is the 21st sub-harmonic of the S-band RF, is distributed around the machine by stabilized fibre links. The RF distribution, the photo-injector laser [28], the laser-based beam arrival monitors and the experimental lasers are all synchronized to the clock pulses.

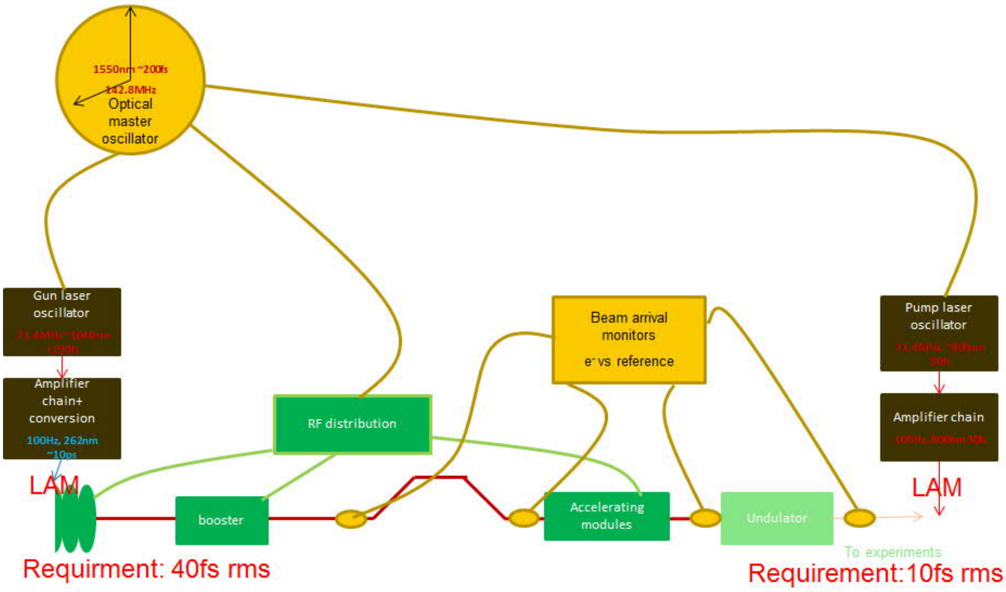


Fig. 8: Schematic of timing distribution for SwissFEL

At the injector the arrival time of the laser will determine the arrival time to the initial chicane and effect compression to produce shorter bunches. This in turn will influence the overall FEL lasing performance [29]. For the experimental lasers the users would ideally like to see their pump-probe experiment online, with the FEL and laser pulses arriving at a well-controlled delay at the femtosecond level. This problem is currently bridged by measuring the laser arrival time relative to the FEL at the experiment and binning the data afterwards for analysis. The ultimate aim is to apply a feedback system, which corrects the lasers' arrival over time. This is aided by locking the laser oscillator to the distributed reference. Large, ultra-short pulse laser systems include a long chain of optical components, however. As an example the Ti:Saph laser in the SwissFEL test facility [30, 31] includes 3 EO switches, 2 acousto-optical pulse shapers, 140 reflective surfaces and more than 7 crystals, and amounts to a total of 86 m of propagation path until it reaches the cathode for electron production. To reach the required 40 fs accuracy this length has to be stabilized to 12 μm accuracy. Mechanical vibrations, air-flow, electrical noise in the switches, thermal distortions and fluctuations, both in the cooling system for the crystals and in the environment, have detrimental effects on the arrival time. In a typical regenerative amplifier the expansion of the baseplate at 0.1°C change can induce drift at the 100 fs scale. Air temperature and pressure changes over long distances have the same effect. Therefore the direct control of the laser timing is necessary to maintain synchronicity with the rest of the system. Slow temperature-related drifts can easily be controlled with feed-forward systems, while fast variations related to vibrations are best passively tackled by design. Using kilohertz repetition rates and fast feedback, some of these noise sources can also be damped. Laser- and beam-arrival monitors along the system are used to stabilize the subsystems. The optical clock becomes very useful, as direct comparison of these pulses with the laser pulses optically allows for measurements at 100 Hz.

2.2 Photo-injector lasers

In a photo-injector the electrons are produced by the photo-emission process. A photo-cathode is illuminated by laser pulses and the electrons are extracted by a high electric field inside the gun [15]. PI allows for approximately two orders of magnitude higher brightness than a conventional thermionic source, where electrons are produced by heat. Short pulses with 4D shaping have the potential to produce ultra-bright electron beams. Apart from the high particle beam quality achievable from these sources they also take advantage of the wide range of pulse length, repetition rates and pulse train structures that can be generated with laser oscillators, amplifiers and optical gating systems described in the previous section.

The main considerations for laser for photo-injectors include:

i) Wavelength (tuned close to the work-function of the material):
   – Cs_2Te and Cu require UV ~260 nm;
   – GaAs needs IR and gives polarized electrons;
   – New type of alkaline cathodes, requiring visible light;

ii) Timing:
   – The laser has to be synchronized to external reference/sub-harmonic of the RF;
   – Noise of the oscillator architecture, active elements need to support this;

iii) Single pulse/burst mode to match machine operation:
   – Determines the chosen laser architecture (regenerative or multi-pass amplifier);

iv) 3D–4D shaping to reduce emittance:
   – Beer-can/truncated Gaussian/ellipsoidal;

v) Pulse length:
   – To mitigate space-charge effects at the gun;
   – To allow for shaping techniques;

vi) Pulse energy:
   – Dependent on cathode choice/quantum efficiency and operational charge plus transport losses;

vii) Reliability/reproducibility/stability:
   – Architecture used;

viii) Running cost/service support.

When specifying a laser for a certain injector the cathode material will have a certain bandgap, which will determine the required laser wavelength. The relationship between the produced charge and the laser parameters can be described by the following simple equation

$$C\,[\mathrm{nC}] = 8 \cdot QE\,[\%] \cdot W\,[\mu\mathrm{J}] \cdot \lambda\,[\mathrm{nm}] \tag{1}$$

where C is the produced charge in nC, QE is the quantum efficiency of the cathode as a percentage, W is the energy/micropulse in μJ and λ is the wavelength of the laser in nm (8 includes elementary charge, the Plank constant and speed of light).

Copper is usually used, when a charge below 1 pC is required for its robustness, long lifetime and easy production. This requires pulse energies in the tens of microjoule range. Alkaline cathodes, most typically Cs_2Te, are used for a higher charge at a cost of degrading efficiency and shorter lifetime, as well as more strict vacuum requirements. As the quantum efficiency is orders of magnitude higher, the laser energy can be in the nanojoule range, which also allows pulse trains to be delivered to the cathode. Both of these cathodes operate at a wavelength ~260 nm, which means fourth harmonic generation from the Nd and Yb doped lasers or third harmonic from a Ti:Saph laser. Producing and propagating UV light to the gun can be challenging due to losses or the conversion process and the degradation of the beam while propagating through air. Often, part of the beam transport has to be in vacuum.

As an example, Fig. 9 shows the SwissFEL photo-injector laser system [28]. The SwissFEL gun is a 2.5 cell S-band cavity, running at 100 Hz repetition rate. The peak acceleration is 120 MV/m and the required charge is up to 200 pC with the option to reduce down to a few picocoulombs for ultra-short FEL pulses. To maintain the space-charge conditions this in turn requires the scaling of the laser beam cross-section (0.1 nm to 0.27 mm) and pulse-length (4 ps to 10 ps) to match the charge. The normalized emittance requirement is 0.275/0.114 mm for the 100 pC and the 10 pC operational modes. To reach these values shaping of the laser pulse longitudinally and transversally to flat top is necessary. Both copper and Cs_2Te cathodes are used, therefore a large energy range has to be covered, up to 60 μJ. The stability is also crucial at the gun, as with single pulse operation the noise induced here cannot be reduced by feedback. The energy stability in the UV has to be below 0.5% RMS and the timing jitter below 40 fs. To reach these requirements the following architecture was chosen. The front-end oscillator is an ultra-low noise compact Yb fibre oscillator operating at 1040 nm. The oscillator is synchronized to the optical reference using RF locking to less than 30 fs integrated jitter. This can be further reduced by optical locking, which is planned for the future. The chirped pulse amplifier is a single-box regenerative amplifier with stabilized base-plates. The amplifier material is $Yb:CaF_2$, which allows short enough pulses for pulse-shaping purposes, but can still be directly diode-pumped. Part of the beam is used in the IR for the laser heater (described in the following section), while the rest of the output is converted to its fourth harmonic.

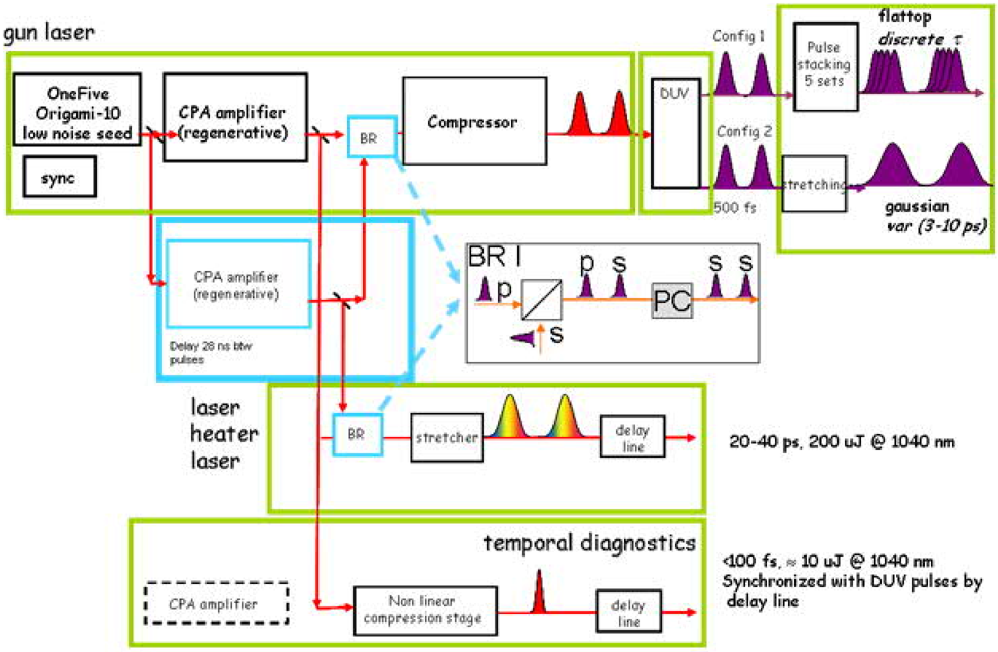


Fig. 9: The SwissFEL photo-injector laser system

Another example is the Nd:YLF system for FLASH [32] (Fig. 10). What makes this system more complicated is that there is a programmable train of pulses required at the cathode. This means that the single-box regenerative amplifier has to be replaced by multi-pass amplifiers, which run in a steady-state configuration to provide equal gain for each pulse in the train. After pre-amplification the pulses, which are amplified during the build-up of the gain, are rejected by an electro-optic switch (pulse picker) to produce a pulse train. All amplifiers are pumped by fibre-coupled pump diodes for accuracy, reliability and to remove the heat sources from the laser table.

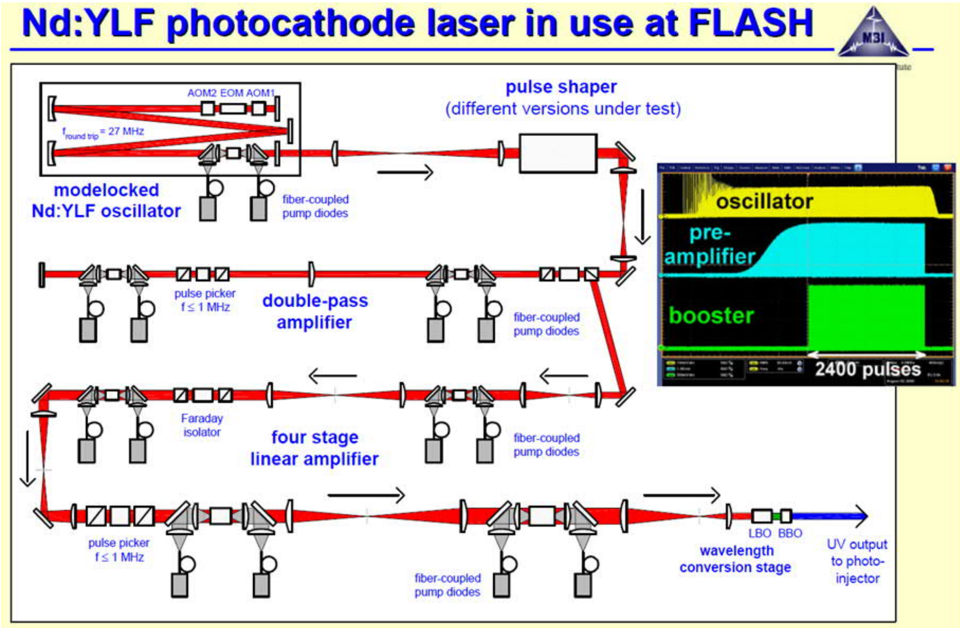


Fig. 10: Nd:YLF injector laser for FLASH

Pulse shaping is a very important aspect for achieving low emittance. Emittance at the gun is

$$\sigma = \sqrt{\sigma^2_{\text{spacecharge}} + \sigma^2_{\text{thermal}} + \sigma^2_{\text{RF}}} \tag{2}$$

The space-charge forces, which are determined by the electron distribution at the exit of the cathode, can be controlled by the laser pulse shape and the cathode response. Thermal emittance will be dependent on the work-function and the laser wavelength. When approaching the bandgap of the cathode material with the wavelength, lower thermal emittance can be achieved, but at the cost of reduced *QE*. The RF field will also make a contribution to the emittance. This has been demonstrated in the SwissFEL Tests Injector, where a Ti:Saph laser with tunable wavelength was converted to its third harmonic to illuminate the cathode. By reducing the wavelength from 267.5 nm to 260 nm the thermal emittance has increased from 584 nm/mm to 626 nm/mm and *QE* increased by ~50% [33]. In a linac with proper focusing, emittance due to linear space-charge force can always be compensated. Hence, the idea that the laser should be shaped to be flat top both in space and time. This is applied in many working machines, such as LCLS, FLASH, SPARC, PITZ and SwissFEL. More information on the techniques used can be found in [34]. A great deal of effort was invested to produce such flat top pulses in time. While the spatial distribution can be achieved by simply aperturing the laser beam and projecting this plane to the cathode, the temporal distribution is harder to achieve. Fourier transformation shows that when sharp edges are created in time they require infinite frequencies, which is not possible in the practical world. Gaussian ‘wings’ are therefore accepted at the rising and falling edges of the pulse. A simple technique applied for UV pulse shaping is a passive, so-called pulse-stacking system, where the pulses are propagated through birefringent crystals, which induce replica pulses with a certain delay. The sum of the pulses then gives a quasi-flat top distribution in time. This, however, suffers from ripples on the top of the pulse. Care needs to be taken that the frequency of the modulation caused by the finite number of pulses and interference between them is not enhancing micro-bunching instabilities in the machine. The technique was further developed at BMI, where the shaping takes place in the IR, directly after the oscillator and the crystals are temperature-controlled to maintain, and also to actively program, the pulse shape [35]. A new approach to shaping the laser pulse into a 'rugby ball' provides even better emittance on paper, but the production and propagation of such a beam in the UV is still challenging. Tests were performed in PITZ in this direction [36]. Finally, Table 2 gives a summary of the different types of photo-injector laser currently in use with FELs [37–49].

Table 2: Photo-injector lasers around the world

	ELSA	**FLASH (FEL)**	**TESLA**	**LCLS**	**ELETT RA**	**European XFEL**	**SwissFEL**
Cathode	K_2CSSb	Cs_2Te	Cs_2Te	Cu	Cu	Cs_2Te	Cs_2Te/Cu
Wavelength on cathode [nm]	532	262	263	253	261	262	260
Pulse length on cathode [ps]	30	4.4	20 square	3 to 20 square	6 to 15	6	4 to 10 square
Material	Nd:YAG	Nd:YLF	Nd:YLF; Nd:glass	Ti:Saph	Ti:Saph	Nd:YLF	Yb:CaF2
Harmonic	2nd	4th	3rd and mixing	3rd	3rd	4th	4th
Macropulse rep. rate [Hz]	1	≤10	10	30 to 120	≤50	1 to 5	≤100
Micropulse rep. rate [MHz]	14.4	27	1	NA	NA	1500	NA
Pulse train length	150 µs	≤800 µs	800 µs	NA	NA	1.3 µs	2 pulses
Pumping[†]	FL	D, FL	FL	L	L	L	D
Energy/pulse IR	10 µJ	300 µJ	200 µJ	25 mJ	15 mJ	5 µJ	2 mJ
Macro-pulse stability	3%	–	3% (<10%)	–	0.8%	1.5% to 3% (<0.5% in IR)	Na
Micro-pulse stability	?	1% to 2%	(<5%)	–	–	–	0.7%

[†]FL, flashlamp; D, diode; L, laser.

2.3 Laser heater

Following the order of the lasers found in the machine, the next is the laser heater [16, 17]. If short pulses are generated in an RF gun with small momentum spread, due to ripples and oscillations in the pulse and their interaction with the accelerating field, the pulses can suffer from so-called micro-bunching instabilities. The laser heater superimposes a polarized laser beam and the electron beam in a properly tuned undulator. This produces a momentum modulation which smeared out in a chicane to obtain the desired momentum spread increase. To match the bunch length and the physical size of the undulator the lasers used for the laser heater are working in the IR range (~0.8 µm to 1 µm). The pulse length needs to be in the tens of picoseconds Gaussian, with cross-section to match or overfill the size of the electron beam in the undulator. The energy required is ~100 µJ. The pulse structure matches the electron beam structure. Usually they can rely on the fundamental radiation from the gun laser system. This has the advantage of the lasers being inherently locked together in timing and having the same pulse structure. As the heater is usually a few tens of metres from the gun, propagating the beam to it from the gun laser room is possible, but sometimes requires active beam-steering and stabilization to get through the undulator.

2.4 Experimental lasers

Laser used for experiments are the most sophisticated and versatile as they have to cover many different applications. Biological studies require visible and ultraviolet light, femtosecond chemistry relies on 800 nm Ti:Saph lasers, molecular studies need mid-infrared radiation, while magnetization studies reach into the terahertz range.

General requirements are synchronizability to the machine, matching of the pulse structure of the FEL, femtosecond accuracy and scanning capability and pulse length at least as short as the FEL's. There are also resonance experiments where wavelength tuning, although with longer pulses, is a requirement.

I bring here as an example the European XFEL laser [50], used for pump-probe and molecular alignment studies (Fig. 11). Because of the burst mode operation of the machine, the laser has been developed specially for this machine. The intra-pulse repetition rate is up to 4.5 MHz with up to 2700 pulses in each burst, delivered at up to 10 Hz repetition rate. At 800 nm, 15 fs to 300 fs long pulses are required with arbitrary pulse pattern selection. The pulses are then converted to different wavelengths. This is all achieved by a fibre–solid-state and OPA system. A Yb fibre oscillator, which is synchronized to the external reference, is further amplified in a CPA system before splitting part of the beam to further amplify it in a 400 W InnoSlab amplifier and 5 kW InnoSlab booster. The pulses here are selected by an electro-optical modulator and compressed from 400 ps to 800 fs. This part of the beam is converted to its second harmonic and is used to pump the OPA stages. A 2 μJ fraction of the 1030 nm booster output is compressed down to 300 fs. This high intensity allows for supercontinuum generation to generate a wider range of wavelengths for the OPA stages, which can be tuned by the crystal angle, as they work in a non-collinear arrangement. The final system delivers up to 2 mJ pulses.

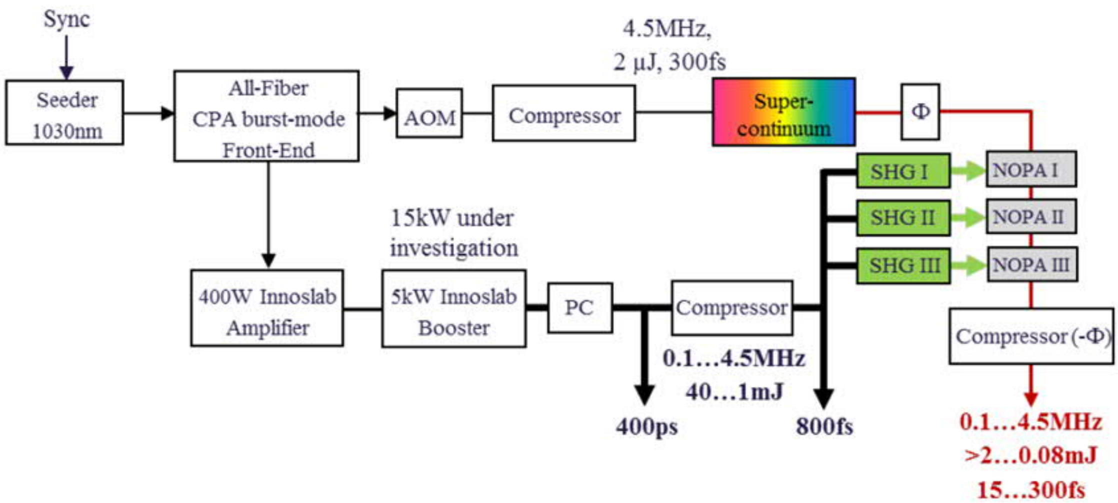


Fig. 11: The experimental laser of the European XFEL

Other experiments around the XFEL require less unique configurations. For the high intensity (HI) and high energy (HE) stations a commercial TW class Ti:Saph system and a kilojoule nanosecond laser are used, respectively.

Another example from SwissFEL is shown in Fig. 12. The core laser system is a commercial Ti:Saph system from Coherent Inc., delivering 20 mJ pulses compressed to <30 fs at 100 Hz. Two identical lasers will be used to ensure redundancy in the case of failure. The lasers will be housed in the room above the experimental stations. The local installation in the experimental station includes a Light Conversion Ltd OPA, delivering pulses from 1.1 mm to 15 mm with ~1 mJ to 10 mJ of energy, respectively. The pulse length is below 100 fs. The output of the OPA will also be used to pump organic crystals for terahertz pulse generation with 1 THz to 10 THz of >1 MV/cm electric field strength in ~10 µJ single-cycle pulses. For pulses shorter than the standard 30 fs output of the main laser system, hollow core fibre, kagome fibre and halogen gas chambers are being investigated. The hall also houses the laser arrival monitor and the THz streak camera, to ensure accurate timing between the laser and the X-rays.

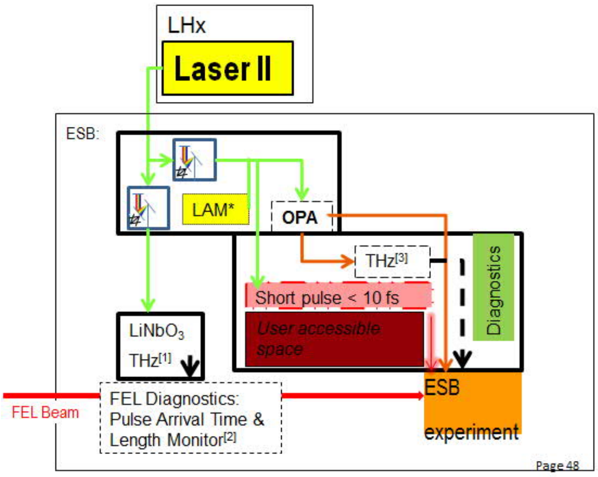


Fig. 12: The laser in the experimental hall of SwissFEL

Finally, mention is made of the requirements for infrared FELs: the FELBE facility is used as an example [52]. For tuneability a range of Ti:Saph lasers are used in conjunction with parametric oscillators and amplifiers, as well as different frequency stages. For the low energy pulses the repetition rates are at 78 MHz, and pulse energies range from 4 nJ to 150 nJ with pulse lengths from 15 fs to 100 fs. Regenerative amplifiers provide higher pulse energies above 200 µJ with 1 kHz repetition rate. In addition, the laser systems are suitable for generating broadband terahertz pulses, which can be used for probing the dynamics excited with FEL pulses (FEL-pump – broadband terahertz probe). All lasers can be synchronized to the FEL.

2.5 Seeding laser

As radiation in an X-ray FEL builds up from the initial noise, there is a lack of temporal coherence in the SASE beam. This can be helped by seeding the FEL with a wavelength that is tuned to a sub-harmonic of the FEL [23–26]. Pulses should also be short and therefore the application of high harmonics from lasers is a natural choice. This way the input signal, the so-called seed, is coherently amplified in the FEL. Seeding was demonstrated at FERMI@Elettra down to 10 nm wavelength.

The advantages of such seeding arrangement are:

- Very high peak flux and higher 6D brightness than with SASE;
- Temporal and transverse coherence of the FEL pulse;
- Control of the time duration, polarization, wavelength and bandwidth of the FEL pulse;
- Inherent synchronization of the FEL pulse to the seed laser, which is also often used for experiments;

- Reduction in undulator length needed to achieve saturation as compared to starting from noise as in SASE FELs.

To achieve efficient FEL seeding a spatial overlap between the electron and laser beams is required, and therefore good pointing stability from the optical laser is needed. The harmonic parameters, such as energy, chirp and wavelength, need to be within the tolerance of the undulator. Temporal overlap between the laser and the electron beam needs to be maintained and the jitter has to be below the sigma bunch length for stable output.

A White Paper was written by the ICFA-ICUIL Joint Task Force to identify high power laser technological needs for accelerators [54]. It was shown, that to reach the required energy levels for seeding at the 30 eV to 0.25 keV range, tens of μJ energies in 100 kHz bursts would be needed. This requires a beast of a 100 GW Ti:Saph laser with a pulse length at 10 fs. Either of these parameters on its own is a challenge. For hard X-rays, similar energy and power levels are necessary, but at a different wavelength, which allows the use of diode-pumped solid-state lasers in conjunction with OPCPA. To develop such lasers at 100 GW levels, however, still requires R&D.

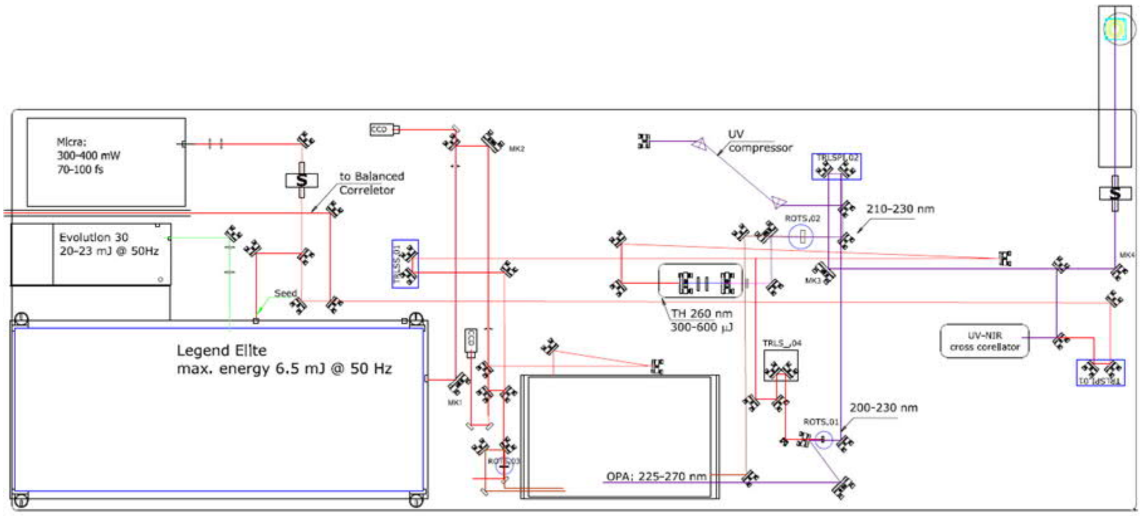


Fig. 13: Seed laser at Fermi@ELETTRA

Figure 13 shows the layout of the Fermi seeding laser for a High Gain Harmonic Generation (HGHG) scheme [55]. It is possible to reach the wavelengths required here with conventional harmonic generation crystals (>200 nm). The system is delivered by Coherent Inc. The Micra Ti:Saph oscillator delivers <100 fs pulses with 400 mW average power. An Evolution 23 mJ green laser provides the pumping at 50 Hz for the amplifier, where a single pulse is recirculated and 6.5 mJ is produced in the infrared. A standard third harmonic generation stage delivers 260 nm, while a tunable OPA gives the seed for the variable UV output stage. Table 3 summarizes the required and achieved parameters.

Table 3: Fermi@ELETTRA seed laser parameter requirements, with the achieved values in brackets

Parameter	Tunable UV	Fixed UV
Tunability range [nm]	210 to 280 (230 to 260)	261 197
Peak power [MW]	100	>400
Pulse duration [fs]	100 (180)	<150 (150 to 500)
Pulse energy stability [RMS, 5000 shots]	<4%	<2%
Timing jitter RMS [fs]	<50 (100)	<50 (100)
Spot in undulator $1/e^2$ [mm]	1	1 to 1.2
Wavelength stability	10^{-4}	$<10^{-4}$
Beam quality [M^2]	<2	<1.5

To seed hard X-rays shorter wavelengths are necessary: 1 nm to 20 nm. Here, high harmonic generation from the laser is used. At high laser intensities the laser field is strong enough to suppress the coulomb barrier and therefore an electron is able to tunnel out of the atom. The laser's electric field then accelerates this electron in a half-cycle. The electron can then be steered back by the opposite electric field from the laser and recombine with the parent ion, while emitting a photon at a higher energy. The high harmonic radiation forms a 'comb', where the energies are separated by half of the drive-laser period, due to the acceleration and recombination process. Such a comb structure, when Fourier-transformed, corresponds to an attosecond pulse train in the time domain. At sFLASH such a scheme was used to seed the FEL, using an off-the-shelf laser system at 800 nm, with 20 mJ laser energy and 35 fs pulses. HGHG targets usually consist of a noble gas and a guiding structure.

Most FELs are now aiming for self-seeding arrangements, where a monochromator is used between subsequent undulators and the residual beam is used for seeding [56, 57]. The process, however, is much more lossy and the output after the monochromator is much more unstable than a laser seed.

2.6 Diagnostics lasers

The electro-magnetic field of the laser is used often as a tool for diagnostics due to its non-invasive nature. Lasers are used to measure beam size, electron bunch length and time of arrival as well as to characterize the final photon beam of the FEL. The following section will not give an exhaustive summary, but more of a taster on how lasers can be used for diagnostics. The references should be studied in more detail for each application.

A laser wire scanner is used at CAEP FEL to measure the transverse beam size of the electron beam at 250 keV [18]. The energy of the beam in this particular case is very low and therefore a non-invasive solution is very attractive. The bunch length is 15 ps and the repetition rate is very high at ~54 MHz, with a high bunch charge of 100 pC. The laser is propagated perpendicularly to the beam and is at 5 ps length to ensure accurate timing overlap. The focusing arrangement is chosen so that the Rayleigh range of the laser beam overlaps with the longitudinal size of the electron bunch. As only about 55.4 nJ energy is required an oscillator and a booster amplifier are sufficient. The laser is operating at 532 nm, using the second harmonic light of a neodymium laser. The interaction is based on Compton scattering.

Another well-established application is electro-topical bunch length and arrival time measurement [19–22]. All of these devices take advantage of the coulomb field generated by the electron beam. When using specific electro-optical crystals, which change their properties due to this field, one can map the properties of the electron beam in time by encoding the information into a short laser pulse passing through the crystal. The encoding can take place in the spectral domain, making the readout simple. For scanning methods the coincidence of the laser and the electron-pulse can be read out by a photodiode, while for single shot measurements a CCD is used to map the spectrum. Spatial encoding is also possible, though the imaging limits the resolution and the electron beam has to pass through the crystal. Temporal encoding mixes a reference pulse with the spectrally encoded pulse to achieve tens of femtoseconds resolution. Most methods are limited by the crystal size, as the signal strength is proportional to this, but the encoding is smeared out when thicker crystals are used. Relatively low energy but broadband pulses are needed, with a pulse structure matching the FEL's Ti:Saph laser, so pulses that are used for experiments can be split off for this purpose, or from fibre lasers with a broadening stage.

Beam arrival monitors are very important to keep the whole machine, often spread over several hundred metres, in synchronization. They use RF antennae to pick up the signal from the electron beam and use this signal with appropriate attenuation to drive an electro-optical modulator. The signal to be modulated is the optical pulsed reference. The signal, which is proportional to the relative delay between the optical reference and the pickup signal, is read out by a photodiode. Here 20 fs resolution can be achieved.

A THz streak camera is used to measure the relative timing between the X-ray and the pump laser pulses, as well as to characterize the FEL pulse length [58–60]. The technique has been adapted from the attosecond world, where an electric field of a few cycles of IR pulses are used to sweep the electrons [61], which are generated by ionization from the X-ray pulses, as shown in Fig. 14. The linear part of the laser's electric field has to match the pulse length. As FEL pulses are longer, one has to move to longer laser wavelengths, hence to terahertz pulses. The other requirement is the field strength, which has to be strong enough to deviate the electrons by an amount that is detectable by a time-of-flight device. The challenge also lies in the initial temporal and spatial overlap of the terahertz and X-ray pulses, both propagating in vacuum. This technique is also used for the AMO at LCLS.

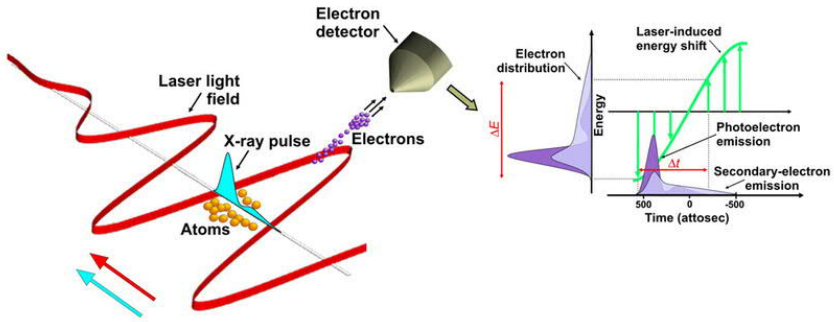


Fig. 14: A schematic of the operation of a THz streak camera

The laser system at FERMI@Elettra is a fine example of how a laser can be utilized for most tasks in the FEL. The same laser is used for seeding and experiments as well as for the time-of-arrival measurements. The long optical passes are actively stabilized to keep the synchronization between the different parts of the machine.

3 The future of lasers in FEL

Figure 15 shows the state-of-the-art for X-ray short pulse sources based on lasers and on FELs. With new projects, such as Extreme Light Infrastructure (ELI) [63] and Berkeley Lab Laser Accelerator (BELLA) [64] the photon energy gap between the two types of sources is closing. Both of these facilities utilize a petawatt laser, giving multiple tens of joules of laser energy in a single pulse at 1 Hz. The aim is to produce accelerating gradients reaching 100 GV/m by laser plasma acceleration. Eventually, robust fibre technology developed for telecommunications could provide large numbers of lasers locked together to reach joules of energy to drive laser plasma accelerators as well.

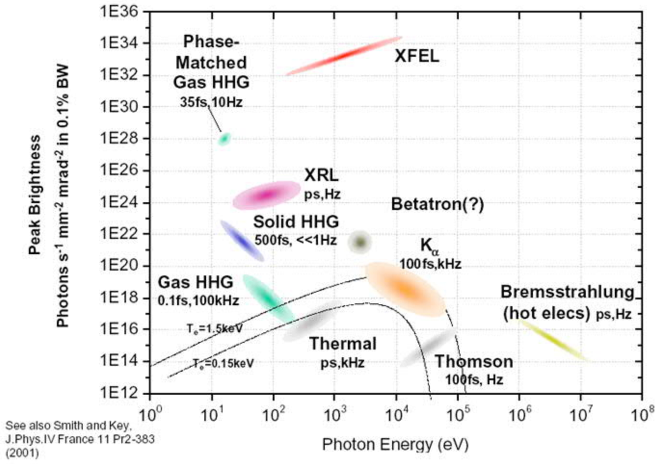


Fig. 15: Short pulse X-ray sources

The Advanced Proton Driven Plasma Wakefield Acceleration Experiment (AWAKE) project at CERN is aiming to combine the best of both accelerator and laser technology and accelerate an electron beam produced by a photo-injector using proton-driven plasma wakefield acceleration. The plasma is produced by a laser and the modulation, which drives the acceleration, is induced by self-modulation instabilities from the proton beam. The aim is to create GV/m accelerating strengths.

Scaling to higher repetition rates on the laser side has already been achieved, and commercial high pulse energy ultra-short pulse laser systems at a few kilohertz are already available on the market, while similar scaling for high power RF distribution is not clear. In conjunction with laser wakefield acceleration techniques, these lasers can provide tabletop X-ray sources for universities, small laboratories and medical treatment centres. I am certainly looking forward to reading the school notes in ten years' time on the same subject.

References

[1] A. Siegman, *Lasers* (University Science Books, 1986).

[2] Gould and R. Gordon, The LASER, light amplification by stimulated emission of radiation, Ann Arbor Conference on Optical Pumping, Michigan, 1959, Eds. P.A. Franken and R.H. Sands. OCLC 02460155, p. 128.

[3] Einstein A, *Verh. Dsch. Phys. Ges.* **18** (1916) 318.

[4] K.R. Spring, T.J. Fellers and M.W. Davidson, Introduction to lasers, http://www.olympusmicro.com/primer/techniques/confocal/laserintro.html (Olympus, 2012).

[5] W. Koechner, *Solid-State Laser Engineering* (Springer-Verlag, New York, 2006).

[6] P. Polynkin, A. Polynkin, D. Panasenko, N. Peyghambarian, M. Mansuripur and J. Moloney, *Opt. Lett.* **31**(5) (2006) 592. https://doi.org/10.1364/OL.31.000592

[7] D. Strickland and G. Mourou, *Opt. Commun.* **56**(3) (1985) 219. https://doi.org/10.1016/0030-4018(85)90120-8

[8] D. Strickland, P. Bado, M. Pessot and G. Mourou, *IEEE J. Quant. Electr.* **24**(2) (1988) 398. https://doi.org/10.1109/3.137

[9] A. Doyuran *et al.*, *NIMA in Phys. Res. A* **528**(1–2) (2004) 467. https://doi.org/10.1016/j.nima.2004.04.133

[10] F. Krausz and M. Ivanov, *Rev. Mod. Phys.* **81**(1) (2009) 163. https://doi.org/10.1103/RevModPhys.81.163

[11] Sutherland: *Handbook of Nonlinear Optics* (CRC Press).

[12] J.-C. Diels and W. Rudolph, *Ultrashort Laser Pulse Phenomena* (Elsevier, Amsterdam, 2006).

[13] X. Chen *et al*., *Sci. Rep* **5** (2015) 18343. https://doi.org/10.1038/srep18343

[14] B. Ning *et al*., *Sci. Rep*. **4** (2014) 5109. https://doi.org/10.1038/srep05109

[15] R.L. Sheffield *et al.*, *Science* **274**(5285) (1996) 236. https://doi.org/10.1126/science.274.5285.236

[16] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *NIM A* **490**(1–2) (2002) 1. https://doi.org/10.1016/S0168-9002(02)00905-1

[17] Z. Huang *et al.*, *Phys. Rev. ST Accel. Beams* **7**(7) (2004) 074401. https://doi.org/10.1103/PhysRevSTAB.7.074401

[18] A. Bosco *et al.*, *Nucl. Instrum. Meth. A* **592**(3) (2008) 162. https://doi.org/10.1016/j.nima.2008.04.012

[19] I. Wilke, A.M. MacLeod, W.A. Gillespie, G. Berden, G.M.H. Knippels and A.F.G. van der Meer, *Phys. Rev. Lett.* **88**(12) (2002) 124801. https://doi.org/10.1103/PhysRevLett.88.124801

[20] A.L. Cavalieri *et al.*, *PRL* **94**(11) (2005) 114801. https://doi.org/10.1103/PhysRevLett.94.114801

[21] S.P Jamison *et al*., *Opt. Lett.* **28**(18) (2003) 1710. https://doi.org/10.1364/OL.28.001710

[22] J. van Tilborg *et al*., *Opt. Lett.* **32**(3) (2007) 313. https://doi.org/10.1364/OL.32.000313

[23] M. Ferray *et al*., J. *Phys. B At. Mol. Opt. Phys.* **21**(3) (1988) L31. https://doi.org/10.1088/0953-4075/21/3/001

[24] G. Lambert *et al*., *Nat. Phys. Lett.* **295** (2008).

[25] L. Gianessi *et al*., *Nucl. Inst. & Meth. A* **594** (2008) 132. https://doi.org/10.1016/j.nima.2008.04.073

[26] C. Lechner *et al*., First direct seeding at 38 nm, Proc. FEL Conf., Nara, 2012.

[27] Ultra low timing jitter performance & characterization of Origami femtosecond laser series, P/N 09-001 (Onefive GmbH, Regensdorf, 2016) http://www.onefive.com/uploads/ultra_low-noise_whitepaper_onefive_v9_rev1.2.pdf.

[28] A. Trisorio *et al*., New concept for the SwissFEL Gun laser, Proc. FEL, New York, 2013.

[29] B. Beutner and S. Reiche, Proc. FEL, Malmö, Sweden, 2010, WEPB17.

[30] M.C. Divall, A. Romann, P. Mutter, S. Hunziker and C.P. Hauri, Laser arrival measurement tools for SwissFEL, Proc. SPIE 9512, Advances in X-ray Free-Electron Lasers Instrumentation III, Prague, 2015, 95121T (SPIE, 2015). https://doi.org/10.1117/12.2179016

[31] M.C. Divall, P. Mutter, E.J. Divall and C.P. Hauri, *Opt. Express* **23**(23) (2015) 29929. https://doi.org/10.1364/OE.23.029929

[32] I.Will, H.I. Templin, S. Schreiber and W. Sandner, *Opt. Express* **19**(24) (2011) 23770. https://doi.org/10.1364/OE.19.023770

[33] M. C. Divall, E. Prat, S. Bettoni, C. Vicario, A. Trisorio, T. Schietinger and C. P. Hauri, *Phys. Rev. ST Accel. Beams* **18**(3) (2015) 033401. https://doi.org/10.1103/PhysRevSTAB.18.033401

[34] M. Divall, Beam shaping, LA3NET School, Caen, 2012 (Paul Scherrer Institut, 2012) https://indico.cern.ch/event/177701/contributions/1441170/attachments/232573/325458/beam_shaping_Marta_Divall.pdf.

[35] I. Will and G. Klemz, *Opt. Express* **16**(19) (2008) 14922. https://doi.org/10.1364/OE.16.014922

[36] M. Bakr , G. Vashchenko, M. Khojoyan, M. Krasilnikov and F. Stephan, Proc. FEL2015, Daejeon, 2015, TUP065 (2015).

[37] A. Gallo *et al*., Laser and RF synchronization measurements at SPARC, Proc PAC07, Albuquerque, 2007 (2007). https://doi.org/10.1109/PAC.2007.4440959

[38] T.J. Maxwell *et al*., Synchronization and jitter studies of the titanium-sapphire laser at the A0 photoinjector, Proc PAC11, New York, 2011, MOP285 (2011).

[39] H. Qian *et al*., Timing jitter characterization at the NSLS SDL, Internal note for Brookhaven Science Associates BNL-82276-2009-CP (2009).

[40] N. Cutic *et al*., On-line arrival time and jitter measurements using electro-optical spectral decoding, Proc. FEL2010, Malmoe, 2010, THOA4, (2010).

[41] Schulz *et al*., Precision synchronization of the Flash Photoinjector Laser, IPAC10, Kyoto, 2010, WEPWB076 (2010).

[42] A. Azima *et al*., Jitter reduced pump-probe experiments. Proc of DIPAC 2007, Venice, Italy.

[43] A. Sakumi *et al*., Synchronization between laser and electron beam at photocathode RF gun, Proc PAC05, Knoxville, 2015 (2005).

[44] J.W. Bähr *et al*., BIW10, TUPSM103 (2010).

[45] M. Krasilnikov *et al*., DESY-M-04-03U (2004).

[46] V. Le Flanchec, J.-P. Blésès, S. Striby and J.-P. Laget, *Appl. Opt.* **36**(33) (1997) 8541. https://doi.org/10.1364/AO.36.008541

[47] A.R. Fry *et al*., Proc. PAC1997, 4w022 (1997).

[48] M. Danailov *et al*., Proc. International FEL Conf. 2007.

[49] P. Singalotti *et al*., Ultrafast laser synchronization at the FERMI@Elettra FEL, Proc. SPIE Vol. 8778 87780Q (2013). https://doi.org/10.1117/12.2020488

[50] M. Lederer, Laser systems for science instruments, European XFEL Users' Meeting, Hamburg, 2017, http://www.xfel.eu/sites/site_xfel-gmbh/content/e63594/e65073/e275358/e278336/04LasersystemsforscienceinstrumentsML_eng.pdf.

[51] C. Erny and C.P. Hauri, *J. Synchrotron Radiat*. **23** (2016) 1143. https://doi.org/10.1107/S1600577516012595

[52] M. Helm and P. Michel, *Notiziario Neutroni E Luce Di Sincrotrone* **11** (2006) 33.http://statistics.roma2.infn.it/~notiziario/2006/11_2_06/pag33.pdf

[53] E. Allaria *et al*., *New J. Phys*. **14** (2012) 11300. https://doi.org/10.1088/1367-2630/14/11/113009

[54] W. Leemans, High power laser technology for accelerators, White Paper of the ICFA-ICUIL Joint Task Force.

[55] FERMI. Elettra and FERMI lightsources. 2013-10-24. http://www.elettra.trieste.it/lightsources/fermi.html.

[56] J. Amann *et al*., *Nat. Photon*. **6** (2012) 693. https://doi.org/10.1038/nphoton.2012.180

[57] I. Gorgisyan, P.N. Juranic, R. Ischebeck, A. Stepanov, V. Schlott, C. Pradervand, L. Patthey, M. Radovic, R. Abela, C. P. Hauri, B. Monoszlai, R. Ivanov, P. Peier, J. Liu, T. Togashi, S. Owada, K. Ogawa, T. Katayama, M. Yabashi and L. Rivkin, The new design of the THz streak camera at PSI, Proc. SPIE 9512, Advances in X-ray Free-Electron Lasers Instrumentation III, 95120D (SPIE, 2015). https://doi.org/10.1117/12.2182204

[58] I. Grguras *et al*., *Nat. Photon*. **6** (2012) 852. https://doi.org/10.1038/nphoton.2012.276

[59] U. Frühling *et al*., *Nat. Photon*. **3** (2009) 353. https://doi.org/10.1038/nphoton.2009.160

[60] P. Juranic *et al*., *JINST* **9** (2014) P03006. https://doi.org/10.1088/1748-0221/9/03/P03006

[61] M. Hentschel *et al.*, *Nature* **414** (2001) 509. https://doi.org/10.1038/35107000

[62] Helml *et al*., *Nat. Photon.* **8** (2014) 950. https://doi.org/10.1038/nphoton.2014.278

[63] D. Jaroszynski, Ultra-short duration electron and radiation pulses from the laser-plasma wakefield accelerator, ELI-NP, 2015, http://www.eli-np.ro/2015-summer-school/presentations/24.09/Jaroszynski_ELI-NP_ss.pdf

[64] BELLA, Berkeley Lab, http://bella.lbl.gov/.

[65] Advanced wakefield experiment, (2013) http://awake.web.cern.ch/awake/

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Motion in the Undulator

S. Reiche
Paul Scherrer Institute, Villigen, Switzerland

Abstract
This paper gives an introduction to the theoretical framework for the motion of an electron in the periodic field of an undulator and wiggler. The equations of motion are derived and solved for planar and helical devices.

Keywords
Free-electron laser; theory; undulator; electron motion.

1 Electron motion in an undulator or wiggler

The hardware part of a free-electron laser is an undulator or wiggler. Its main purpose of the undulator or wiggler is to force the electrons to oscillate ('wiggle') as they move through it. This transverse motion causes the electron beam to emit synchrotron radiation. For relativistic electrons, the synchrotron radiation is confined to a forward cone. The opening angle is the inverse of the Lorentz factor $\gamma = E/mc^2$, where E is the electron energy, m is the electron mass, and c is the speed of light.

The main feature of an undulator and wiggler is a series of paired magnets along the main axis. They are placed opposite to each other, separated by a gap of width g. The magnetic flux has only a component transverse to the undulator axis. If the plane of the gap is fixed, the undulator or wiggler is planar. Another type of undulator involves rotation of the magnets along the main axis, in the form of a double helix. This type of undulator is called helical.

A Cartesian co-ordinate system, where the z-axis coincides with the undulator axis, will be used throughout this paper. The transverse co-ordinates x and y are chosen so that the magnetic field for a planar undulator or wiggler is parallel with the y-axis. Owing to rotational symmetry, the choice of co-ordinate system orientation for the helical undulator is arbitrary. Here, it is defined so that the magnetic field at the undulator entrance ($z = 0$) has only field components in the y-direction.

A higher magnetic field strength can be achieved by hybrid magnets, in which iron poles with high permeability are placed between permanent magnets [1]. Figure 1 shows a schematic cross-section of a planar undulator or wiggler based on hybrid magnets. The magnetic field of the permanent magnets points in either the positive or negative z-direction. The flux of two adjoining magnets is bent into the transverse direction by the iron pole. The advantage of this method is that the cross-section of the iron pole faces is smaller than that of the permanent magnets themselves. Therefore, the maximum achievable magnetic field can be increased by compressing the magnetic flux. A magnetic field strength larger than 2 T can then be obtained.

Wigglers and undulators differ in the deflection strength of the magnetic field. If the maximum deflection angle is larger than the opening angle of the spontaneous emission, there is no continuous emission in the forward direction, resulting in a wiggler. The spectrum observed is enriched by higher harmonics of the periodic signal of the detected radiation. Undulator radiation is modulated but not pulsed in the forward direction and the number of higher harmonics in the spectrum is reduced. A typical spectrum for the TESLA Test Facility is shown in Fig. 2. A more quantitative criterion to distinguish undulators and wigglers is given later in this section. Although both undulators and wigglers are used for free-electron lasers, for the sake of simplicity, the remaining part of this paper refers only to undulators, unless necessity requires that the two types must be distinguished.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.347

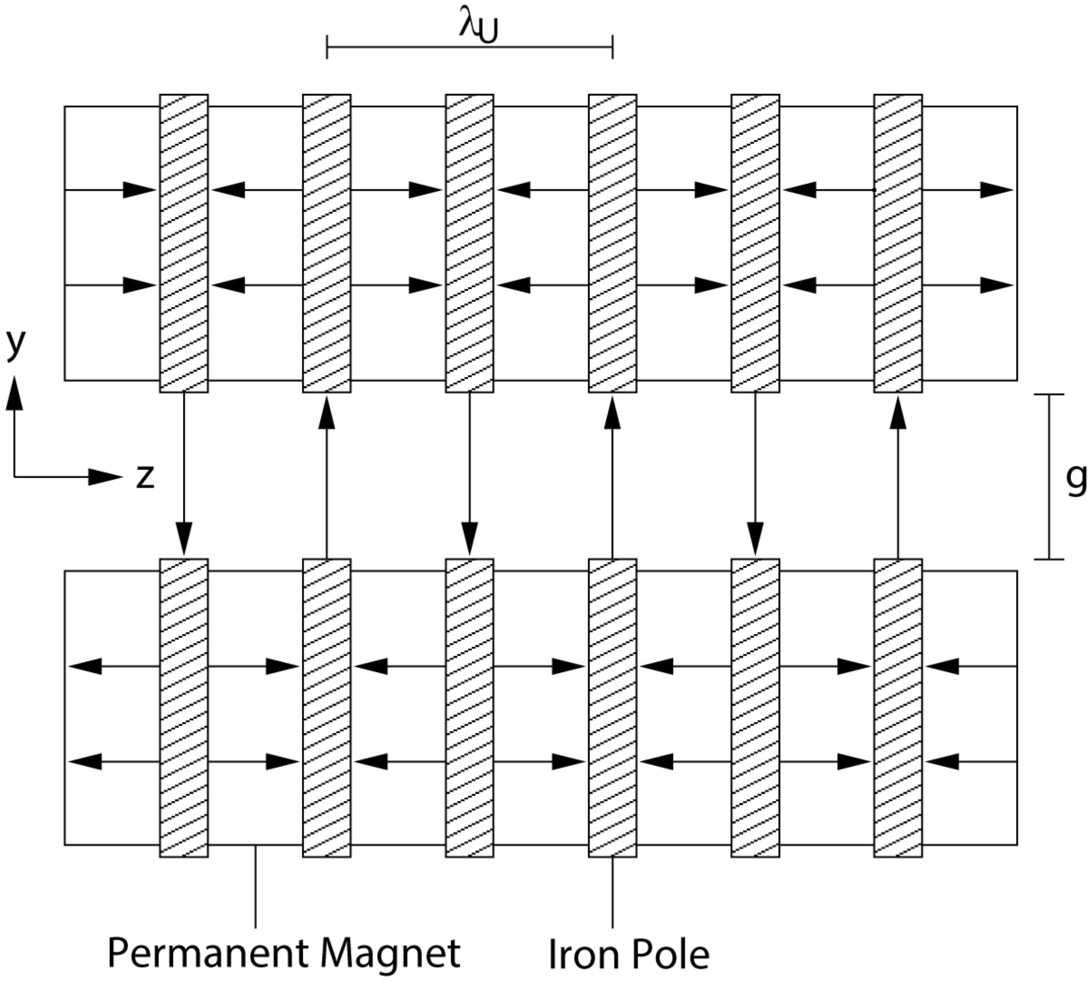


Fig. 1: Cross-section of planar undulator with gap width g and periodicity λ_U. The direction of the magnetic field is indicated by arrows.

1.1 The planar undulator

The discussion begins with the derivation of the electron trajectories within the planar undulator. The calculation for the helical case is similar and is given in the next subsection in a more compressed form.

The magnetic field at the undulator axis is a harmonic function of the longitudinal position z:

$$B_y(z, x = 0, y = 0) = B_0 \cos(k_U z) \, .$$

The field points in the y-direction and has an amplitude B_0 and a wavenumber $k_U = 2\pi/\lambda_U$. Although it might be desirable, the field cannot be constant over the whole transverse plane. Within the free space of the undulator gap, Maxwell's equations for a static magnetic field require that the divergence and curl vanish ($\vec{\nabla} \cdot \vec{B} = 0$ and $\vec{\nabla} \times \vec{B} = 0$). The second condition determines the dependence of the magnetic field on the transverse co-ordinates. It also allows the magnetic field to be derived from a scalar potential ϕ with $\vec{B} = -\vec{\nabla}\phi$. To fulfil Maxwell's equations, the scalar potential ϕ must be a solution of the Laplace equation $\Delta\phi = 0$.

A good starting assumption is

$$\phi = -\frac{B_0}{k_y} \cosh(k_x x) \sinh(k_y y) \cos(k_U z) \, , \tag{1}$$

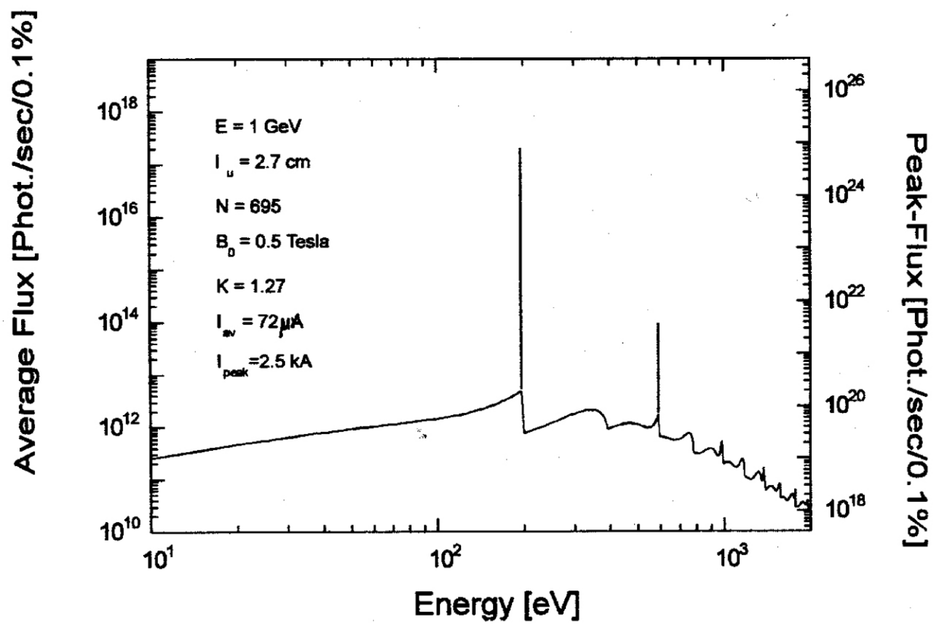


Fig. 2: Radiation spectrum of the free-electron laser at the TESLA Test Facility

which gives the desired magnetic field at the axis. Inserting Eq. (1) into the Laplace equation, the scalar potential is a physically reasonable solution if the relation

$$k_x^2 + k_y^2 = k_U^2 \tag{2}$$

is valid [2]. In general, to a good approximation, a magnetic field is perpendicular to the pole faces. This implies that the pole faces can be identified with equipotential surfaces, where the scalar potential ϕ is constant. For any arbitrarily chosen position z, the curvature of the equipotential surface is defined by the relation $\cosh(k_x x)\sinh(k_y y) =$ constant. It can be seen that y must be constant for $k_x = 0$ and that the pole faces are plane. The case of an outward bent pole face is covered by an imaginary value of k_x or, which is equivalent, by replacing the cosh function in Eq. (1) with the cosine function. In this case, k_y^2 becomes larger than k_U^2. For real values of k_x with $k_x > 0$, the two opposite poles are bent towards each other and k_y is either reduced ($k_x < k_U$), zero ($k_x = k_U$), or imaginary ($k_x > k_U$).

For x and y small compared with the undulator period length, so that $k_x x, k_y y \ll 1$, the hyperbolic function can be expanded into a Taylor series up to the second order. In this approximation, which is reasonable for most undulators up to a beam radius of typically 1 mm, the magnetic field becomes

$$\vec{B} = B_0 \begin{pmatrix} k_x^2 xy \cos(k_U z) \\ \left(1 + \frac{k_x^2 x^2}{2} + \frac{k_y^2 y^2}{2}\right)\cos(k_U z) \\ -k_U y \sin(k_U z) \end{pmatrix} . \tag{3}$$

The extra field caused by the curved pole faces is equivalent to a sextupole field with amplitude $B_0 k_x^2$. As shown later in this section, it provides focusing of the electron beam in the x-direction.

For further discussion, it is useful to know the vector potential $\vec{A}$ of the undulator field. This is given by

$$\vec{A} = \frac{B_0}{k_U} \begin{pmatrix} \left(1 + \frac{k_x^2}{2}x^2 + \frac{k_y^2}{2}y^2\right)\sin(k_U z) \\ -k_x^2 xy \sin(k_U z) \\ 0 \end{pmatrix} , \tag{4}$$

with $\vec{B} = \vec{\nabla} \times \vec{A}$.

The equations of motion for the position $\vec{r}$ and canonical momentum $\vec{P}$ of a single electron [3] are obtained from the Hamilton formalism, using the Hamilton function of a relativistic electron,

$$H = \sqrt{(\vec{P} - e\vec{A})^2 c^2 + m^2 c^4} + e\Phi \,, \tag{5}$$

where Φ is the scalar potential of the electric field $\vec{E}$ with $\vec{E} = -\vec{\nabla}\Phi - \partial\vec{A}/\partial t$.

If the electron is relativistic with $\gamma \gg 1$, the motion of the electron is mainly defined by the magnetic field of the undulator. Interaction with a radiation or electrostatic field can be regarded as a perturbation. These effects, which are important for the free-electron lasing process, are discussed in later sections.

With this assumption, the Hamilton function is a constant of motion because it does not depend explicitly on the time t. Owing to the absence of an electric field ($\Phi = 0$), the electron energy γmc^2 is also constant and identical in value to the Hamilton function.

It is difficult to solve the equations of motion directly. Therefore, the electron motion is split into two parts,

$$\vec{r}(t) = \vec{r}_0(t) + \vec{R}(t) \,,$$

separating the main oscillation $\vec{r}_0(t)$ due to the periodic undulator field from a drift $\vec{R}(t)$ in the transverse position. The drift is slow compared with the quickly varying term $\vec{r}_0(t)$ and has a characteristic length of the scale of many undulator periods. As a first step, the solution for $\vec{r}_0$ is obtained by assuming that $\vec{R}(t)$ is constant.

The equations of motion for the transverse canonical momentum $\vec{P}$ are

$$\dot{P}_x = -\frac{\partial}{\partial x} H = \frac{e}{\gamma m}\left(\frac{\partial}{\partial x}\vec{A}\right) \cdot \left(\vec{P} - e\vec{A}\right) \,, \tag{6}$$

$$\dot{P}_y = -\frac{\partial}{\partial y} H = \frac{e}{\gamma m}\left(\frac{\partial}{\partial y}\vec{A}\right) \cdot \left(\vec{P} - e\vec{A}\right) \,. \tag{7}$$

For the vector potential of Eq. (4), the lowest-order term of the time derivative $\dot{\vec{P}}$ is linear in $k_x x$ or $k_y y$, respectively. As mentioned at the expansion of the hyperbolic function in Eq. (3), these linear terms are small compared with unity. Thus, the change in the canonical momentum contributes either to the 'slow' motion $\vec{R}(t)$ or to the higher-order solutions of $\vec{r}_0$, which are not regarded in this discussion.

The remaining equations of the transverse motion,

$$\dot{x} = \frac{\partial}{\partial P_x} H = \frac{P_x - eA_x}{\gamma m} \,, \tag{8}$$

$$\dot{y} = \frac{\partial}{\partial P_y} H = \frac{P_y - eA_y}{\gamma m} \,, \tag{9}$$

have only one dominant and quickly oscillating source term, given by the x-component of the vector potential in Eq. (8). The resulting motion takes place in the xz-plane with the 'fast' velocity

$$\dot{x}_0 = -\frac{\sqrt{2}cK}{\gamma}\sin(k_U z) \,. \tag{10}$$

Equation (10) suggests the definition of the dimensionless undulator field:

$$K = \frac{e\hat{B}}{mck_U}\left(1 + \frac{k_x^2}{2}X^2 + \frac{k_y^2}{2}Y^2\right) \,, \tag{11}$$

depending, to second order, on the transverse position $X = X(t)$ and $Y = Y(t)$ of the 'slow' trajectory $\vec{R}(t)$. This definition differs from that given in other publications, where the on-axis peak field B_0 is used instead of the r.m.s. value $\hat{B}$. In the case of a planar undulator, $\hat{B}$ is $B_0/\sqrt{2}$. The advantage of this definition is that many equations remain the same for the case of the helical undulator. The value of K at the undulator axis ($X, Y = 0$) defines the undulator parameter. Because the second-order corrections to the undulator field are of the order of 10^{-3}, the transverse dependence of the undulator field has a negligible impact on most of the calculations. Therefore, it is sufficient to use the constant value of the undulator parameter instead.

Equation (10) exhibits the distinction between a wiggler and an undulator. If the electron is relativistic ($z \approx ct$), the maximum divergence $x' = \dot{x}_0/c$ of the electron is $\sqrt{2}K/\gamma$. The opening angle of the synchrotron radiation is γ^{-1}; thus, the device is an undulator for $K \leq 1/\sqrt{2}$ and a wiggler otherwise.

There is no dominant component of the vector potential in y and the motion in this direction consists only of the 'slow' motion ($y_0(t) = 0$).

Owing to energy conservation, the longitudinal velocity can be obtained directly from the definition of the Lorentz factor γ and the normalized velocity $\vec{\beta} = \mathrm{d}\vec{r}/c\mathrm{d}t$. Then the longitudinal velocity is

$$\begin{aligned} \beta_z &= \sqrt{1 - \frac{1}{\gamma^2} - \beta_x^2 - \beta_y^2} \\ &\approx 1 - \frac{1+K^2}{2\gamma^2} - \frac{\beta_R^2}{2} + \frac{K^2}{2\gamma^2}\cos(2k_U z)\,, \end{aligned} \tag{12}$$

where β_R is the transverse velocity of the slow drift, normalized to c. The cross term proportional to $\beta_R K/\gamma \cdot \sin(k_U z)$ has been neglected because it is either small compared with the leading oscillating term ($\propto K^2\cos(2k_U z)$) or not resonant with variation of β_z, as is the case for $\beta_R^2/2$.

The transverse motion within the undulator slows down the electron by roughly $\Delta\beta_z = K^2/2\gamma^2$ with a superimposed longitudinal oscillation with a period half as long as the transverse oscillation.

To obtain the trajectory $x_0(t)$, the longitudinal position is approximated by $z = c\beta_z t \approx c\beta_0 t$ and then Eq. (10) is integrated in first order, using the averaged velocity

$$\beta_0 = 1 - \frac{1+K^2}{2\gamma^2}\,. \tag{13}$$

The integration yields

$$x_0(t) = \frac{\sqrt{2}K}{\gamma k_U \beta_0}\cos(ck_U\beta_z t)\,. \tag{14}$$

The longitudinal oscillating term in Eq. (12) is the source of a phase modulation in the cosine function in Eq. (14). As a consequence, the transverse oscillation exhibits higher harmonics of the fundamental wavenumber k_U. In addition, the synchronization of the electron position with a phase front of an electromagnetic wave, propagating along the undulator axis, is reduced. The impact of both facts will be discussed in the next section. Only slowly varying terms in the equations of motion can contribute to $\vec{R}$. By averaging over one undulator period, Eqs. (8) and (9) are reduced to $\dot{X} = P_x/\gamma m$ and $\dot{Y} = P_y/\gamma m$. The vector potential $\vec{A}$ has only terms proportional to $\sin(k_U z)$ or $\sin(2k_U z)$ and these vanish after averaging. In the remaining equations, Eqs. (6) and (7), all terms are zero except for $(\partial A_x/\partial x)A_x$ and $(\partial A_x/\partial y)A_x$, respectively.

The resulting differential equations

$$\dot{P}_x = -\gamma mc^2\frac{K^2k_x^2}{\gamma^2}X\,, \tag{15}$$

$$\dot{P}_y = -\gamma mc^2\frac{K^2k_y^2}{\gamma^2}Y\,, \tag{16}$$

describe reaction forces proportional to the displacement.

The magnetic field of the undulator provides a natural focusing of the electrons if the pole faces are flat or bent towards each other ($k_x^2 \leq 0$). Although the focusing strength in both planes depends on the curvature of the magnetic poles, the combined strength $K^2 k_U^2/\gamma^2$ does not, owing to Eq. (2). For flat horizontal pole faces, there is no focusing in the x-plane. Increasing the focusing strength in this plane involves a reduction in the y-plane. A more precise calculation shows that the finite width of the undulator magnets introduces a small change in the magnetic field, so that a slight defocusing term is noticeable in the x-direction for $k_x = 0$ [4].

The trajectories of the transverse slow motion are harmonic functions with frequencies $\Omega_x = Kk_x/\gamma$ and $\Omega_y = Kk_y/\gamma$, for the x- and y-planes, respectively. The period length λ_β is typically of the order $\lambda_\beta \approx (\gamma/K)\lambda_U$ and thus much larger than the undulator period for a highly relativistic electron ($\gamma \gg 1$). The index β refers to the definitions used in accelerator physics, where this oscillation is called a betatron oscillation [5].

The transverse beam size is strongly related to the focusing strength. The calculations for the y-direction are identical to those for the x-direction, which are presented here. The general betatron oscillation of a single electron is given by $X(t) = X_0 \cos(\Omega_x z) + (X_0'/\Omega_x)\sin(\Omega_x z)$, where X_0 is the initial offset of the electron and X_0' is the initial angle relative to the undulator axis.

The emittance

$$\epsilon_x = \sqrt{\overline{(x-\overline{x})^2}\;\overline{(x'-\overline{x'})^2} - \overline{(x-\overline{x})(x'-\overline{x'})}^{\,2}}\,, \tag{17}$$

where a bar over a parameter denotes an average over all electrons, is a constant of motion in linear optics [6]. Regarding this definition of the emittance, $\pi\epsilon_x$ can be identified as an equivalent volume of the electron distribution in the transverse (x, x') phase space.

In contrast with the emittance, the r.m.sq. envelope of the electron beam is usually not a constant of motion [5]. The general expression of the envelope $\sigma_x(z)$ for $k_x^2 > 0$ within an undulator is

$$\sigma_x(z) = \sqrt{\sigma_x(0)^2\cos^2(\Omega_x z) + \frac{\sigma_x(0)\sigma_x'(0)}{2\Omega_x}\sin(2\Omega_x z) + \frac{\epsilon_x^2 - \sigma_x^2(0)\sigma_x'^2(0)}{\sigma_x^2(0)\Omega_x^2}\sin^2(\Omega_x z)}\;, \tag{18}$$

where $\sigma_x(0)$ and $\sigma_x'(0)$ are the initial beam size and its derivative in z, respectively. For a matched beam, when the beam size remains constant over the full undulator length, the electron beam must go through a waist directly at the entrance of the undulator ($\sigma_x'(0) = 0$) with an r.m.s. size of $\sigma_x(0) = \sqrt{\epsilon_x/\Omega_x}$. If the undulator focuses equally in both planes with $k_x = k_U/\sqrt{2}$, the constant size is

$$\sigma_x(0) = \sqrt{(\sqrt{2}mc/e)\epsilon_x\gamma/\hat{B}}\,.$$

All other initial settings cause a modulation of the envelope. If a smaller beam size is desired, it can be achieved by superimposing a lattice of quadrupoles. Normally, this is referred to as strong focusing, in contrast with the natural or weak focusing given by the undulator field itself.

1.2 The helical undulator

The treatment of the helical undulator is very similar to that of the planar one. Indeed most of the results are the same. The magnetic field $\vec{B}$, as well as the vector potential $\vec{A}$, consists of a linear combination of the first-order modified Bessel functions I_0 and I_1 [7], depending only on $k_U r$, where r is the transverse distance between the electron position and the undulator axis. Using the assumption that $k_U r$ is much smaller than unity, the Bessel functions are expanded into a Taylor series. Up to second order in $k_U r$, the vector potential in the Cartesian co-ordinate system is given by

$$\vec{A} = \frac{B_0}{k_U}\begin{pmatrix} \left[1 + \frac{k_U^2}{8}(3y^2 + x^2)\right]\sin(k_U z) - \frac{k_U^2}{4}xy\cos(k_U z) \\ \left[1 + \frac{k_U^2}{8}(3x^2 + y^2)\right]\cos(k_U z) - \frac{k_U^2}{4}xy\sin(k_U z) \\ 0 \end{pmatrix}\,. \tag{19}$$

The magnetic field is derived in the usual way by evaluating $\vec{B} = \vec{\nabla} \times \vec{A}$.

The trajectory is split into a quickly oscillating term $\vec{r}_0(t)$ and the slow betatron oscillation $\vec{R}(t)$. The velocity of the fast motion $\dot{\vec{r}}_0$ is proportional to the vector potential. Close to the undulator axis, $\dot{x}_0$ and $\dot{y}_0$ have the same amplitude of oscillation but they have a phase difference of $\pi/2$. This is an obvious result, which can be expected because of the symmetry of a helical undulator. Off-axis, the velocity differs in both directions, owing to the higher-order terms in x and y of the vector potential. The transverse motion of the electron becomes more elliptical, with the short axis pointing in the radial direction. To eliminate this azimuthal dependence, and for a better comparison with the results for the planar undulator, the vector potential is averaged in the azimuthal direction.

The normalized longitudinal velocity is

$$\beta_z \approx 1 - \frac{1+K^2}{2\gamma^2} - \frac{\beta_R^2}{2}\,, \tag{20}$$

with the undulator field

$$K = \frac{e\hat{B}}{mck_U}\left(1 + \frac{k_U^2}{4}(X^2+Y^2)\right) \tag{21}$$

in the Taylor series expansion up to second order in X and Y.

The major difference between helical and planar undulators becomes apparent here. Because the electron oscillates in both transverse directions but with a $\pi/2$ phase difference, the longitudinal velocity is almost constant. The terms proportional to $\beta_R \cos(k_U z)$ or $\beta_R \sin(k_U z)$ are negligible and not included in Eq. (20). The absence of a longitudinal oscillation excludes the generation of higher harmonics in the transverse motion of the electron. The helical undulator field of Eq. (21) agrees with that of a planar undulator (Eq. (11)) if the planar undulator provides equal focusing in both planes, with $k_x^2 = k_y^2 = k_u^2/2$. This similarity is an advantage of the undulator field definition based on the r.m.s. value $\hat{B}$.

With the definition of β_0 in analogy with Eq. (13), the transverse velocity can be integrated to obtain the trajectory $\vec{r}_0$. The electrons move along a helix with a pitch of λ_U. Owing to the asymmetry in the azimuthal and radial motion for larger transverse offsets, the helix is slightly distorted [8]. The average radius of the motion is independent of the azimuthal angle, with

$$r_0 = \frac{K}{\gamma k_U \beta_0}\,. \tag{22}$$

By averaging the transverse equations of motion over the length of one period, the fast oscillation drops out. Some basic algebra yields the differential equations:

$$\dot{X} = \frac{P_x}{\gamma m} - c\frac{\Omega_U^2}{k_U}Y\,, \tag{23}$$

$$\dot{P}_x = -\gamma mc^2\Omega_U^2 X - c\frac{\Omega_U^2}{k_U}P_y\,, \tag{24}$$

$$\dot{Y} = \frac{P_y}{\gamma m} - c\frac{\Omega_U^2}{k_U}X\,, \tag{25}$$

$$\dot{P}_y = -\gamma mc^2\Omega_U^2 Y - c\frac{\Omega_U^2}{k_U}P_x\,, \tag{26}$$

with $\Omega_U = Kk_u/\sqrt{2}\gamma$.

These differential equations describe two coupled oscillations but they can be decoupled into ordinary differential equations for harmonic oscillations with the frequencies

$$\hat{\Omega} = \Omega_U\left(\sqrt{1+\frac{\Omega_U^2}{k_U^2}} \pm \frac{\Omega_U}{k_U}\right), \tag{27}$$

by transforming to the variables $X \pm \mathrm{i}Y$. The ratio Ω_U/k_U is of the order $1/\gamma$. The characteristic length of the orbit beat by the coupling is roughly $\gamma^2\lambda_U$ and, even for moderately relativistic electrons, much longer than the undulator length itself. This term is only important for storage-ring-based undulators because it is a major source of coupling of the betatron motion [8]. By neglecting the coupling term, Eqs. (23)–(26) become identical to the corresponding equations for the planar undulator, with $k_x^2 = k_y^2 = k_U^2/2$. The conditions for optimum matching of the electron beam are also valid for the helical undulator.

References

[1] K. Halbach, *J. Phys. (Paris)* **C1 44** (1983) 211.

[2] E.T. Scharlemann, *J. Appl. Phys.* **58** (1985) 2154. https://doi.org/10.1063/1.335980

[3] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980).

[4] B. Diviacco and P. Walker, *Nucl. Instrum. Methods*, **A292** (1990) 517. https://doi.org/10.1016/0168-9002(90)90409-Y

[5] D.A. Edwards and M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (John Wiley and Sons, New York, 1993). https://doi.org/10.1002/9783527617272

[6] K. Wille, *Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen* (B.G. Teubner Verlag, Stuttgart, 1992). https://doi.org/10.1007/978-3-663-11850-3

[7] G. Rowlands, J. *Phys. A: Math. Gen.* **13** (1980) 2839. https://doi.org/10.1088/0305-4470/13/8/031

[8] J.P. Blewett and R. Chasman, *J. Appl. Phys.* **48** (1977) 2692. https://doi.org/10.1063/1.324119

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Pendulum Equations and Low Gain Regime

S. Reiche
Paul Scherrer Institute, Villigen, Switzerland

Abstract
This paper introduces the theoretical framework for the motion of an electron in the periodic field of an undulator and wiggler. It is a continuation of the previous article on motion in the undulator but includes the interaction with a co-propagating radiation field. The longitudinal motion of each electron is similar to a pendulum and, in resonance with the co-propagating field, energy can be exchanged between the electron beam and the radiation field in the small signal low-gain regime of the free-electron laser.

Keywords
Free-electron laser; theory; low-gain regime; resonance condition.

1 Pendulum equations and low-gain regime

In this section, the interaction of electrons with a radiation field while they move through the undulator is analysed. The approach to this problem is similar to that in the previous paper, except that an additional term in the Hamilton function describes the vector potential of the radiation field [1]. If the emission of radiation is stronger than the absorption, the electrons are losing energy, on average, and the radiation field is amplified. As long as this amplification is small, the radiation field amplitude can be assumed to be constant in the Hamilton function for deriving the equations of motion. The limitations of this model of a 'low-gain' free-electron laser are given at the end of this section. A more self-consistent model of a free-electron laser can be found in the next section, including Maxwell's equation for the radiation field description. Nevertheless, a discussion of the low-gain free-electron laser is fruitful, because it shows the basic principle of how a free-electron laser works, using rather simple equations.

The interaction of charged particles with a radiation field shows two major aspects. The first is the change of the particle momentum and energy. The Hamilton equations of motion are the mathematical representation of this process. The method for solving these equations is very similar to the treatment in the previous paper, but differs in that the electron energy is no longer constant, owing to the electric field components of the radiation field.

The second aspect is the change of the radiation field itself. The fast transverse oscillation of the electrons is a source of radiation. For relativistic particles, this radiation points mainly in the forward direction of the electron beam motion. If the radiation wavelength is shorter than the electron bunch length, the electrons emit at almost all phases and the radiation adds up incoherently. The emission is strongly enhanced if the longitudinal beam profile is modulated on the scale of the radiation wavelength.

Under special conditions, both processes—the change of the particle energy and the emission of radiation—are the source of a collective bunching of the electrons on a resonant frequency, and the radiation field is strongly amplified. The next section analyses this instability—the working principle of the 'high gain' free-electron laser. In contrast with the high gain free-electron laser, the low-gain free-electron laser provides amplification without the necessity of a strong modulation in the electron density.

The discussion begins with the assumption on the radiation field. If a radiation field propagates along the undulator together with the electron bunch, the interaction time is maximized. The electric field components are lying in the transverse xy-plane; thus, only a transverse motion, along or against the field orientation, changes the electron energy. Owing to the symmetry of the magnetic field, the

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.355

radiation emitted in a planar undulator is linearly polarized, while it is circularly polarized for the case of a helical undulator. In this section, the case of a planar undulator is considered. Most of the results are similar or identical for a helical undulator and only important differences are mentioned in the text.

The electric field component of the radiation field

$$\vec{E} = \vec{E}_0 \cos(k(z - ct) + \Psi)\,, \tag{1}$$

is defined by its amplitude $\vec{E}_0$, its wavenumber $k = 2\pi/\lambda$ or wavelength λ, and its initial phase Ψ at the undulator entrance.

The magnetic field component is perpendicular to $\vec{E}$ as well as to the unit vector in the direction of propagation, which mainly coincides with $\vec{e}_z$. Compared with the strong undulator field, the magnetic field of the radiation field is negligible and can be ignored in further discussion. The amplitude $\vec{E}_0$ and the phase Ψ depend on z, because of diffraction. The dependence becomes negligibly small if the transverse extension of the radiation wavefront is much larger than the radiation wavelength.

The change of the electron energy is caused only by the electric field components, which, depending on the radiation phase, accelerate or decelerate the electron with

$$\dot{\gamma} = \frac{e\vec{E}\cdot\vec{\beta}}{mc}\,. \tag{2}$$

Only the parallel components of $\vec{E}$ and $\vec{\beta}$ contribute to Eq. (2). For the planar undulator, they are pointing in the x-direction resulting in a linear polarization of the radiation field.

To obtain the transverse velocities, $\vec{\beta}$ the vector potential $\vec{A}_r$ of the electromagnetic wave must be added to the Hamiltonian,

$$H = \sqrt{(\vec{P} - e\vec{A})^2 c^2 + m^2 c^4} + e\Phi\,. \tag{3}$$

From the potential,

$$\vec{A}_r = \frac{1}{ck}\sin(k(z - ct) + \Psi)\begin{pmatrix} E_0 \\ 0 \\ 0 \end{pmatrix}\,. \tag{4}$$

the electric field is derived by the time derivative $\vec{E} = -\partial\vec{A}_r/\partial t$. Here, the Lorentz gauge is chosen, which enables the scalar potential to be omitted in the derivation of the electric field.

For an assumed pulse length $L \gg \lambda$, the dependence of the amplitude $\vec{E}_0$, as well as the phase Ψ, on time is negligible and A_r is a valid vector potential for the radiation field of Eq. (1). Inserting the vector potential of the radiation field and the undulator field into the Hamilton function, the transverse velocities are

$$\dot{x} = -\frac{\sqrt{2}cK}{\gamma}\sin(k_{\mathrm{U}}z) - \frac{\sqrt{2}cK_r}{\gamma}\sin(k(z - ct) + \Psi) + \dot{X}\,, \tag{5}$$

$$\dot{y} = \dot{Y}\,. \tag{6}$$

The dimensionless radiation amplitude,

$$K_r = \frac{e\hat{E}}{mc^2 k}\,, \tag{7}$$

is defined in an analogous way as the undulator parameter K. The motivation to use the r.m.s. value $\hat{E}$ of the electric field is the same. Most results will be identical for the helical undulator. The velocity terms $\dot{X}$ and $\dot{Y}$ of the betatron oscillation are the same as before.

For the sake of simplicity, any transverse variation of the radiation field is excluded. A radiation field with a finite transverse extension is more difficult to analyse. For small transverse momenta, the longitudinal velocity is approximately

$$\beta_z \approx 1 - \frac{1+K^2+K_r^2}{2\gamma^2} - \frac{\beta_{\mathrm{R}}^2}{2} + \frac{K^2}{2\gamma^2}\cos(2k_{\mathrm{U}}z) + \frac{K_r^2}{2\gamma^2}\cos\left(2k(z-ct)+2\Psi\right)$$
$$- \frac{2KK_r}{\gamma^2}\sin(k_{\mathrm{U}}z)\sin(k(z-ct)+\Psi)\,. \tag{8}$$

This expression is very similar to that for the electron motion in a pure magnetic field of an undulator, except for three additional terms. The electric field forces an additional transverse oscillation with the frequency of the electromagnetic wave. As for the undulator field, the longitudinal velocity is slowed down and modulated with an oscillation of twice the frequency of the radiation field. It will be shown later that the longitudinal modulation by the radiation field is much smaller than the longitudinal modulation by the undulator field and can be neglected.

The cross term, $\propto KK_r$, can be split into two independent oscillations. If one of them has a small frequency, it can significantly change the longitudinal velocity β_z on a time-scale different to the dominant oscillating term, $\propto K^2$. The explicit calculation of this term is postponed until β_z is further discussed (see Eq. (15)).

Combining all constant or slowly varying terms to β_0, the integration of Eq. (8) up to first order yields

$$z = \beta_0 ct + \frac{K^2}{4\gamma^2 k_{\mathrm{U}}\beta_0}\sin(2k_{\mathrm{U}}\beta_0 ct)\,. \tag{9}$$

With the given expression of the transverse velocities $\dot{x}$ and $\dot{y}$, Eq. (2) can be evaluated. Most of the cross terms between E_x and β_x are quickly oscillating. Over many undulator periods, the net change of the electron energy is negligible. The only possible term that might be constant is the product of $\cos(k(z-ct)+\Psi)$ and $\sin(k_{\mathrm{U}}z)$, similar to the term in Eq. (8). This term is split into two independent oscillations, with the phases $(k \pm k_{\mathrm{U}})z - kct + \Psi$. If one of the phases remains almost constant, the energy change is accumulated over many periods.

With an average longitudinal velocity of $c\beta_0$, the phase relation between the electron and the radiation field remains unchanged if the condition

$$\beta_0 = \frac{k}{k \pm k_{\mathrm{U}}} \tag{10}$$

is fulfilled. As shown later, the interaction between the electron beam and the radiation field needs to add up resonantly over many undulator periods to result in a significant change of the electron energy or the radiation amplitude and phase. This implies that, for a given beam energy and undulator wavelength, the radiation wavelength of the radiation field is well defined according to Eq. (10). The case of the '$-$' sign is excluded because it would demand an electron velocity faster than the speed of light to keep the electrons in phase with the radiation field for any time. The restriction to a well-defined resonant radiation wavelength is called the resonance approximation.

In the limit of a weak electric field ($K_r \to 0$) and a small beam emittance, the resonant radiation wavelength is

$$\lambda_0 = \frac{\lambda_{\mathrm{U}}}{2\gamma^2}(1+K^2)\,. \tag{11}$$

This important equation is also valid for a planar and a helical undulator. A transverse betatron motion and a stronger radiation field shift the resonance condition slightly towards longer wavelengths. If Eq. (11) is exactly fulfilled, the energy change is constant over many undulator periods, pushing the electron off resonance.

So far, the longitudinal oscillation of the electron has not been taken into account. As mentioned previously, it induces higher harmonics in the motion of the electrons.

Inserting Eqs. (1) and (5) into Eq. (2) yields the resonant term

$$\dot{\gamma} = -\frac{2ckKK_r}{\gamma}\cos(k(z-ct)+\Psi)\sin(k_{\mathrm{U}}z)\,. \tag{12}$$

Note that the choice of the radiation wavenumber k is free and does not need to agree with the resonant wavenumber $k_0 = 2\pi/\lambda_0$, defined by the undulator properties and the particle energy. To evaluate Eq. (12), the sine and cosine function are replaced by complex exponential functions. The oscillating part of the longitudinal motion (Eq. (9)) can be expanded into a series of Bessel functions [2] by the identity

$$\mathrm{e}^{\mathrm{i}a\sin b} = \sum_{m=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}mb} J_m(a)\,.$$

The result is a sum of exponential functions with frequencies $[(k+(2m+1)k_{\mathrm{U}})\beta_0 - k]c$. Beside the ground mode with $m = 0$, some terms are resonant at different wavelengths. The frequencies of these are the odd harmonics of the resonant frequency $\omega_0 = ck_0$.

Collecting all terms belonging to one mode, Eq. (12) becomes

$$\begin{aligned}\dot{\gamma} = -\frac{2ckKK_r}{\gamma}\frac{1}{4i}\Bigg[&\mathrm{e}^{\mathrm{i}\theta+\mathrm{i}\Psi}\sum_{m=-\infty}^{\infty}\mathrm{e}^{\mathrm{i}2mk_{\mathrm{U}}\beta_0 ct}(J_m(\chi)-J_{m+1}(\chi))\\ &-\mathrm{e}^{-\mathrm{i}\theta-\mathrm{i}\Psi}\sum_{m=-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}2mk_{\mathrm{U}}\beta_0 ct}(J_m(\chi)-J_{m+1}(\chi))\Bigg]\,,\end{aligned} \tag{13}$$

with $\chi = kK^2/4\gamma^2 k_{\mathrm{U}}$ and the so-called ponderomotive phase,

$$\theta = (k+k_{\mathrm{U}})z - ckt\,. \tag{14}$$

For completeness, it is noted that a transverse non-uniform radiation field also couples the particle motion to the even harmonics of ω_0 [3,4]. If the radiation field is expanded into a Taylor series around the electron position of the betatron oscillation ($x = X + x_0$),

$$\vec{E}(x) = \vec{E}(X) + \left.\frac{\mathrm{d}\vec{E}}{\mathrm{d}x}\right|_X x_0\,,$$

the factor $x_0\dot{x_0}$ is proportional to $\sin(2k_{\mathrm{U}}z)$ in Eq. (2). Using the same calculation as for Eq. (13), the complex exponential functions have the arguments $[(k+(2m+2)k_{\mathrm{U}})\beta_0 - k]ct$, being resonant at all even harmonics. The additional pre-exponential factor is $(K/2K_r\gamma k_{\mathrm{U}}\beta_0)\mathrm{d}K_r/\mathrm{d}x$.

The postponed calculation of the cross term $\sin(k(z-ct)+\Psi)\sin(k_{\mathrm{U}}z)$ in Eq. (8) is performed in a very similar way. If the phase Ψ is temporarily replaced by $\tilde{\Psi} = \Psi - \pi/2$ to convert the sine function into a cosine function, the expansion into Bessel functions yields

$$\begin{aligned}\beta_z = 1 &- \frac{1+K^2+K_r^2}{2\gamma^2} - \frac{\beta_{\mathrm{R}}^2}{2}\\ &+ \frac{KK_r}{2\gamma^2}\Bigg[\mathrm{e}^{\mathrm{i}\theta+\mathrm{i}\Psi}\sum_{m=-\infty}^{\infty}\mathrm{e}^{\mathrm{i}2mk_{\mathrm{U}}\beta_0 ct}(J_m(\chi)-J_{m+1}(\chi))\\ &\qquad + \mathrm{e}^{-\mathrm{i}\theta-\mathrm{i}\Psi}\sum_{m=-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}2mk_{\mathrm{U}}\beta_0 ct}(J_m(\chi)-J_{m+1}(\chi))\Bigg]\,.\end{aligned} \tag{15}$$

The resonant frequencies are well separated, such that only one resonance frequency is of importance for a given radiation field. The coupling factor is smaller for higher modes. Thus, the interaction is the strongest for the fundamental mode [5], which is the only mode considered in the following discussion.

Where the free-electron laser operates at the fundamental frequency, the non-linear terms in the free-electron laser equations will induce an enhanced bunching in the longitudinal position at higher harmonics. This bunching increases more quickly than operating on the higher frequency itself.

For a helical undulator, the amplification of higher modes is much smaller because the dominant longitudinal oscillation, which is why coupling to higher harmonics is strongly suppressed. At the fundamental frequency, the synchronization of the phase front of the ponderomotive wave and the electrons is almost perfect, while it is reduced by a factor $(J_0(\chi) - J_1(\chi))$ for the planar undulator.

Compared with the fast-changing position of the electron, $z \approx \beta_0 ct$, the ponderomotive phase $\theta = (k + k_{\mathrm{U}})z - ckt$ of the electron is almost constant. It is convenient to change to a moving coordinate system, which is synchronized with the ponderomotive wave. With a simple canonical transformation [6], which keeps the energy unchanged, the equation of motion for the new variable θ becomes $\dot{\theta} = (k + k_{\mathrm{U}})c\beta_z - kc$. Replacing β_z with Eq. (15), the differential equations for the low-gain free-electron laser are obtained:

$$\dot{\theta} = ck_{\mathrm{U}} - \omega \frac{1 + K^2 + K_r^2 - 2f_{\mathrm{c}}KK_r\cos(\theta + \Psi)}{2\gamma^2} - \omega\frac{\beta_{\mathrm{R}}^2}{2}\,, \tag{16}$$

and

$$\dot{\gamma} = -\omega f_{\mathrm{c}}\frac{KK_r}{\gamma}\sin(\theta + \Psi)\,. \tag{17}$$

With the definition of the coupling factor

$$f_{\mathrm{c}} = \begin{cases} J_0(\chi) - J_1(\chi) & \text{planar undulator,} \\ 1 & \text{helical undulator.} \end{cases} \tag{18}$$

and $\chi = kK^2/4\gamma^2 k_{\mathrm{U}} = K^2/2(1+K^2)$ for the fundamental resonant wavelength, the free-electron laser equations are valid for both types of undulator.

Another way to derive the differential equations is the rigorous canonical and Legendre transformation of the Hamilton function of Eq. (3) [7]. The new Hamilton function, depending on the canonical variable and momentum θ and γ, respectively, is

$$H = ck_{\mathrm{U}}\gamma + \omega\frac{1 + \gamma^2\beta_{\mathrm{R}}^2 + K^2 + K_r^2 - 2f_{\mathrm{c}}KK_r\cos(\theta + \Psi)}{2\gamma}\,. \tag{19}$$

The independent variable is time t. As long as the electric field and the transverse momenta do not change significantly, they can be kept constant in the Hamiltonian. This is the basic assumption of the low-gain free-electron laser. The limitation of this model will be given at the end of this section.

In the limit of a low-gain free-electron laser, the Hamilton function is regarded as independent of t and therefore a constant of motion. Setting the Hamiltonian to $H = 2ck_{\mathrm{U}}(1 + \alpha)\gamma_{\mathrm{R}}$ with $\gamma_{\mathrm{R}}^2 = k(1 + \gamma^2\beta_{\mathrm{R}}^2 + K^2 + K_r^2)/2k_{\mathrm{U}}$, the particle energy γ depends on θ as

$$\gamma = \gamma_{\mathrm{R}}(1 + \alpha) \pm \sqrt{\gamma_{\mathrm{R}}^2\alpha(2 + \alpha) + \frac{kf_{\mathrm{c}}KK_r}{k_{\mathrm{U}}}\cos(\theta + \Psi)}\,. \tag{20}$$

The lowest boundary of α is $\alpha > -1$, to avoid unphysical negative values of the energy. Other limitations are given by the square root in Eq. (20). Two values of α are of particular interest, for the lowest possible value of the Hamilton function and for an existing solution of γ for all phases θ, respectively.

The smallest value of α is found if the cosine function in the argument of the square root is unity. At $\theta = -\Psi$ the root becomes real for

$$\alpha_0 = -1 + \sqrt{1 - \frac{kf_\mathrm{c}KK_r}{k_\mathrm{U}\gamma_\mathrm{R}^2}}\,. \tag{21}$$

Inserting α_0 in Eq. (20) yields the corresponding energy

$$\gamma_0 = \sqrt{\gamma_\mathrm{R}^2 - \frac{kf_\mathrm{c}KK_r}{k_\mathrm{U}}}\,.$$

The position $(-\Psi, \gamma_0)$ in the longitudinal phase space is a stable fixed point, where the electron remains in its position. For any small deviation, the differential equations, Eqs. (16) and (17), can be linearized and combined to produce a second-order differential equation of $\Delta\theta = \theta + \Psi$ with

$$\Delta\theta'' + \Omega^2\Delta\theta = 0 \tag{22}$$

and $\Omega = \sqrt{2f_\mathrm{c}kk_\mathrm{U}KK_r}/\gamma_0$.

This equation is solved by any sine or cosine function with the frequency Ω. The motion in the longitudinal phase space is bound. This is typical for a stable fixed point. For a larger amplitude of $\Delta\theta$, non-linear terms are no longer negligible and the frequency depends on the initial condition of the electron.

Solutions of γ for all phases θ are found for α larger than

$$\alpha_1 = -1 + \sqrt{1 + \frac{kf_\mathrm{c}KK_r}{k_\mathrm{U}\gamma_\mathrm{R}^2}}\,. \tag{23}$$

The trajectory in phase space is not closed and the electrons have either energy above or below γ_R. A transition is not possible.

The phase space surface for $H = 2ck_\mathrm{U}(1 + \alpha_1)\gamma_\mathrm{R}$ is called a separatrix. It separates the bound and unbound motion. Any electron within the separatrix is trapped in the ponderomotive wave and oscillates around $-\Psi$. Referring to the acceleration of charged particle in RF cavities, this enclosed area of the separatrix is often called a 'bucket' [8]. The width of the bucket is given by the properties of the undulator and the radiation field and is $\Delta\gamma = \sqrt{8kf_\mathrm{c}KK_r/k_\mathrm{U}}$. Electrons outside the separatrix are moving unlimited in θ either faster or slower than the ponderomotive wave.

Figure 1 shows several phase space trajectories for different initial conditions calculated by Eq. (20). Within the bucket, the electrons are moving clockwise; above zero, they move towards larger phases ($\dot{\theta} > 0$), while below zero, they move towards smaller phases. This implies that an electron injected at the ponderomotive phase $0 < \theta + \Psi < \pi$ loses energy. If the undulator length is shorter than the period length of the phase space oscillation $2\pi/\Omega$, the electron will mainly remain in this phase region. Owing to energy conservation, the radiation field has been amplified. This can be generalized for the whole electron bunch. As long as the initial distribution in the longitudinal phase space changes to a final distribution of a mean energy smaller than the initial energy, the gain of the free-electron laser is positive.

Unfortunately, the most obvious way by injecting all electrons at $0 < \theta + \Psi < \pi$ is not realizable. The radiation wavelength depends on the energy as γ^{-2} (Eq. (11)) and is much smaller than a typical bunch length of about 1 mm. The initial ponderomotive phases of the electrons are almost uniformly distributed over 2π. Owing to the finite number of electrons over one radiation wavelength, a small modulation of the electron beam remains. This spontaneous emission provides the initial radiation field for self-amplified spontaneous emission free-electron lasers, discussed in the end of this paper.

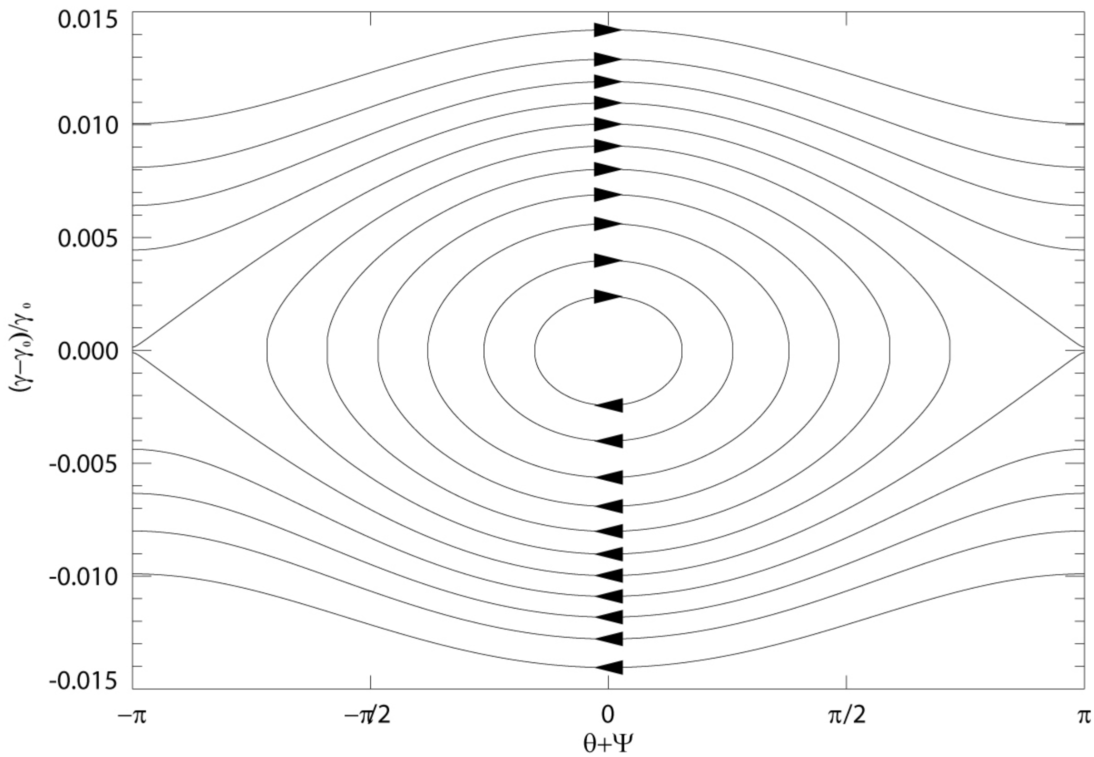


Fig. 1: Electron trajectories in the longitudinal phase space for different initial settings

With an RF photo gun driving the injector for a free-electron laser, relative energy spreads smaller than 1% can be achieved. This width is typically smaller than the width of the bucket and fills it unevenly. For a large energy spread, the bucket is filled almost homogeneously. Any motion of the electrons within the homogeneously filled bucket would not change the mean energy, because the phase space density remains constant, according to Liouville's theorem [9].

Operating as a free-electron laser amplifier, the injection at resonance energy γ_{R} would not provide any gain at all. For the unmodulated beam, the energy change of one electron is always compensated by a complementary electron, which moves on the same trajectory but which has a phase difference of $2(\theta+\Psi)$. The only visible effect is the increase of the energy spread, because electrons at $-\pi < \theta+\Psi < 0$ gain energy while the complementary electrons at $0 < \theta + \Psi < \pi$ lose energy.

If the injection is off-resonance ($\gamma \neq \gamma_{\mathrm{R}}$) the change of the phase space distribution is no longer symmetrical. For $\gamma > \gamma_{\mathrm{R}}$, electrons at $-\pi < \theta + \Psi < 0$ tend to change the phase rather than the energy, while the opposite is true for the remaining electrons. Averaging over all electrons, the electron beam loses energy and the radiation field is amplified. For injection below the resonant energy, the electron beam will gain energy and the radiation field is weakened.

The gain dependence on the injection energy can be calculated by perturbation theory [10]. The rather long but straightforward calculation is not presented here. The dependence on the injection energy is

$$G \propto -\frac{\mathrm{d}}{\mathrm{d}(\eta/2)} \frac{\sin^2(\eta/2)}{(\eta/2)^2}\,, \tag{24}$$

where $\eta = 4\pi N_{\mathrm{U}}(\gamma - \gamma_{\mathrm{R}})/\gamma_{\mathrm{R}}$ and N_{U} is the total number of undulator periods. The gain of the low-gain amplifier is related to the spectrum of the spontaneous undulator radiation [11, 12] by taking the frequency derivative of the intensity spectrum of the spontaneous radiation. This relation is known as Madey's theorem [13].

For the free-electron laser oscillator, as well as for the self-amplified spontaneous emission free-electron laser, the situation is slightly different, because both types of free-electron laser start from spontaneous emission with a broad bandwidth in the frequency domain. As a consequence, the electron beam is always in resonance with the frequency of the largest gain. Using an energy dependence as the argument of Eq. (24) is no longer meaningful and the energy dependence must be replaced with the frequency dependence. The results are made similar by redefining η as $\eta = 2\pi N_{\mathrm{U}}(\omega - \omega_0)/\omega_0$, with ω_0 as the resonant frequency.

In this low-gain approximation, the interaction between the electrons is almost negligible and the gain is proportional to the total number of electrons. In this one-dimensional model of a low-gain free-electron laser, a higher beam current means a larger amplification of the radiation field. Unless the gain does not exceed several per cent of the usage of the free-electron laser, Eqs. (16) and (17) are justified. Otherwise, the assumption of a constant field K_r is no longer valid. The radiation power can increase, which might change the strength of the electron interaction. To cover this aspect, a self-consistent set of free-electron laser equations must be derived.

References

[1] S. Reiche, Motion in the Undulator, these proceedings.

[2] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Function* (Dover Publication, New York, 1972).

[3] D. Noelle, Ph.D. thesis, DELTA Int. Rep. 91–31.

[4] M.J. Schmitt and C.J. Elliott, *Phys. Rev.* **A41** (1990) 3853. https://doi.org/10.1103/PhysRevA.41.3853

[5] J.B. Murphy *et al.*, *Opt. Commun.* **53** (1985) 197. https://doi.org/10.1016/0030-4018(85)90331-1

[6] F. Kuypers, *Klassische Mechanik* (VCH, Weinheim, 1990).

[7] T.M. Tran and J.S. Wurtele, LRP 392/90, CRPP report, Lausanne, Switzerland, 1990.

[8] D.A. Edwards and M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (John Wiley and Sons, New York, 1993). https://doi.org/10.1002/9783527617272

[9] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980).

[10] C. Pellegrini, in *Free Electron Lasers*, Eds. S. Martellucci and A.N. Chester (Plenum Press, New York, 1983).

[11] B.M. Kincaid, *J. Appl. Phys.* **48** (1977) 2684. https://doi.org/10.1063/1.324138

[12] W.B. Colson, *IEEE J. Quantum Electron.* **QE17** (1981) 1417. https://doi.org/10.1109/JQE.1981.1071273

[13] J.M.J. Madey, *Il Nuovo Cimen.* **50B** (1979) 64. https://doi.org/10.1007/BF02737622

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Bunch-length Compressors

S. Di Mitri
Elettra – Sincrotrone Trieste, Trieste, Italy

Abstract
An introduction to radio-frequency and magnetic bunch length compression of ultra-relativistic particle beams in linear accelerators is given, with a treatment of the single-particle motion up to the second order, and attention to the production of high peak current bunches for free-electron lasers.

Keywords
Electron bunches; radio-frequency linear accelerators; electron optics.

1 Introduction

There is a growing demand for generating and transporting very short, high charge density electron bunches. Applications range from light sources driven by radio-frequency linear accelerators (RF linacs) such as free-electron lasers (FELs), to future linear colliders and novel electron-beam-driven acceleration schemes, e.g., dielectric- and plasma-wake-field-driven accelerators. The generation of hundreds of amperes peak current electron bunches directly out of an electron source is in conflict with the production of small transverse emittance beams, due to the repulsive interparticle Coulomb interactions ('space-charge' forces) that are especially effective at low beam energies. It is therefore preferable to create only a few tens of amperes peak current bunches at the source, such as an RF photo-injector, in order to dilute the charge density, and thereby ensure small transverse emittances. Beam manipulations are implemented then in the downstream transport line, at higher beam energies, in order to obtain short electron bunches while preserving the transverse emittance at the injector level. The process of manipulating an electron beam so to enhance its peak current is called, in short, bunch compression.

In this chapter, we introduce the reader to bunch-length compression by means of RF and magnetic insertions, with an analytical treatment of the single-particle motion up to second order in the particle coordinates. We pay special attention to the production of high peak current bunches for high-gain FELs. Other schemes aimed to produce short bunches have been proposed, either with a special design of the electron source or by selecting only one part of the bunch, e.g., via energy-dispersive collimation or spoiling. These latter techniques, however, are not addressed in this chapter.

2 Why FELs require high peak current bunches

Over the last decade several linac-driven FELs in the ultra-violet (UV) and X-ray wavelength ranges have been built, have met their design specifications and are now operating reliably in several laboratories around the world [1, 2]. One of the main factors contributing to this successful development has been the ability to create, accelerate, transport and control electron bunches of very high brightness.

The six-dimensional (6D) energy-normalized electron-beam brightness is defined as the total bunch charge divided by the product of the root-mean-square (rms) horizontal, vertical and longitudinal normalized emittances, barring numerical factors that can be found in the literature. Essentially, it is the beam charge density in the 6D phase space. For simplicity, particle motion is intended to be uncoupled, and each transverse emittance is meant to be 'projected', i.e., it is computed over the particles' coordinates projected onto the longitudinal z-coordinate internal to the bunch [3]. The normalized

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.363

transverse emittances scale as the product of beam size and angular divergence. The normalized longitudinal emittance scales as the product of bunch length and energy spread, the latter being in fact the particles' spread in longitudinal momentum. The transverse normalized emittances are invariant under acceleration and linear transport, presuming that collective effects, such as space-charge forces, can be neglected. The same is true for the longitudinal normalized emittance if the energy spread is purely uncorrelated (i.e., not correlated with z) and particles are in the ultra-relativistic approximation: in this case, neither the bunch length, σ_z , nor the beam energy spread, σ_{E} , vary during acceleration.

The presence of non-linear motion and collective effects along the beam-delivery system may dilute the rms normalized emittances from their values at the injection point [2]. By introducing an effective degradation factor $\varsigma \geq 1$ in each plane of particle motion so that $\varepsilon_{\mathrm{n}x,\mathrm{f}} = \varsigma_x \gamma_0 \varepsilon_{x,0}, \varepsilon_{\mathrm{n}y,\mathrm{f}} = \varsigma_y \gamma_0 \varepsilon_{y,0}$ and $\varepsilon_{\mathrm{n}z,\mathrm{f}} = \sigma_{z,\mathrm{f}} \sigma_{\mathrm{E,f}} = \varsigma_z \sigma_{z,0} \sigma_{\mathrm{E},0}$, we are able to relate the 6D normalized brightness at the undulator, $B_{\mathrm{n,f}}$, to the one at the linac injection, $B_{\mathrm{n},0}$:

$$B_{\mathrm{n,f}} = \frac{Q}{\varepsilon_{\mathrm{n}x,\mathrm{f}} \varepsilon_{\mathrm{n}y,\mathrm{f}} \varepsilon_{\mathrm{n}z,\mathrm{f}}} = \frac{Q}{\varsigma_x \varsigma_y \varsigma_z \gamma_0^2 \varepsilon_{x,0} \varepsilon_{y,0} \sigma_{z,0} \sigma_{\mathrm{E},0}} = \frac{B_{\mathrm{n},0}}{\varsigma_x \varsigma_y \varsigma_z} \ . \tag{1}$$

In the ideal case of vanishing non-linear and collective effects, $\varsigma_x, \varsigma_y, \varsigma_z \to 1$, and the 6D normalized brightness is *preserved* at the injector level under acceleration and bunch-length compression.

The importance of electron-beam brightness for linac-driven high-gain FELs is underlined by the characteristic FEL parameter ρ , through which most of the one-dimensional (1D) FEL dynamics can be depicted [4]. In the so-called 1D, 'cold-beam' limit, where the effects on the FEL output of electron-beam energy spread, transverse emittance and radiation diffraction are all neglected, the radiation peak power at the resonant wavelength grows exponentially along the undulator with a gain length $L_{\mathrm{G}} = \lambda_{\mathrm{u}} \big/ \left(4\pi\sqrt{3}\rho\right)$, where

$$\rho = \left(\frac{\Omega_{\mathrm{p}} \lambda_{\mathrm{u}} a_{\mathrm{w}} [JJ]}{8\pi c \gamma} \right)^{2/3} = \frac{1}{2\gamma} \left(\frac{I}{I_{\mathrm{A}}} \right)^{1/3} \left(\frac{\lambda_{\mathrm{u}} a_{\mathrm{w}} [JJ]}{2\pi \sigma_x} \right)^{2/3} \ , \tag{2}$$

Ω_{p} being the plasma frequency, γ the relativistic Lorentz factor for the electron-beam mean energy, λ_{u} the undulator period length, I the electron-bunch peak current, I_{A} = 17045 A the Alfven current, σ_x the standard deviation of the (assumed round) electron-beam transverse size; [JJ] is the undulator–radiation coupling factor [5], equal to 1 for a helically polarized undulator and to $\left[J_0(\xi) - J_1(\xi)\right]$ for a plane-polarized undulator, where J_0 and J_1 are Bessel functions of the first kind with argument $\xi = K^2 / \left(4 + 2K^2\right)$. Here a_{w} and K are defined in the expression for the FEL fundamental wavelength of emission [6]:

$$\lambda = \frac{\lambda_{\mathrm{u}}}{2\gamma^2} \left(1 + a_{\mathrm{w}}^2\right) \ , \tag{3}$$

with a_{w} = K for helically and a_{w} = $K/\sqrt{2}$ for plane-polarized undulators, and $K = eB_0\lambda_{\mathrm{u}} / \left(2\pi m_{\mathrm{e}} c\right) = 0.934 B_0[\mathrm{T}] \lambda_{\mathrm{u}}[\mathrm{cm}]$, in practical units, is the so-called undulator parameter, B_0 the undulator peak magnetic field, e and m_{e} the electron charge and rest mass, respectively, and c the speed of light in vacuum. K is linearly proportional to the electron's amplitude of transverse oscillation in the undulator field and is typically in the range 1–5. Equation (3) is often referred to as the FEL 'resonance condition' since it selects, for any undulator period and magnetic field strength, the necessary electron beam energy for lasing at λ .

For kA-current beams, typically $\rho \approx 10^{-3}$ in the UV range, and it drops to ~10^{-4} in the X-ray range. If the undulator length $N_u \lambda_u$, with N_u the number of undulator periods, is equal to or longer than ~$20L_G$, the conversion of electrons' kinetic energy to photon energy considerably enlarges the electron-beam energy spread, with an eventual reduction in the FEL gain [6]. When the FEL process starts up from noise in the electron-charge distribution – it is therefore said to operate in self-amplified spontaneous emission (SASE) mode [7, 8] – the associated FEL power saturates at a level $P_{\text{sat}} \approx 1.6\rho EI/e$. In spite of the low FEL extraction efficiency relative to the electron-beam power (because $\rho << 1$), an electron beam at multi-GeV energies and kA-scale peak currents is able to produce multi-GW-scale radiation peak power. For SASE devices, the value of ρ also defines the approximate number of undulator periods $N_{\text{sat}} \approx 1/\rho$ and the length $L_{\text{sat}} \approx \lambda_u/\rho$ necessary to reach power saturation. The normalized spectral bandwidth at saturation is $\Delta\omega/\omega \approx \rho$, presuming a more or less mono-energetic electron beam with little *z*-correlated energy spread.

Equation (2) suggests that a smaller beam transverse emittance, which is proportional to the square of the beam transverse size, is associated with a higher FEL gain. In fact, the most efficient electron–photon beam interaction occurs when the transverse beam phase space area and distribution match those of the emitted radiation. This is also the condition for maximum transverse coherence of FEL radiation, and it translates into an electron-beam emittance at the undulator smaller than or of the same order as the one of the diffraction-limited photon beam [9]:

$$\varepsilon_{x,y} \le \frac{\lambda}{4\pi} \ . \tag{4}$$

By substituting Eqs. (3) and (4) (with the beam emittance at the diffraction limit) into Eq. (1), we can establish a relationship between $B_{\text{n,f}}$ and λ [10]:

$$B_{\text{n,f}} = \frac{Q}{\varepsilon_{\text{n}x,\text{f}}\varepsilon_{\text{n}y,\text{f}}\varepsilon_{\text{n}z,\text{f}}} = \frac{I}{c\sigma_{\text{E}}\gamma_0^2\varepsilon_0^2} \approx \frac{I}{\lambda\sigma_{\text{E}}}\frac{32\pi^2}{c\lambda_{\text{u}}\left(1+a_{\text{w}}^2\right)} \ . \tag{5}$$

It is worth noticing that the ratio I/σ_{E} is invariant under acceleration and compression, when collective effects are ignored. So, for any given undulator, efficient lasing at shorter λ requires a higher $B_{\text{n,f}}$. This is confirmed by Fig. 1, where $B_{\text{n,f}}$ of designed and existing single-pass linac-driven FEL facilities is shown as a function of the maximum photon energy (i.e., minimum fundamental wavelength) from UV to X-rays.

At this point one might wonder why a high peak current is required at the undulator, if the FEL dynamics appears to be so tightly related to the 6D normalized electron-beam brightness, the latter one being approximately invariant for a well-set beam-delivery system. The fact is that, as long as the beam effective rms *relative* energy spread is smaller than ρ (here 'effective' refers to the contribution of both beam transverse emittance and spread in longitudinal momentum to the lack of synchronism between the electrons and the radiation emitted in the undulator [6]), the FEL dynamics is well depicted by the 1D cold-beam model in Eq. (2). We therefore see that the higher the peak current, the larger the FEL gain, the shorter the saturation length. The relatively weak dependence of ρ on *I* is one of the reasons why ρ typically spans over one order of magnitude only, for lasing from UV to hard X-rays, whereas *I* has typically to be increased from a few tens of amperes out of photo-injectors to kA level at the undulator. With the common prescription of *relative* energy spread $\sigma_\delta < 0.5\rho$ in mind for maximum FEL gain [6], it makes sense to collapse the electron-beam quality into the *5D* normalized brightness, which is just $B_{\text{n,f}} \times \sigma_{\text{E}}$. This quantity is *not* invariant under compression, and is actually linearly proportional to the bunch length total compression factor.

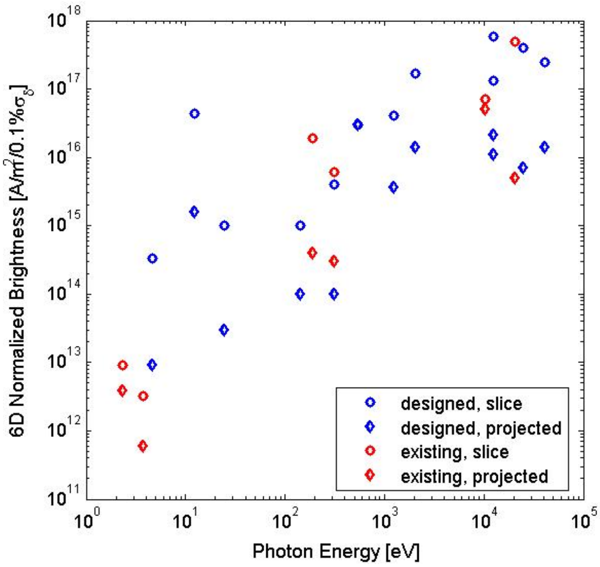


Fig. 1: 6D normalized electron-beam brightness (Eq. (1)) vs. maximum photon energy at fundamental FEL emission, for facilities in the UV to X-rays, in design stage (blue) or operational (red), up to 2013. From lower to higher photon energies, "existing" facilities are: SPARC (Italy), SDUV-FEL (China), FLASH-I (Germany), FERMI (Italy), LCLS (USA), SACLA (Japan). The brightness refers to the projected (circle) or slice value in the bunch (diamond). Published in Ref. [2]. Copyright of Elsevier.

In summary, bunch-length compression is required for increasing the bunch peak current from the injection level by usually one to three orders of magnitude, and eventually driving a *high-gain* FEL, given that transverse emittances and relative energy spread are kept small at the undulator. In the following we will focus on RF and magnetic bunch-length compression of ultra-relativistic electrons. As we will see, RF compression is achieved by exploiting the longitudinal slippage of electrons in an RF linac, at beam energies typically lower than ~100 MeV. A magnetic compressor is made of an RF linac followed by a magnetic insertion including dipole magnets, and the electrons' longitudinal slippage happens in the magnetic insertion only. Magnetic compression is commonly achieved at beam energies higher than ~100 MeV. An exhaustive literature on bunch-length compression is available, and some fundamental references are provided in this chapter. Still, we will introduce the reader to the salient topics related to the single-particle dynamics, and will illustrate the basic equations of motion.

3 Particle motion in a RF linac

We consider the motion of ultra-relativistic particles in an RF linac made, for example, of copper structures with inner iris. We assume each structure made of identical cylindrical cells; the RF power flows through the cells, and is eventually extracted on a load. Such structures behave like waveguides of cylindrical symmetry, and the longitudinal electric field component, which has in general a radial dependence, is a superposition of n field harmonics characterized by an angular RF frequency ω and by an RF wave-number k [11]:

$$\begin{aligned}E_z^{\mathrm{TW}} &= \sum_{n=-\infty}^{+\infty} a_n J_n\left(k_{r,n} r\right)\cos\left(\omega t\left(z\right) - k_n s + \varphi\right) \cong E_{z,0}^{\mathrm{TW}}\cos\left(\omega t_{\mathrm{syn}} + \omega\Delta t - ks + \varphi\right) \\ &= E_{z,0}^{\mathrm{TW}}\cos(\omega t_{\mathrm{syn}} - ks + \varphi + kz) \equiv E_{z,0}^{\mathrm{TW}}\cos(\phi_{\mathrm{rf}} + kz) \ ,\end{aligned} \tag{6}$$

where the generic particle time coordinate $t(z)$ was expanded in the arrival time t_{syn} of the reference (or synchronous) particle, e.g. the bunch centroid, plus the arrival time of the generic particle with respect to it. We then used the identity $\omega\Delta t = kz$. The z-coordinate runs inside the bunch, with z = 0 for the reference particle. The s-coordinate runs along the electric axis of the cell. The arbitrary phase φ determines the arrival time of the reference particle relative to the electric field inside the cell. Finally, we defined the RF phase $\phi_{\mathrm{rf}} = \omega t_{\mathrm{syn}} - ks + \varphi$, which tends to be constant for ultra-relativistic beams.

The last term of Eq. (6) describes the fundamental on-axis mode of the longitudinal electric field in a 'travelling wave' (TW) accelerating structure. In fact, we assume that the transverse beam sizes are much smaller than the structure inner radius, and that the beam is well centred on the structure's electric axis. Moreover, most of the acceleration is provided by the fundamental mode of the field.

If the structure is made in order to allow reflections of the RF power, we calculate the resulting net positive interference of two counter-propagating waves as follows:

$$\begin{aligned} E_z^{\mathrm{SW}} &\cong E_{z,0}^{\mathrm{TW}} \cos\left(\omega t(z) - ks + \varphi\right) + E_{z,0}^{\mathrm{TW}} \cos\left(\omega t(z) + ks + \varphi\right) \\ &= 2E_{z,0}^{\mathrm{TW}} \cos\left(\omega t_{\mathrm{syn}} + \omega \Delta t + \varphi\right) \cos\left(-ks\right) \\ &\equiv E_{z,0} \cos(\omega t_{\mathrm{syn}} + kz + \varphi) \cos\left(ks\right) \ . \end{aligned} \tag{7}$$

Equation (7) describes the on-axis longitudinal electric field in a 'standing wave' (SW) structure. We now assume that the electric field is approximately uniform across the cell gap (i.e., cos(ks) ≈ const.), and that the reference particle's velocity v, as well as those of all other particles, does not change substantially during acceleration (ultra-relativistic limit). Then the generic particle energy gain through a cell of coordinates [–g/2, g/2] is

$$\begin{aligned} \Delta E(g,z) &= -e \int_{-g/2}^{g/2} E_z^{\mathrm{SW}} \mathrm{d}s = -eE_{z,0} \int_{-g/2}^{g/2} \mathrm{d}s \left[\cos(\omega t_{\mathrm{syn}}) \cos(kz+\varphi) - \sin(\omega t_{\mathrm{syn}}) \sin(kz+\varphi)\right] \\ &= -eE_{z,0} \int_{-g/2}^{g/2} \mathrm{d}s \left[\cos(\omega s/v)\cos(kz+\varphi) - \sin(\omega s/v)\sin(kz+\varphi)\right] \\ &= -eE_{z,0} g \frac{\sin\left(\omega g/2v\right)}{\left(\omega g/2v\right)} \cos(kz+\phi_{\mathrm{rf}}) \equiv e\Delta V_0\left(g\right) T \cos(\phi_{\mathrm{rf}}+kz) \ . \end{aligned} \tag{8}$$

T is called the 'transit-time factor'; it is always less than 1 (typically in the range 0.85–0.95) and it describes the reduction of energy gain because of the time variation of the electric field along the cell, when the beam traverses the cell with a finite velocity $v < c$. In the following, for the sake of brevity, we will collapse T into an effective electric potential ΔV_0. It is worth noticing that the final expression for the energy gain in Eq. (8) applies to a TW as well with $T = 1$, where the TW structure is assumed to be tuned in order to maintain the synchronism between the RF field and the reference particle.

Finally, we notice that the first equality in Eq. (7) can be rewritten as $E_z^{\mathrm{SW}}\left(z=0\right) \cong E_{z,0}^{\mathrm{TW}}\left[\cos\left(\phi_{\mathrm{rf}}\right) + \cos\left(\phi_{\mathrm{rf}} + 2ks\right)\right]$. This tells us that a particle travelling in synchronism with the forward wave at the light speed experiences both a constant accelerating force and an oscillating force from the backward wave. The latter has a double oscillation frequency, and it does not contribute to beam motion on average. Henceforth, we keep the notation according to which beam acceleration (i.e., acceleration sampled by the reference particle) is maximum for $\phi_{\mathrm{rf}} = 0$: in this case the beam is said to be 'on-crest' of the RF wave. As we will see in the next section, magnetic bunch-length compression requires a correlation of the particles' energy with their longitudinal positions inside the bunch (see Figs. 3 and 4 below), and such a correlation is established by operating the linac 'off-crest', namely at an RF phase $-\pi < \phi_{\mathrm{rf}} < 0$ or $0 < \phi_{\mathrm{rf}} < \pi$, depending on the geometry of the downstream magnetic insertion. The special point $\phi_{\mathrm{rf}} = \pm\pi/2$ is commonly called 'zero crossing'. Accelerated either on-crest or off-crest, we assume that the beam longitudinal phase space (z, E) is mainly determined by the curvature imposed by the cosine-like behaviour of the accelerating field.

The coefficient used to quantify the linear correlation in (z, E) is named 'linear energy chirp', and it can be evaluated by expanding the electric field-induced energy gain in Eq. (8) to first order in z:

$$\begin{aligned} h &\equiv \frac{1}{E_0}\frac{\mathrm{d}E}{\mathrm{d}z} \cong \frac{1}{E_\mathrm{i} + e\Delta V_0 \cos\phi_{\mathrm{rf}}}\frac{\mathrm{d}}{\mathrm{d}z}\left[E_\mathrm{i} + e\Delta V_0 \cos(\phi_{\mathrm{rf}}) - e\Delta V_0 kz \sin(\phi_{\mathrm{rf}}) + o(z^2)\right] \\ &= -\frac{e\Delta V_0 k \sin\phi_{\mathrm{rf}}}{E_\mathrm{i} + e\Delta V_0 \cos\phi_{\mathrm{rf}}} \ , \end{aligned} \tag{9}$$

where the beam is injected into the linac with a mean energy E_i. When the beam energy spread induced by the RF curvature is much larger than the uncorrelated energy spread, which depends on the process of beam generation, we may estimate $|h| \approx \sigma_\delta / \sigma_z$. As a consequence, as long as the beam correlated energy spread is constant when the bunch length is shortened (or lengthened) in a magnetic insertion, the energy chirp is increased (or lowered) by the same compression factor.

4 Particle motion in a magnetic chicane

By evaluating the Lorentz force for a particle with longitudinal momentum p_z traversing a dipole magnet with uniform vertical field B_0, one finds that the radius of curvature R of the particle's trajectory depends on the magnetic field and the momentum according to $p_z[\mathrm{GeV}/c] = 0.2998 B_0[\mathrm{T}] R[\mathrm{m}]$. This suggests that particles with different momentum will follow different (longer vs. shorter) orbits. Since the longitudinal velocity of all particles is assumed to be very close to c independently from their spread in energy, the particles will arrive at a longitudinal position s downstream of the magnet at different times. In other words, the longitudinal coordinate z of particles inside the bunch is changed. We therefore envision a way to shorten or lengthen the bunch, with a suitable arrangement of energy spread and dipole magnets. The former is manipulated with an RF linac as depicted in the previous section. In particular, the energy spread is correlated along the bunch so that, for example, less energetic particles in the bunch head will follow orbits longer than more energetic particles in the bunch tail, as shown in Fig. 2. At the exit of the magnetic insertion, the bunch head and tail will have been caught towards the bunch centre, and the bunch length will be shortened. The final aim of bunch-length compression, as explained in Section 2, is that of increasing the bunch peak current.

In linacs driving single-pass FELs, it is usually convenient to maintain the beam trajectory on a straight path, which is also the electric axis of the accelerating structures. For this reason dipole magnets devoted to bunch compression are arranged in geometries that do not provide a net beam deflection, such as a four-dipole chicane. We first consider a symmetric geometry made of identical dipoles, as shown in Fig. 2. In each dipole of length l_d, the bending angle of the reference (on-momentum) particle is $|\theta_0| = l_\mathrm{d}/R = eB_0 l_\mathrm{d}/p_{z,0}$, and the total deflection angle through the chicane is $\theta_{0,1} + \theta_{0,2} + \theta_{0,3} + \theta_{0,4} = 0$. For a generic off-momentum particle the bending angle is $\theta = eB_0 l_\mathrm{d}/(p_{z,0} + \Delta p_z) \equiv \theta_0/(1+\delta)$, and still the net deflection through the chicane is zero. The same result holds for an expansion of the total bending angle to any order in δ.

We now assume that when particles are travelling in the drift section upstream of the chicane, their transverse position and angular divergence do not depend on their energy difference, i.e. the energy-dispersion function and its first derivative w.r.t. s, evaluated at the entrance of the chicane, are both zero [12]: $\eta_x(s=0)=\Delta x/\delta\equiv 0, \eta'_x(s=0)=\Delta x'/\delta\equiv 0$. For symmetry, those functions are also zero at the chicane exit. Naively, this means that particles lying on a line at the entrance of the chicane and with no angular divergence still lie on a line at the chicane exit, regardless of their spread in energy (see Fig. 2). Such a property defines the chicane as an 'achromatic' line. Since the total bending angle of an off-momentum particle through the chicane is zero at all orders in δ, as demonstrated above, such a chicane is achromatic at all orders (barring magnets' errors or geometry imperfections).

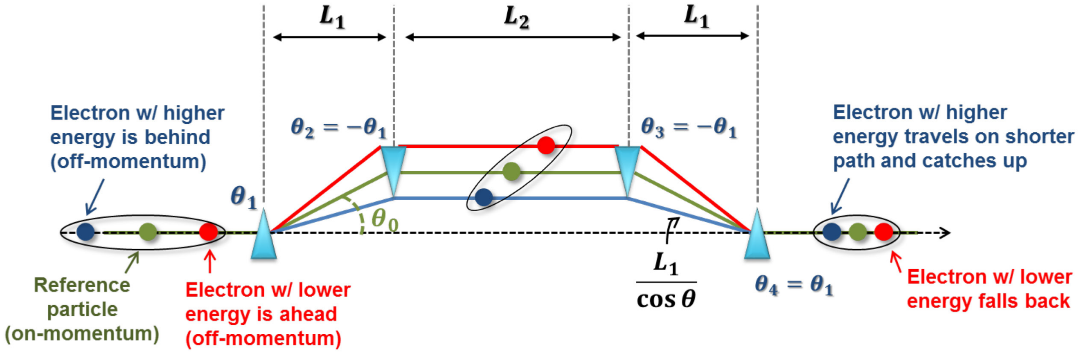


Fig. 2: Geometry (not to scale) of a four-dipole symmetric magnetic chicane, and particles' motion through it for the case of bunch-length compression. See context for the meaning of symbols.

In accelerator physics, the evolution of particle 6D coordinates through an arbitrary beam line is commonly depicted through the matrix formalism [12], i.e., each element of the beam line is depicted through a matrix whose terms depend on the element's parameters and geometry. A beam line made of consecutive elements is represented by a matrix that is the ordered product of the individual ones. Thus, the dependence of a particle's z-coordinate at the exit of the chicane on its momentum deviation can be written as $z_f(\delta)=z_i+R_{56}\delta$, with R_{56} the chicane matrix element. We now calculate R_{56} looking at the particle longitudinal slippage $\Delta z = z_f - z_i$, through the chicane, for the geometry shown in Fig. 2, and assuming once more ultra-relativistic particles.

We neglect for the moment the length of dipole magnets with respect to other lengths of the chicane involved. The path length of an off-momentum particle through the chicane is $s_z=\frac{2L_1}{\cos\theta}+L_2$, and the one of the on-momentum particle $s_{z,0}=\frac{2L_1}{\cos\theta_0}+L_2$. Their path-length difference, i.e. the longitudinal slippage of the off-momentum particle w.r.t. the on-momentum one, at the exit of chicane, is

$$\Delta s = s_z - s_{z,0} = 2L_1\left(\frac{1}{\cos\theta}-\frac{1}{\cos\theta_0}\right)\cong L_1\left(\theta^2-\theta_0^2\right)+o\left(\theta^4\right)=-L_1\theta_0^2\left[1-\frac{1}{(1+\delta)^2}\right]+o\left(\theta^4\right) \tag{10}$$
$$\cong -2L_1\theta_0^2\delta+o\left(\theta^4,\delta^2\right).$$

Equation (10) tells us that $R_{56}=\frac{\Delta z}{\delta}\cong -2L_1\theta_0^2$ for $\theta_0 << 1$; namely, at first order in δ the bunch-length shortening is quadratic with the dipoles' bending angle, and does not depend on the drift length in between the inner dipoles (in that region, particles are travelling on parallel trajectories at the same velocity, and therefore they do not slip one with respect to the others). When the path length in non-zero-length dipoles is included, we find $R_{56} \cong -2\theta_0^2\left(L_1 + \frac{2}{3}l_{\mathrm{d}}\right)$. More accurate expressions can be found in the literature [13].

There is an intrinsic connection between R_{56} and the energy-dispersion function (henceforth, simply dispersion), which we explicit below. Let us introduce the 'momentum compaction' factor of a dispersive beam line, i.e., the particle's relative variation of path length per relative momentum deviation, $\alpha_{\mathrm{c}} \equiv \frac{\Delta L/L_0}{\delta}$. In a dipole magnet we have $\Delta L = (R_0 + \Delta x)\theta - R_0\theta_0 \cong \Delta x\theta_0$ for $\Delta\theta = (\theta_0 - \theta) \to 0$, with R_0 the curvature radius of the on-momentum particle, and Δx the lateral distance in the bending plane of the off-momentum particle from the reference trajectory. Hence, we obtain $\alpha_{\mathrm{c}} \cong \frac{1}{R_0}\frac{\Delta x}{\delta}$. For R_{56} we find

$$R_{56} = \frac{\Delta z}{\delta} \cong \frac{\Delta L}{\delta} = \alpha_{\mathrm{c}} L_0 \to R_{56}(s) = \int_0^s \alpha_{\mathrm{c}}(s')\mathrm{d}s' = \int_0^s \frac{1}{R(s')}\frac{\Delta x(s')}{\delta}\mathrm{d}s' = \int_0^s \frac{\eta_x(s')}{R(s')}\mathrm{d}s' \, , \tag{11}$$

where we have retained a generic dependence of the bending radius on the s-coordinate, and we used the definition of dispersion function introduced above. Equation (11) holds for an arbitrary beam line, and it shows that longitudinal slippage of particles only happens in the presence of curvature, i.e., inside dipole magnets. At the same time, manipulation of the dispersion function in between consecutive dipoles (e.g., through a suitable distance between dipoles of a chicane, or with additional quadrupole magnets in between them) allows R_{56} of the system to be tuned.

We finally point out that if particles are not in the ultra-relativistic regime, i.e., their longitudinal velocity varies with their longitudinal momentum, then an effective particles' slippage also happens in a drift section. It can be shown that the drift section is characterized by a matrix element:

$$R_{56} = \frac{\Delta z}{\delta} = -\Delta L \frac{p_{z,0}}{\Delta p_z} = -L_0 \frac{\Delta\beta_z}{\beta_{z,0}} \frac{\beta_{z,0}}{\gamma_0^2 \Delta\beta_z} = -\frac{L_0}{\gamma_0^2} \, , \tag{11a}$$

where the suffix '0' refers to the reference particle. At low beam energies, this term may become important. If applied to a chicane, L_0 refers to the path length through the whole line.

5 Bunch-length linear compression factor

We now consider a particle motion in the chicane of Fig. 2. We differentiate the particle longitudinal slippage, evaluated through the whole chicane, and keep only terms to first order in the particle coordinates (linear approximation):

$$\begin{aligned} \mathrm{d}z_{\mathrm{f}} &= \mathrm{d}z_{\mathrm{i}} + R_{56}\mathrm{d}\delta \cong \mathrm{d}z_{\mathrm{i}} + R_{56}\frac{\mathrm{d}E}{E_0} = \mathrm{d}z_{\mathrm{i}}\left(1 + R_{56}\frac{1}{E_0}\frac{\mathrm{d}E(z)}{\mathrm{d}z_{\mathrm{i}}}\right) + R_{56}\frac{\mathrm{d}E_{\mathrm{unc}}}{E_0} \\ &= \mathrm{d}z_{\mathrm{i}}\left(1 + h_{\mathrm{i}}R_{56}\right) + R_{56}\delta_{\mathrm{unc}} \equiv \mathrm{d}z_{\mathrm{i}}/C + R_{56}\delta_{\mathrm{unc}} \, . \end{aligned} \tag{12}$$

In Eq. (12), E_0 is the electron beam mean energy at the compressor and δ is the energy deviation relative to E_0. We have split the particle energy deviation into two terms, one for the energy deviation correlated with z, which translates into the initial linear energy chirp h_{i}, and the other one for the initial uncorrelated energy deviation, δ_{unc}. Equation (12) defines the *linear* compression factor $C = (1 + h_{\mathrm{i}}R_{56})^{-1}$. It is worth noticing that $C \to \infty$ for $R_{56} = -1/h_{\mathrm{i}}$. However, even in that limit, the actual bunch length is finite and reaches the minimum rms value $\sigma_{z,\min} = R_{56}\sigma_{\delta,\mathrm{unc}}$ by virtue of a non-zero δ_{unc}. Thus, 'full' compression at higher beam energies would result in shorter minimum bunch lengths. With accepted convention, the chicane geometry in Fig. 2 provides $R_{56} < 0$, and therefore the bunch length is shortened if $h_i > 0$, namely if the bunch head has a lower energy than the bunch tail. If a non-linear energy chirp is present (E depends on z at higher orders in z), we expect $h_\mathrm{i}(z)$ to vary along the bunch, and so will C.

The uncorrelated energy spread plays an important role in the build up of the FEL instability, and for this reason it is convenient to point out its evolution during the compression process. In practical situations the linearly correlated energy spread is controlled through a proper setting of the linac RF phase. Quadratic and cubic components may require more sophisticated beam manipulations – such as acceleration through higher harmonic RF frequency structures or shaping the bunch current profile – which are not considered at the moment.

We write down the total relative energy spread of a generic particle as the sum of an uncorrelated term (δ_u) and a z-correlated term (δ_c), the latter being the beam energy chirp times the particle z-position. The *total* energy spread is assumed constant through the chicane (only magnetic fields are involved, and no frictional forces), and the final bunch length is expressed as a function of the initial one through the definition of C given above:

$$\delta_{\mathrm{tot}} = \delta_{\mathrm{u,i}} + \delta_{\mathrm{c,i}} = \delta_{\mathrm{u,i}} + h_\mathrm{i}\Delta z_\mathrm{i} \equiv \delta_{\mathrm{u,f}} + h_\mathrm{f}\Delta z_\mathrm{f} = \ldots = \delta_{\mathrm{u,f}} + Ch_\mathrm{i}\left(\frac{\Delta z_\mathrm{i}}{C} + R_{56}\delta_{\mathrm{u,i}}\right) + o\left(\delta^2, \Delta z^2\right) \; . \quad (13)$$

By equating the third and the last terms of Eq. (13), and then passing to the rms value of the quantities involved, we find $\sigma^2_{\mathrm{u,f}} = \sigma^2_{\mathrm{u,i}}\left(1 - Ch_\mathrm{i}R_{56}\right)^2 = C^2\sigma^2_{\mathrm{u,i}}$, i.e., the uncorrelated energy spread is increased by the same factor C by which the bunch length is shortened. This result is often referred to as 'preservation of longitudinal emittance' because, when the energy chirp is removed (virtually or in reality) from the phase space, the longitudinal emittance is just the product of bunch length and uncorrelated energy spread. The product is constant, in fact, in the approximation of linear motion and absence of frictional forces. An illustration of the preservation of the longitudinal emittance for 'undercompressed', 'fully compressed' and 'overcompressed' beams is given in Fig. 3.

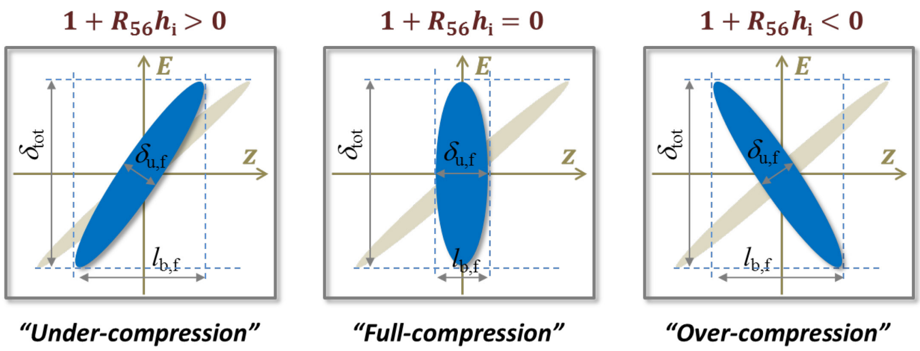


Fig. 3: Sketch of beam longitudinal phase space (ellipses) before (grey shadow) and after (blue) a magnetic chicane. Following accepted convention, $R_{56} < 0$ and bunch head is at $z < 0$; therefore, a positive chirp h_i at the entrance of the chicane (bunch head at lower energy) leads to bunch shortening. The 'undercompression' scenario on the left is the prevalent mode of operation of FEL linac drivers. When $h_\mathrm{i} = -1/R_{56}$, the longitudinal phase space at the chicane exit is upright (plot at centre), and the bunch length l_b reaches its minimum value as set by the uncorrelated energy spread δ_u. If $h_\mathrm{i} > 0$ but R_{56} is so negative that $1 + R_{56}\,h_\mathrm{i} < 0$, then the bunch head and tail flip their longitudinal positions, and the energy chirp at the chicane exit has changed its sign. This is the case of 'overcompression' (right-hand plot), in which the bunch surpasses the point of 'full compression', and therefore the final bunch length is longer than its minimum value. In the absence of frictional forces (collective effects), the total energy spread δ_{tot} is constant through the chicane, as well as the beam longitudinal emittance, represented by the area of the ellipses.

6 Bunch-length compression at second order and linearization

Quadratic and even cubic components of the energy chirp, as anticipated above, may play an important role in the compression process, as C is no longer constant through the bunch, and different longitudinal portions of the bunch (slices) may be compressed in a different manner. Such a dynamics would imply that the current profile before compression (e.g., uniform, parabolic, Gaussian, etc.) is not preserved by the compression process. The situation is additionally deteriorated by a higher order dispersion function that translates into a higher order momentum compaction (T_{566} term at second order, U_{5666} at third order etc). In order to evaluate such a non-linear dynamics, we start expanding the expression for the energy gained by a generic particle in an RF linac to second order in z. For illustration, we ignore at this stage the second-order momentum compaction in the chicane. As already done for Eq. (8), we find

$$E_1 \cong E_\mathrm{i} + e\Delta V_0 \cos\phi_\mathrm{rf} - e\Delta V_0 kz \sin\phi_\mathrm{rf} - \frac{e\Delta V_0}{2} k^2 z^2 \cos\phi_\mathrm{rf} + o\left(z^3\right) \ . \tag{14}$$

The second-order term in z of Eq. (14) can be cancelled by means of an additional RF component, but with different RF wavenumber:

$$\begin{aligned} E_2 &= E_1 + e\Delta V_\mathrm{H} \cos\left(k_\mathrm{H} z + \phi_\mathrm{H}\right) \\ &\cong E_1 + e\Delta V_\mathrm{H} \cos\phi_\mathrm{H} - e\Delta V_\mathrm{H} k_\mathrm{H} z \sin\phi_\mathrm{H} - \frac{e\Delta V_\mathrm{H}}{2} k_\mathrm{H}^2 z^2 \cos\phi_\mathrm{H} + o\left(z^3\right) \ . \end{aligned} \tag{15}$$

By comparing Eqs. (14) and (15), we find that the quadratic term generated by the additional structure(s) has to be positive, i.e., $\cos\phi_\mathrm{H} < 0$ (say, $\phi_\mathrm{H} = \pi$) and therefore the zeroth-order term from the additional linac is *decelerating* the beam. The new linac voltage has to satisfy $\Delta V_\mathrm{H} = \frac{k^2}{k_\mathrm{H}^2} \Delta V_0 \cos\phi_\mathrm{rf}$.

Thus, compensation of the second-order energy chirp ('RF curvature') and net beam acceleration can only be achieved simultaneously if the RF wavenumber of the additional linac (often named 'linearizer') is larger than the one of the baseline accelerator. The scaling of the linearizer peak voltage with the wavenumber favours this approach, as long as the ratio of wavenumbers is 1/3 or smaller. For example, a baseline RF linac running in the S-band 3 GHz RF and providing 200 MeV energy gain, can be supplied by an additional X-band 12 GHz RF structure with peak voltage at ~15 MV level.

As anticipated above, we have so far ignored the non-linear z-motion of particles through the chicane, which is depicted by $z_\mathrm{f}\left(\delta\right) = z_\mathrm{i} + R_{56}\delta + T_{566}\delta^2$ up to second order. In a more complete analysis, the energy deviation in this expression combines with the expression for the energy chirp up to second order. In this more general and realistic case, linearization does not apply to the longitudinal phase space at the entrance of the chicane only, but to the compression process as a whole, through the RF linac *and* the chicane. As a result of cancellation of all second-order terms in the particles' dynamics, we expect that the current profile at the exit of the chicane resembles the one at its entrance, just squeezed in the z-coordinate. It can be shown [14] that cancellation of all the second-order terms for the special case $\cos\phi_\mathrm{H} = -1$ implies $bR_{56} + a^2 T_{566} = 0$, with $a = -\frac{e\Delta V_0}{E_\mathrm{i}} \sin\phi_\mathrm{rf}$ and $b = -\frac{e\Delta V_0 k^2 \cos\phi_\mathrm{rf} + e\Delta V_\mathrm{H} k_\mathrm{H}^2}{2E_\mathrm{i}}$.

By imposing that the beam mean energy at the chicane, E_BC, and the final bunch length do not change w.r.t. the case of purely linear motion, and additionally ignoring the contribution of the uncorrelated energy spread to the final bunch length, we find the necessary peak voltage of the harmonic cavity [14]:

$$e\Delta V_{\mathrm{H}} = \frac{1}{\left(k_{\mathrm{H}}^2/k^2 - 1\right)} \left\{ E_{\mathrm{BC}} \left[1 + \frac{2}{k^2} \frac{T_{566}}{|R_{56}|^3} \left(1 - \frac{1}{C} \right)^3 \right] - E_{\mathrm{i}} \right\} . \tag{16}$$

Figure 4 shows beam longitudinal phase-space and peak-current profiles simulated with the 1D tracking code LiTrack [15] up to second order, in the first stage of the FERMI FEL linac. The linac upstream the four-dipole chicane is set at 26 deg S-band far from the crest. With no X-band cavity, the RF curvature leads to a current spike in the bunch head, and to a ramped current profile at much lower level. A uniform current profile is recovered with an X-band cavity voltage of –15 MV.

Although Eq. (16) is valid for a one-stage compression only, the dependence of the linearizer peak voltage on the RF wavenumber is the same for multistage compression schemes. Moreover, when two chicanes or more are adopted, the peak-voltage setting of the linearizer does not vary much because after the first chicane, at lower energy, the bunch is shorter and less vulnerable to RF curvature [16]. A semi-analytical treatment of linearization of the compression process through RF cavities in the presence of higher order beam dynamics and single-bunch collective effects (e.g., short-range geometric wake fields [17]) can be found in Ref. [18].

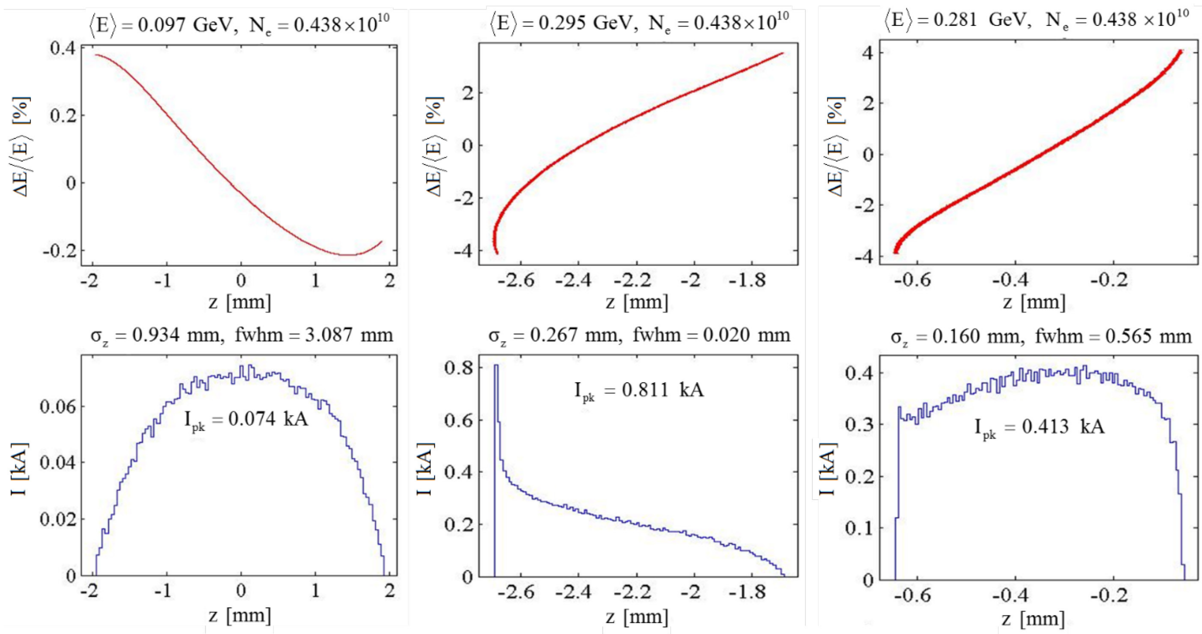


Fig. 4: Longitudinal phase space (top row) and current profile (bottom row) of a 700 pC charge bunch injected into the S-band FERMI FEL linac at 97 MeV (left). Centre plots are for the beam after magnetic compression in a four-dipole chicane, with no linearizer included. Right-hand plots are for the same beam when an X-band cavity in the upstream linac is set to –15 MV, for linearization of the compression process. Simulations were done with LiTrack code and up to second order.

Compensation of third-order terms is also possible by running the linearizer off-crest. However, the third-order energy chirp is commonly generated in the beam injector by space-charge forces (at energies typically lower than 5 MeV for photocathode RF injectors), and is of a sign [19] that is difficult to cancel without also partially cancelling the linear energy chirp necessary for compression, resulting in either inefficient acceleration or insufficient compression factor.

Alternative methods for the linearization of the compression process include passive dielectric-lined insertions or magnetic elements. In the former case, an optimum longitudinal voltage loss over the length of the bunch can be provided in order to compensate both the second-order RF time curvature and the second-order momentum compaction term [20]. Removal of second-order non-linearities in the longitudinal phase space through optical elements is typically dealt with by sextupole magnets [21–23]. Sextupoles introduce a quadratic dependence of the particle path-length difference on energy deviation through an effective T_{566} term that, if supplied with the appropriate sign, 'stretches' the curvature in phase space. This mechanism is illustrated in Fig. 5. However, if the beam has to enter the undulator chain for lasing, tight tolerances on the final beam transverse emittance make the sextupole correction in a four-dipole chicane less attractive due to possible high-order magnetic aberrations. Moreover, the use of a higher-harmonic RF field does not introduce coupling between longitudinal and transverse phase-space coordinates, unlike optical manipulation of R_{56} and T_{566} terms does. For this reason, to date most of the FEL facilities have chosen to linearize the magnetic compression process with up-frequency RF structures. In principle, sextupole-induced aberrations can be counteracted with a suitable betatron phase advance between those magnets. This approach, however, implies a more sophisticated design of the chicane [24] or a different magnetic insertion [25].

It is worth noticing that a larger T_{566} term, such as the one provided by a multistage compression scheme, may be helpful in the reduction of the quadratic energy chirp induced by longitudinal geometric wake fields excited in small-iris accelerating structures [2]. The multistage compression, however, tends to amplify the so-called microbunching instability, which implies a finally increased energy spread and modulated current profile, as discussed in the next chapter.

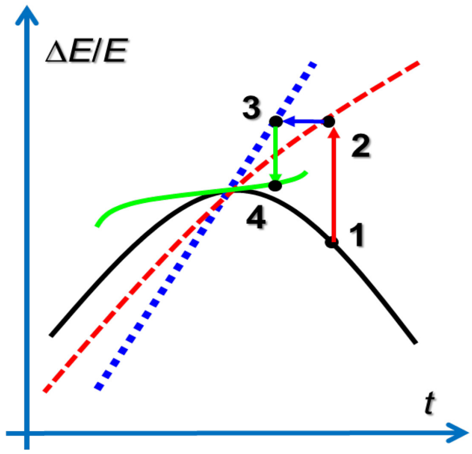


Fig. 5: Linearization of longitudinal phase space with T_{566} transport matrix element provided, e.g., by a sextupole magnet installed in a dispersive region. (1) RF curvature imprints second-order non-linearity onto the phase space ('RF curvature'). (2) Linear energy chirp is imparted to the beam with off-crest acceleration upstream of a chicane. (3) Beam passes through a sextupole magnet in the middle of the chicane, and it is time-compressed at the chicane exit. (4) Linear chirp is removed with (opposite) off-crest phasing in a linac downstream of the chicane.

Most common geometries of magnetic insertions for bunch-length compression are shown in Fig. 6. C-shape symmetric chicanes are very common because they allow remote control of the bending angle through a translation stage of the inner dipoles, for a tuning of the compression factor and balance of momentum compaction vs. coherent synchrotron radiation (CSR) instability, which is discussed in the next chapter. The inner drift section does not contribute to the compression, but it offers room for hosting beam diagnostics and scrapers or masks for beam shaping. The chicane lateral arms may host weak quadrupole magnets for the correction of spurious dispersion function due to dipole magnet errors. Different geometries (S-shape, asymmetric tuneable C-shape and double C-shape) of the chicane have been explored in order to minimize the impact of CSR emission on the beam emittance.

In symmetric C-shape geometries, all dipoles provide the same bending angle. For any given $R_{56} \cong -2\theta_0^2\left(L_1+\frac{2}{3}l_\mathrm{d}\right)$ (see Section 4 for notation), we have $T_{566} \cong -1.5 \times R_{56}$, $U_{5666} \cong 2 \times R_{56}$. For compactness, the inner dipoles can be collapsed to one magnet with double bending angle than the outer ones. In S-shape geometries, the inner dipoles provide a bending angle larger than the outer ones. Quadrupole and sextupole magnets can interleave dipole magnets.

Arcs usually provide an R_{56} term with sign opposite to that of four-dipole chicanes, and are a natural choice for compression in recirculating machines, such as energy-recovery linacs. They may offer the chance of accommodating sextupole magnets for the linearization of the compression process, with a phase advance suitable for the cancellation of geometric and chromatic aberrations (the latter ones commonly dominate because of the relatively large relative energy spread required for compression). However, additional constraints on the linear optics functions in the bending plane are required in the arcs in order to minimize or cancel the otherwise CSR-induced projected emittance growth [26]. An arc composed of N_c fodo cells (focusing and defocusing quadrupoles alternate, interleaved by identical dipoles), with betatron phase advance μ_x per cell in the bending plane, and extending over a total length L_{arc}, is characterized by $R_{56} \cong \dfrac{\theta_0^2 L_{arc}}{4N_c^2 \sin^2(\mu_x/2)}$; with no sextupoles included, $T_{566} \approx 2 \times R_{56}$ or larger. A dog-leg can be built with two consecutive arc-fodos. For the simplest two-dipole symmetric geometry, the dog-leg features $R_{56} \cong \theta_0^2 l_d/3$. A series of double- or multibend achromatic cells can be used to build up an arc of arbitrary bending angle. In a periodic arc made of N_c identical symmetric double-bend achromatic cells, $R_{56} \cong 2N_c\theta_0^2 l_d$.

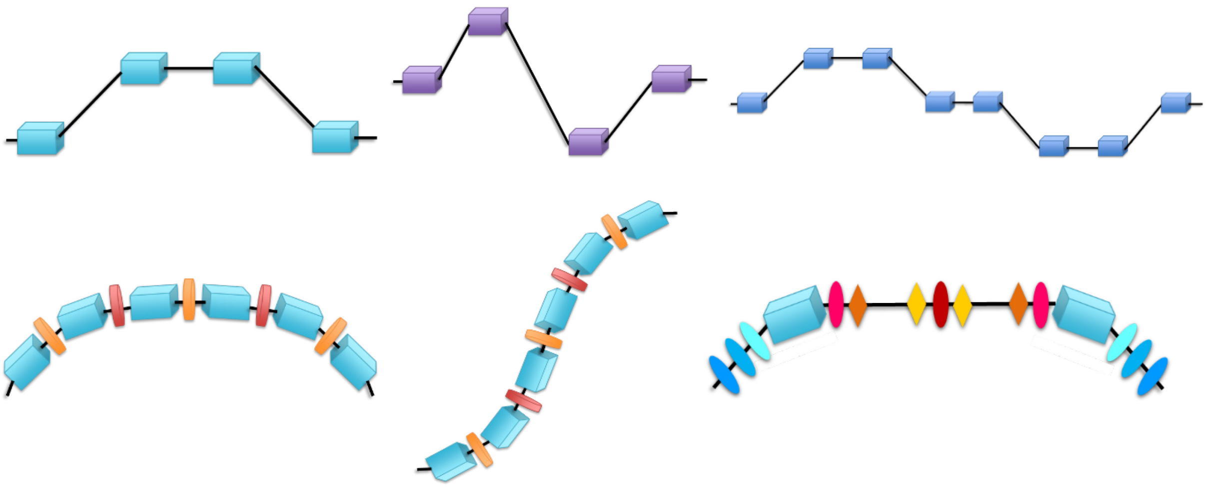


Fig. 6: Most common geometries (not to scale) of magnetic insertions for bunch-length compression. From top, left to right: C-shape, S-shape and double C-shape chicanes; bottom, arc-fodo, dog-leg-fodo and double-bend achromatic cells.

7 Jitter of bunch arrival time and compression factor

FELs usually require tight control of the electron-beam arrival time at the undulator. The shot-to-shot reproducibility of the arrival time of consecutive electron bunches, henceforth named 'arrival time jitter' (ATJ), is of great importance for multishot experiments. On the single-pulse basis, it is even more important for FELs driven by an external laser (externally seeded FELs), in order to ensure synchronism between the laser and the electron bunch. The requirement of small ATJ is particularly stringent when the electron bunch is longitudinally compressed to sub-ps durations, in order for the jitter to be (much) smaller than the bunch duration. Following Ref. [27], we introduce a model for the ATJ in the presence of magnetic compression in a four-dipole chicane, like the one sketched in Fig. 2. The error sources contributing to the ATJ we consider are: photo-injector laser arrival time on the cathode, jitter of phases and voltages of the RF linac and fluctuations of the compressor's dipole field, as may be produced by fluctuations of the power converters.

We adopt the bunch centroid as the reference particle. Its final time coordinate in the laboratory frame is $t_\mathrm{f} = t_\mathrm{i} + \dfrac{\Delta l + L}{c}$, where t_i is the reference initial arrival time, L is the straight trajectory length through the chicane (zero bending angle) and Δl is the path-length difference between the beam trajectory through the chicane with dipoles turned on and the straight trajectory. If θ << 1, one finds [13] $\Delta l \cong -R_{56}/2$. The ATJ after the beam has passed through an RF linac and one chicane is

$$\begin{aligned} c\mathrm{d}t_\mathrm{f} &= c\mathrm{d}t_\mathrm{i} + \mathrm{d}(\Delta l) = c\mathrm{d}t_\mathrm{i} + \frac{\partial(\Delta l)}{\partial E}\mathrm{d}E + \frac{\partial(\Delta l)}{\partial B}\mathrm{d}B \\ &= c\mathrm{d}t_i + \frac{\partial(\Delta l)}{\partial\theta}\frac{\partial\theta}{\partial E}dE + \frac{\partial(\Delta l)}{\partial R_{56}}\frac{\partial R_{56}}{\partial\theta}\frac{\partial\theta}{\partial B}\mathrm{d}B \cong c\mathrm{d}t_\mathrm{i} + \frac{R_{56}}{E}\mathrm{d}E - \frac{R_{56}}{B}\mathrm{d}B\ , \end{aligned} \tag{17}$$

where E is the beam mean energy at the chicane and B is the dipoles' magnetic field. The last equality makes use of the expressions for the dipole length $l_\mathrm{d} = R\sin\theta$ and of the curvature radius $R = E/(eBc)$. The differential of the beam energy w.r.t. the variation of the peak voltage, RF phase and arrival time at the linac entrance, is $\mathrm{d}E = e\cos\phi\mathrm{d}V - eV\sin\phi\mathrm{d}\phi - ecVk_\mathrm{rf}\sin\phi\mathrm{d}t_\mathrm{i}$. Substituting it into Eq. (17), and also introducing the linear compression factor C (see Eq. (12)), we obtain

$$\begin{aligned} \mathrm{d}t_\mathrm{f} &\cong \mathrm{d}t_\mathrm{i} + \frac{R_{56}}{c}\left(\frac{\cos\phi}{E}e\mathrm{d}V - \frac{eV\sin\phi}{E}\mathrm{d}\phi - \frac{eVck_\mathrm{rf}\sin\phi}{E}\mathrm{d}t_\mathrm{i} - \frac{\mathrm{d}B}{B}\right) \\ &= \frac{\mathrm{d}t_\mathrm{i}}{C} + \frac{R_{56}}{c}\left(\frac{\cos\phi}{E}e\mathrm{d}V - \frac{eV\sin\phi}{E}\mathrm{d}\phi - \frac{\mathrm{d}B}{B}\right)\ . \end{aligned} \tag{18}$$

We now move from the single-particle picture in Eq. (18) to the rms value of the ATJ, where all the jitter sources are assumed to be small and independent perturbations to the particles' motion:

$$\sigma_\mathrm{t,f}^2 \cong \left(\frac{\sigma_\mathrm{t,i}}{C}\right)^2 + \left(\frac{R_{56}}{c}\right)^2\left[\left(\frac{eV\cos\phi}{E}\right)^2\left(\frac{\sigma_V}{V}\right)^2 + \left(\frac{eV\sin\phi}{E}\right)^2\sigma_\phi^2 + \left(\frac{\sigma_B}{B}\right)^2\right]\ . \tag{19}$$

If no additional dispersive insertions are foreseen between the chicane and the undulator, the ATJ at the exit of the chicane will be frozen up to the end of the beam line. Reduction of the ATJ at the entrance of the chicane by the compression factor is due to the fact that an earlier (later) arrival of the bunch centroid to the RF field in the upstream linac translates, e.g., to a lower (higher) energy at the chicane, and therefore to a shorter (longer) path length with respect to the reference trajectory.

The linac peak voltage jitter maximally (minimally) contributes to the ATJ for the linac operated on crest (at zero crossing). For on-crest operation, the RF phase jitter term can usually be neglected as long as the bunch length is much shorter than the RF wavelength. This opposite behaviour of the two RF jitter sources as a function of the RF phase, suggests the possibility of choosing the RF linac phase in a way that, for any specified error budget, the ATJ is minimum [27]. Although quite an attractive option in principle, such an optimal linac configuration constrains the compression factor to some specific values or to a limited range, for any given setting of the magnetic chicane. In the case of multistage compression schemes, more RF settings are available that may simultaneously ensure the lowest ATJ for a design compression factor, energy spread and chicane bending angle.

A jitter of the compression factor implies a jitter of the final bunch length or of the final peak current, for initially constant bunch length and bunch charge. Owing to the fact that the linac upstream of the chicane is run off-crest in practical cases, the jitter of C is dominated by the RF phase jitter. We therefore differentiate the expression for C assuming that only the RF phase varies (namely, we neglect any variation of linac peak voltage and dipole field):

$$\Delta\left(\frac{1}{C}\right) = -\frac{\Delta C}{C^2} = \Delta\left(1+hR_{56}\right) \cong R_{56}\frac{\Delta h}{h}h \cong R_{56}\frac{\Delta\left(\sin\phi\right)}{\sin\phi}h = hR_{56}\frac{\Delta\phi}{\tan\phi} \ ,$$
$$\frac{\Delta C}{C} \cong -ChR_{56}\frac{\Delta\phi}{\tan\phi} = \left(C-1\right)\frac{\Delta\phi}{\tan\phi} \ , \tag{20}$$
$$\sigma_C^2 \cong \left(C-1\right)^2\frac{\sigma_\phi^2}{\tan^2\phi} \ .$$

Equation (20) shows that the *relative* jitter of C is proportional to C itself and that, for any given RF phase jitter, it is maximum for the phase set at zero crossing.

8 RF compression

RF compression [28] refers to two techniques of bunch-length shortening that exploit the relative longitudinal slippage of low-energy electrons as induced by a suitable arrangement of the RF linac phase. In RF 'ballistic bunching', an energy chirp is imparted to the beam in a cavity run off-crest. If the beam energy is low enough (particles are *not* in the *ultra*-relativistic limit yet), a difference in longitudinal momentum translates into a difference in longitudinal velocities, and therefore in arrival time at a given position downstream of the cavity. In order for a ~10 MeV bunch to be shortened by, say, a factor ~5, a drift length of the order of ~1 m or longer may be needed after the cavity. Bunch shortening happens if the bunch head is at lower energies than the bunch tail. Most of bunch shortening happens outside the cavity, and the energy–position correlation established in the cavity tends to be removed later in the drift section. However, the final longitudinal phase space usually shows strong non-linearities as induced by both the RF curvature and space-charge forces, which are enhanced by the increased charge density [29].

RF 'velocity bunching' differs from ballistic bunching in that the phase-space rotation happens *inside* an RF linac, still run off-crest, and the energy chirp is smoothly removed in the linac itself through electrons' longitudinal slippage and acceleration. Similarly to the ballistic bunching, the minimum bunch length achievable with this technique is determined by the distortion of the final phase space induced by RF field nonlinearities and space-charge forces.

In order to follow the longitudinal particle motion in the presence of RF compression, we assume the beam to be accelerated in a (series of) SW structure(s), which were introduced in Section 3. The evolution of the beam mean energy gain and the beam arrival time along the beam line is

$$\frac{\mathrm{d}\gamma\left(s\right)}{\mathrm{d}s} = \alpha\left[\cos\left(\phi\right)+\cos\left(\phi+2ks\right)\right] \ ,$$
$$\mathrm{d}t = \frac{\mathrm{d}s}{\beta c} = \mathrm{d}s\frac{\gamma\left(s\right)}{c\sqrt{\gamma\left(s\right)^2-1}} \ , \tag{21}$$

where we introduced the 'electron capture' parameter $\alpha \equiv \frac{eE_{z,0}}{2km_\mathrm{e}c^2}$ and the RF phase φ; the factor '2' in α disappears for TW structures. The physical meaning of such a normalized strength of the accelerating field is that, for values larger than 1, the particle dynamics shows relativistic effects within one period of the RF wave.

Following Ref. [30], we notice that dt (equivalent to beam phase) changes considerably near the cathode, where the electrons are still not or weakly relativistic. In that region s is small and we integrate Eq. (21) as follows:

$$\begin{aligned} &\gamma(s\approx 0)\approx 1+2k\alpha s\cos\phi\ , \\ &\phi(s)=\phi_0+\int_0^s \mathrm{d}\left(\omega t-ks+\varphi\right)\approx \phi_0+k\int_0^s \mathrm{d}t\frac{\gamma(s)}{\sqrt{\gamma(s)^2-1}}\ . \end{aligned} \tag{22}$$

We now insert the upper expression of Eq. (22) in the lower one, to find an approximate expression for the beam phase. The latter will be inserted into Eq. (21) again to find a more accurate expression for the energy gain. Eventually, we find

$$\begin{aligned} &\gamma(s)\cong 1+\alpha\left\{ks\cos\phi(s)-\frac{1}{2}\left[\sin(\phi)-\sin(\phi+2ks)\right]\right\}\ , \\ &\phi(s)\cong\phi_0-\frac{1}{2\alpha\cos\phi_0}\left[\sqrt{\gamma^2-1}-(\gamma-1)\right]_{(s\approx 0)}\to\phi_\infty=\phi_0-\frac{1}{2\alpha\cos\phi_0}\ . \end{aligned} \tag{23}$$

The asymptotic value of the beam phase in Eq. (23) is for $\gamma >> 1$. Figure 7 shows particle trajectories in the longitudinal phase space as depicted by Eq. (23). In that example, acceleration is maximum for $\varphi=\pi/2$.

A rough estimation of the bunch length compression factor in the limit of high beam energy can be obtained by recalling that bunch-length shortening means also compression of an incoming time or phase jitter (see Section 7). We differentiate the phase expression in Eq. (23) and find

$$C\approx\frac{\mathrm{d}\phi_0}{\mathrm{d}\phi_\infty}=\left[1-\frac{\sin\phi_0}{2\alpha\cos^2\phi_0}\right]^{-1}\ . \tag{24}$$

For example, for $\alpha=2$ and $\varphi_0=\pi/3$, $C\approx 10$.

Although neglected so far, the longitudinal particle dynamics at beam energies as low as considered in this section is intrinsically coupled to the transverse one by means of the repulsive 3D space-charge forces. In practical situations, the compression factor achieved through RF compression is limited by the tolerable transverse emittance dilution induced by space-charge forces. This effect can be mitigated by the application of external magnetic focusing, such as solenoidal fields, that counteract the particles' repulsion ('emittance compensation') [31].

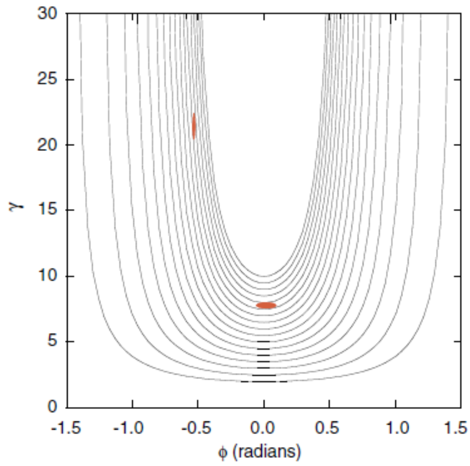


Fig. 7: Phase-space apparent rotation in RF sinusoidal accelerating field, leading to bunch-length shortening for a bunch injected near the phase of zero crossing. Published in Ref. [29]. Copyright of American Physical Society.

Acknowledgement

The author wishes to thank M. Venturini for his invaluable guidance through the physics of bunch-length compressors, and for his contribution to Figs. 2 and 3 of this chapter. P. Craievich and M. Cornacchia are acknowledged for careful reading of this work and for suggestions.

References

[1] K. Cohn *et al*., Proc. 37th Free Electron Laser Conf., WEP016, Daejeon, Korea, 2015.

[2] S. Di Mitri and M. Cornacchia, *Phys. Rep.* **539** (2014) 1. https://doi.org/10.1016/j.physrep.2014.01.005

[3] C. Lejeune and J. Aubert, in *Applied Charged Particle Optics*, Ed. A.L. Septier (Academic Press, New York, 1980), pp. 159–259.

[4] R. Bonifacio, C. Pellegrini and L.M. Narducci, *Opt. Commun.* **50** (1984) 373. https://doi.org/10.1016/0030-4018(84)90105-6

[5] W.B. Colson, *IEEE J. Quant. Electron.* **17** (1981) 1417. https://doi.org/10.1109/JQE.1981.1071273

[6] C. Pellegrini, *Eur. Phys. J. H* **37** (2012) 659. https://doi.org/10.1140/epjh/e2012-20064-5

[7] A.M. Kondratenko and E.L. Saldin, *Part. Accel.* **10** (1980) 207.

[8] P. Sprangle, C.-M. Tang and W.M. Manheimer, *Phys. Rev. A* **21** (1980) 302. https://doi.org/10.1103/PhysRevA.21.302

[9] K.-J. Kim, *Nucl. Instrum. Methods Phys. Res. A* **246** (1986) 71. https://doi.org/10.1016/0168-9002(86)90048-3

[10] S. Di Mitri, *Photonics* **2** (2015) 317. https://doi.org/10.3390/photonics2020317

[11] P.M. Lapostolle and A.L. Septier, Eds., *Linear Accelerators* (North-Holland, Amsterdam, The Netherlands, 1970), pp. 30–47.

[12] K.L. Brown, *Adv. Part. Phys.* **1** (1968) 71 and SLAC-75 Rev. 4 (1982).

[13] P. Castro, DESY Technical Note TN-2003-01 (2003).

[14] P. Emma, LCLS Report LCLS-TN-01-1 (2001).

[15] K.L.F. Bane and P. Emma, Proc. 2005 Particle Accelerator Conf., Knoxville, TN, USA, 2005, pp. 4266–4268. https://doi.org/10.1109/PAC.2005.1591786

[16] K. Flöttmann, T. Liberg and Ph. Piot, DESY Technical Note TESLA-FEL-2001-06 (2001).
[17] K.L.F. Bane, *Int. J. Mod. Phys. A* **22** (2007) 3736. https://doi.org/10.1142/S0217751X07037391
[18] I. Zadorodnov and M. Dohlus, *Phys. Rev. ST Accel. Beams* **14** (2011) 014403. https://doi.org/10.1103/PhysRevSTAB.14.014403
[19] M. Cornacchia *et al.*, *Phys. Rev. ST Accel. Beams* **9** (2006) 120701. https://doi.org/10.1103/PhysRevSTAB.9.120701
[20] P. Craievich, *Phys. Rev. ST Accel. Beams* **13** (2010) 034401. https://doi.org/10.1103/PhysRevSTAB.13.034401
[21] P. Piot, D.R. Douglas and G.A. Krafft, *Phys. Rev. ST Accel. Beams* **6** (2003) 030702. https://doi.org/10.1103/PhysRevSTAB.6.030702
[22] R.J. England *et al.*, *Phys. Rev. ST Accel. Beams* **8** (2005) 012801. https://doi.org/10.1103/PhysRevSTAB.8.012801
[23] P. Evtushenko, S. Benson and D. Douglas, Proc. 10th European Workshop on Beam Diagnostics and Instrumentation, MOOB04, Hamburg, Germany, 2011.
[24] Y. Sun, *et al.*, *Phys. Rev. ST Accel. Beams* **17** (2014) 110703. https://doi.org/10.1103/PhysRevSTAB.17.110703
[25] S. Thorin *et al.*, Proc. 32nd Int. Free Electron Laser Conf., WEPB34, Malmö, Sweden, 2010.
[26] S. Di Mitri and M. Cornacchia, *Eur. Phys. Lett.* **109** (2015) 62002. https://doi.org/10.1209/0295-5075/109/62002
[27] P. Craievich *et al.*, *Phys. Rev. ST Accel. Beams* **16** (2013) 090401. https://doi.org/10.1103/PhysRevSTAB.16.090401
[28] L. Serafini and M. Ferrario, *Physics of, and Science with, the X-Ray Free-Electron Laser*, Eds. S. Chattopadhyay, M. Cornacchia, I. Lindau and C. Pellegrini (AIP Conf. Proc. 581) (AIP, New York, 2001).
[29] S.G. Anderson *et al.*, *Phys. Rev. ST Accel. Beams* **8** (2005) 014401. https://doi.org/10.1103/PhysRevSTAB.8.014401
[30] K.-J. Kim, *Nucl. Instrum. Methods Phys. Res. A* **275** (1989) 201. https://doi.org/10.1016/0168-9002(89)90688-8
[31] M. Ferrario *et al.*, *Phys. Rev. Lett.* **104** (2010) 054801. https://doi.org/10.1103/PhysRevLett.104.054801

Bibliography

Proceedings of the CAS-CERN Accelerator School: Fifth General Accelerator Physics Course, 7 - 18 Sep 1992, Jyväskylä, Finland, edited by S. Turner, CERN-1994-001 (CERN, Geneva 1994). https://doi.org/10.5170/CERN-1994-001.

I.S. Ko and W. Chou, Eds., ICFA Beam Dynamics Newsletter No. 38 (2005).

P. Schmuser, M. Dohlus and J. Rossbach, in *Ultraviolet and Soft X-ray Free-Electron Lasers*, Eds. A.F. Chiba *et al.* (Springer Tracts Mod. Phys. 229) (Springer, Berlin, 2008).

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Coherent Synchrotron Radiation and Microbunching Instability

S. Di Mitri
Elettra – Sincrotrone Trieste, Trieste, Italy

Abstract
Emission of coherent synchrotron radiation (CSR) in dipole magnets by ultra-relativistic electron beams is treated. The CSR impact on beam energy distribution and transverse emittance is discussed, and methods for minimizing emittance growth are recalled. A qualitative explanation of the contribution of CSR to the microbunching instability is given. This instability is treated in some detail in the presence of longitudinal space-charge force and magnetic bunch-length compression. Special attention is given to the aforementioned collective effects such as for the production of high peak current bunches driving free-electron lasers.

Keywords
Electron bunches; linear accelerators; electron optics; synchrotron radiation; impedance; microbunching instability.

1 Coherent synchrotron radiation

The advent of sub-picosecond-long electron beams with very high brightness in ultra-violet (UV) and X-ray free-electron lasers (FELs) driven by radio-frequency linear accelerators (RF linacs) has raised the awareness of the accelerators community to the effect of coherent synchrotron radiation (CSR) emission on the beam energy distribution and transverse emittance (primarily in the bending plane). In fact, CSR is one of the limitations associated to magnetic bunch-length compression (see previous chapter) and to transport of short bunches in multibend lines.

CSR is the low-frequency component (typically up to the THz frequency range) of electromagnetic radiation emitted by ultra-relativistic particles in dipole magnets, see Fig. 1.

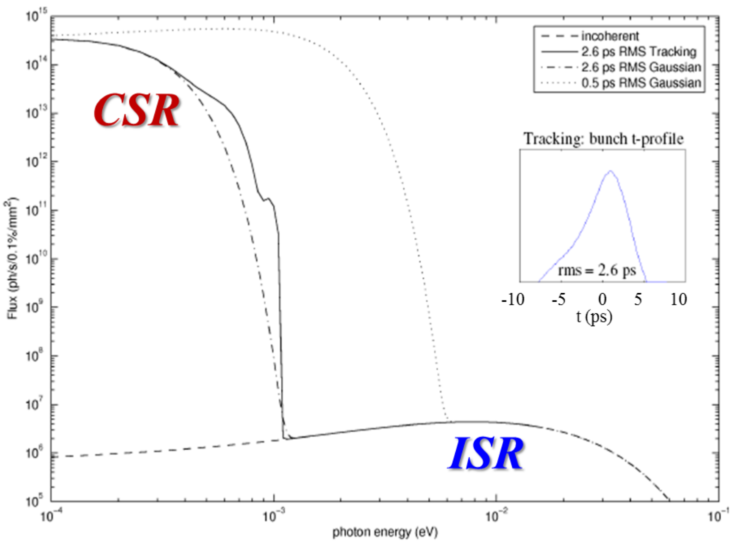


Fig. 1: Synchrotron radiation flux spectral distribution, emitted by the ~70 pC electron-bunch profile in the inset, through a dipole magnet, at the beam kinetic energy of 12.5 MeV. Courtesy of I.V. Bazarov.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.381

Particles are there subject to centripetal acceleration according to the Lorentz force. Radiation emitted by particles accelerated perpendicularly to their velocity is called synchrotron radiation (SR). The low-frequency part of the SR spectrum is amplified w.r.t. the high-frequency one, by the fact that electrons in the bunch are confined to a length scale of the order of, or shorter than, the radiation wavelength. As a consequence, the electric field of radiation produced by individual electrons adds in phase. This gives rise to a total radiation intensity that is proportional to the number of electrons in the bunch *squared*. In contrast, at shorter wavelengths the radiation field adds incoherently, i.e., the total intensity goes linearly with the number of beam particles. For a typical ~100 pC bunch charge, the CSR intensity is amplified by a factor ~5 × 10^8 w.r.t. the incoherent component.

The SR total power emitted in the fully incoherent and fully coherent regime in a dipole magnet of bending radius R is, respectively:

$$P_{\mathrm{FI}} = N\frac{e^2 c}{6\pi\varepsilon_0}\frac{\gamma^4}{R^2} \ , \qquad P_{\mathrm{FC}} = NP_{\mathrm{FI}} \ , \tag{1}$$

with N the number of particles in the bunch, γ the Lorentz factor for the beam mean energy, e the electron charge, c the light speed in vacuum and ε_0 the vacuum permittivity. The choice of beam energy and dipole radius typical of magnetic insertions in linac-driven FELs is a compromise between radiation effects and space-charge effects on the beam quality. As a result, SR is emitted in a regime of partial coherence in which the power of coherent emission is independent of beam energy (this happens for rms bunch lengths of the order of the characteristic SR wavelength $\sim R/\gamma^3$, and up to ~7 orders of magnitude longer wavelengths) [1]. In this case the total CSR power emitted by a Gaussian line-charge distribution moving along an arc is

$$P_{\mathrm{CSR}} \cong \frac{0.028c}{\varepsilon_0}\frac{Q^2}{R^{2/3}\sigma_z^{4/3}} \ . \tag{2}$$

Equation (2) shows that CSR power is highest for shortest bunches. For example, a 70 pC charge bunch with 20 μm rms bunch length emits $P_{\mathrm{CSR}}/(Nc) \cong 0.3\mathrm{MeV/m}$ energy per meter in a 2 m radius dipole. At the energy of 1 GeV, that corresponds to 0.03% loss of kinetic energy in a 1 m long magnet. This is comparable to the equivalent energy bandwidth of a UV or X-ray FEL [2]; thus, it is a significant amount of CSR power.

Owing to particles' curved path, radiation emitted by trailing particles in the bunch is allowed to catch up with leading particles within the dipole magnet (see Fig. 5 below). The primary impact of this tail–head CSR interaction [3] is a slice-by-slice change of longitudinal momentum. Such a change is in fact correlated with the z-coordinate internal to the bunch, and the z–E correlation length scale is of the order of the bunch length. Hence, particles in the same bunch slice, with a slice length much shorter than the bunch length, are subjected to the same longitudinal electric field. The change in particle energy in a dispersive region causes particles to deviate from their design trajectory. This accumulates into transverse offsets and angular divergence of bunch slices, by an amount of the order of $\Delta x \approx \eta_x\delta, \Delta x' \approx \eta'_x\,\delta$, where η_x, η'_x are the energy dispersion function and its longitudinal derivative, respectively, at the location of CSR emission, and δ is the CSR-induced relative energy deviation. This slice-by-slice character is illustrated in Fig. 2. Assuming typical values in magnetic compressors $\eta_x \approx 0.2\,\mathrm{m}, \eta'_x \approx 0.1\,\mathrm{rad}$ and $\delta \approx 0.03\%$, the CSR power corresponds to transverse errors of the order of $\Delta x \approx 60\ \mu\mathrm{m}, \Delta x' \approx 30\ \mu\mathrm{rad}$. These are comparable to typical unperturbed rms beam size and angular divergence. We see from such examples that it is essential to incorporate CSR in the design and optimization stage of a magnetic lattice.

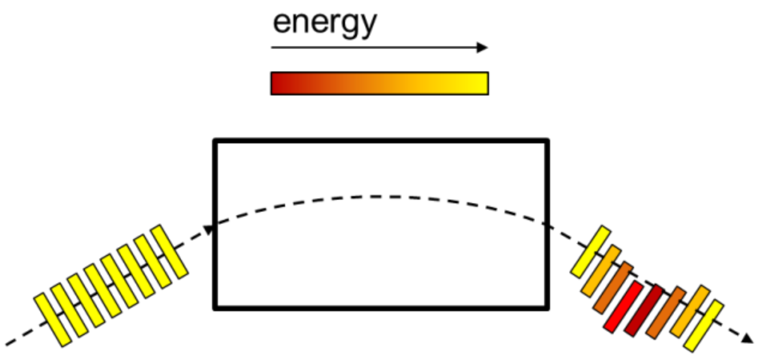


Fig. 2. Top-view representation of longitudinal bunch slices. Passing through a dipole magnet and interacting with CSR, the slices change their mean energy and energy spread. Energy change in a dispersive region causes different slices to end up at different transverse positions and angular divergences.

2 One-dimensional CSR model

The tail-to-head effect depicted in the previous section relies on the finite time that photons take to travel on a straight path from the source particle to the witness particle, in the same bunch. The CSR physics can therefore be described by taking into account the field retardation effect, i.e., the relation between the position r and time t at which a field is observed, and the retarded position r' and time t' at which this field is actually generated: $|r-r'| = c(t-t')$. The electric field E and the magnetic field B at position r and time t due to a point charge q in general motion with instantaneous velocity $\beta' = v'/c$ and acceleration $\beta' = v'/c$, derived from the Liénard–Wiechert retarded potentials, are [4]

$$\begin{aligned} \vec{E}(t) &= \frac{q}{4\pi\varepsilon_0}\left[\frac{\vec{n}-\vec{\beta}'}{\gamma^2|\vec{r}-\vec{r}'|^2\left(1-\vec{n}\vec{\beta}'\right)^3} + \frac{\vec{n}\times\left[\left(\vec{n}-\vec{\beta}'\right)\times\dot{\vec{\beta}}'\right]}{c|\vec{r}-\vec{r}'|\left(1-\vec{n}\vec{\beta}'\right)^3}\right], \\ \vec{B}(t) &= \frac{\vec{n}\times\vec{E}(t)}{c}, \end{aligned} \tag{3}$$

where $n = (r-r')/|r-r'|$. The first term of the electric field in Eq. (3) is commonly named 'velocity' or 'Coulomb' term; the second one, 'acceleration' or 'radiation' term. When the distance of an observer from the emission point (say, ≥ 0.1 m or so) is much larger than the bunch length (say, ≤ 1 mm or so), the physical system of any source particle–witness particle is well aligned with the circular trajectory. For this reason, the one-dimensional (1D) approximation to the description of the CSR field can be applied when the transverse offset of particles from the reference orbit can be neglected. An estimation of the regime in which the 1D approximation applies, also named the 'Derbenev criterion', is obtained by demanding that the retarded bending angle in the 1D model is a good approximation of the actual retarded angle, as shown in Fig. 3. This brings us to the constraint $\kappa \equiv \sigma_x / \left(\sigma_z^{2/3} R^{1/3}\right) << 1$ [3], where the symbols refer to rms transverse beam size in the bending plane, bunch length, and dipole radius. It is worth mentioning that in some cases [5, 6] the 1D model was found to reproduce well experimental evidence of transverse and longitudinal CSR effects, even if κ was approaching unity.

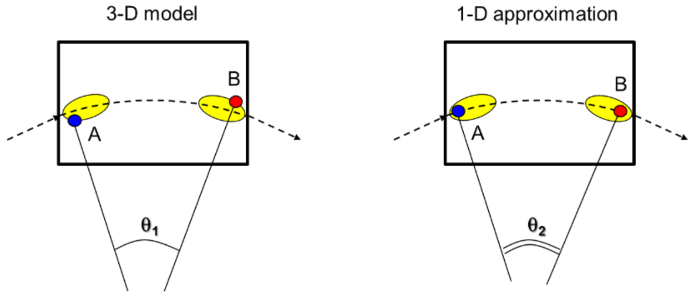


Fig. 3: CSR modelling in 1D approximation neglects the transverse offset of source and witness particles. The condition according to which $\theta_1 \cong \theta_2$ leads to the so-called Derbenev criterion (see context for details).

A rather common 1D CSR model implemented, e.g., in the `elegant` particle-tracking code [7, 8] approximates the electron bunch by a line-charge density $\lambda(z) = \frac{1}{Q}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho(x,y,z)\mathrm{d}x\mathrm{d}y$, where $\rho(x,y,z)$ is the volume-charge density. The calculation of the retarded electromagnetic field due to $\lambda(z)$ is done on a path given by an infinitely long drift, a circular arc with radius R and angle θ, and a drift with finite length. Since any change in $\lambda(z)$ during the CSR interaction is neglected, the system is said to be in the 'frozen-beam' approximation. Namely, it is assumed that the current profile during CSR emission is the same as during CSR interaction.

The point-charge field in Eq. (3) can be used to directly calculate the electric field due to a whole charge distribution $\lambda(z)$ by a Green's function method, i.e., by considering $\lambda(z)$ as a collection of infinitesimal charges $\lambda(z)\mathrm{d}z$ and integrating (in practice, some mathematical elaborations are done in order to speed up this calculation). The result for the component of the (longitudinal) electric field tangent to the beam trajectory at the interaction point, for an observation point on the arc, turns out to be [9]

$$E_{//} = \frac{e}{24^{1/3}\pi\varepsilon_0 R^{2/3}}\left\{\frac{\left[\lambda(z-s_\mathrm{L})-\lambda(z-4s_\mathrm{L})\right]}{s_\mathrm{L}^{1/3}} + \int_{z-s_\mathrm{L}}^{z}\frac{\mathrm{d}\lambda(z')}{\mathrm{d}z'}\frac{\mathrm{d}z'}{(z-z')^{1/3}}\right\}, \tag{4}$$

where we introduced the electron–photon path-length difference or 'slippage length' $s_\mathrm{L} = R\theta - 2R\sin(\theta/2) \cong \frac{R\theta}{2\gamma^2} + \frac{R\theta^3}{24}$ over a circular arc of length $R\theta$. For beam energies of interest here, the term $\sim 1/\gamma^2$ can usually be neglected.

The term in square bracket of Eq. (4) dominates at the entrance of the dipole magnet ('entrance transient' regime of CSR emission) and finds its origin in the Coulomb term of Eq. (3). The integral dominates as θ becomes large ('steady-state' regime of CSR emission), and is mostly due to the radiation term in Eq. (3). The entrance transient effect is opposite in sign to that of the bulk contribution. It describes the physical picture in which some head particles are inside the magnet, but they do not interact with CSR because tail particles still lie outside the magnet (and therefore are not radiating yet). In `elegant`, an infinitely long drift section is always assumed at the magnet entrance. This implies that, by using the full expression of Eq. (4), the global CSR effect might be underestimated if the drift section between two consecutive dipoles is shorter than the characteristic formation length $L_\mathrm{EN} \approx \gamma R^{1/3}\sigma_z^{2/3}$ [10]. Mostly depending on beam energy, L_EN can be as long as ~0.1 m to ~10 m.

When both leading and trailing particles are inside the magnet, CSR interaction can be described in the steady-state regime. Formally, we are allowed to do this if emission happens at wavelengths comparable to bunch length, and much longer than the characteristic wavelength of incoherent synchrotron radiation (ISR), $\lambda_c \approx \frac{4\pi}{3}\left(\frac{R}{\gamma^3}\right)$. Moreover, the bunch has to be short enough to allow photons emitted by trailing particles through the dipole to overcome leading particles. In short, we assume that $R/\gamma^3 << l_b \leq s_L$ [9], with l_b the characteristic bunch length. The last inequality also means that the magnet has to be longer than the characteristic distance from the magnet entrance at which transition to the steady state takes place, i.e., $l_m \geq \left(24R^2\sigma_z\right)^{1/3}$. Equivalently, the transient effect at the dipole entrance extends to approximately an angle $\theta_{tr} \approx \left(24\sigma_z/R\right)^{1/3}$, and it can be ignored w.r.t. the steady-state emission as long as the retarded bending angle $\gamma\theta >> 1$ [9]. Finally, propagation of CSR field through a drift section following the magnet enhances the global CSR effect, where the field is usually assumed to decay exponentially or so with the drift length [10].

In summary, Eq. (4) relies on the following approximations:

i) electrons are ultra-relativistic;

ii) transverse offsets of bunch particles from the reference orbit are neglected;

iii) the charge-density distribution rigidly moves along the curved path;

iv) the CSR field only influences particles ahead of emitters, i.e., radiation emitted backward is neglected;

v) the separation angle between the source particle and the witness particle is << 1;

vi) the CSR effect is assumed to be purely longitudinal, i.e., only the longitudinal component of the emitted electric field at any point tangent to the witness particle's orbit is responsible for change of its longitudinal momentum. Transverse electric field components, as well as magnetic field components, are all neglected.

As already mentioned, CSR physics relies on the CSR slippage length (it is documented that the classical expression for s_L was already known to Ipparco in ancient Greece around 100 B.C.), i.e., on the fact that photons travel straight, and doing so they catch up with electrons ahead of the source, which move on a curved path. Since angles involved are all assumed to be small, one may wonder if the angular divergence of CSR radiation is large enough to justify the tail-to-head interaction. Henceforth, we will discuss this point by limiting our attention to the steady-state regime.

It is known that most of SR is emitted in the laboratory frame in the forward direction, tangent to the source orbit, and with rms value of field intensity at angular divergence $\pm 1/(2\gamma)$, as shown in Fig. 4, left-hand plot. For example, in a case with $\gamma \geq 300$ and $\theta \leq 0.1$ rad, the characteristic angle of emission of ISR would not permit any tail-to-head interaction. Instead, CSR emitted at wavelengths $\lambda >> \lambda_c$ shows an intensity field distribution with angular divergence $\sigma_\varphi \approx 0.4\left(\lambda/R\right)^{1/3}$, and approximately independent of beam energy (see Fig. 4, right-hand plot) [11]. As an example, a 10 μm long bunch bent in a 0.3 m long dipole by an angle of 0.1 rad at the energy of 300 MeV would emit ISR at $\lambda_c \approx 60$ nm with rms opening angle $1/(2\gamma) \approx 0.6$ mrad. CSR would be emitted with a characteristic opening angle from 6 mrad up to several tens of mrad. Roughly speaking, the opening angle of CSR, evaluated at wavelengths equal to or longer than the bunch length, is large enough to be comparable to the dipole bending angle, and thereby to allow photons emitted by the bunch tail to catch up with the head, inside the dipole.

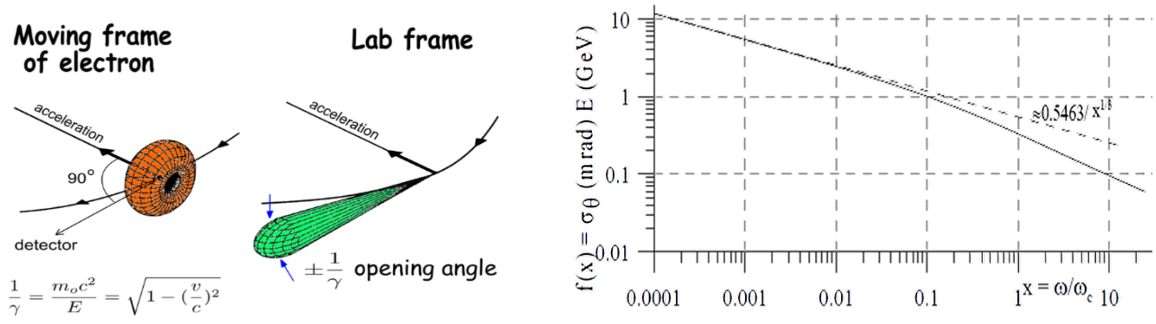


Fig. 4: Left-hand side: angular distribution of SR intensity emitted in a dipole magnet, in the moving frame of the electron and in the laboratory frame. Presented in Ref. [12]. Right-hand side: spectral dependence of SR opening angle in the bending plane. SR emitted forward at wavelengths $\lambda \le \lambda_c \approx 4\pi R/(3\gamma^3)$ shows an intensity distribution approximately Gaussian with rms width $\sigma_\varphi \cong \frac{0.6}{\gamma}\left(\frac{\lambda}{\lambda_c}\right)^{\frac{1}{3}}$. At wavelengths $\lambda >> \lambda_c$, $\sigma_\varphi \cong 0.4\left(\frac{\lambda}{R}\right)^{\frac{1}{3}}$. Published in Ref. [11].

A geometrical representation of the tail-to-head effect is given in Fig. 5, where the vectors for the electric field of Eq. (3) are shown. The source particle has a charge q and the witness particle is at point P. The electric field associated to the emitted radiation is orthogonal to the plane wave's direction of propagation in free space. We aim to estimate the electric field component $E_{//}$, parallel to the longitudinal velocity of the witness particle at P. We see that as long as $\theta << 1$, we can approximate the motion of q as if it had constant velocity, in both modulus and direction. In this case, the expression for the retarded field at P is as if it originated by a line-charge distribution at point D [4] (we are now considering the effect from a whole 1D bunch). Moreover, at large distances and high energies $\gamma >> 1$, only the 'radiation' field term in Eq. (3) contributes at P ('far-field' approximation). Hence, we can write $E(t;\theta << 1) \approx E_\perp = \lambda_z/(2\pi\varepsilon_0 d)$. By substituting the expression of d (see Fig. 5) and of the CSR 'overtaking length's (the slippage length for an arbitrary angle θ within the dipole) into $E_\perp$, and also noticing that $E_{//} \approx E_\perp\theta$, we find $E_{//} \approx E_\perp\theta = \frac{\lambda_z\theta}{2\pi\varepsilon_0 d} \approx \frac{1}{24^{1/3}\pi\varepsilon_0}\frac{\lambda_z}{R^{2/3}s^{1/3}}$, in agreement with [3]. For a uniform line-charge distribution $\lambda_z = Q/l_b$, and for a bunch sufficiently short so that CSR emitted by the bunch tail overtakes the bunch head before the bunch leaves the magnet, we find that the single-particle energy loss per unit length is $\frac{\mathrm{d}U_e}{\mathrm{d}z} \approx -\frac{1}{24^{1/3}\pi\varepsilon_0}\frac{eQ}{R^{2/3}l_b^{4/3}}$. This result can be extended to any line-charge distribution by introducing the steady-state CSR 'wake' or Green's function [3]:

$$\frac{\mathrm{d}U_e}{\mathrm{d}z} \cong -\frac{Ne^2}{24^{1/3}\pi\varepsilon_0 R^{2/3}}\int_{z-s_L}^{z}\frac{\mathrm{d}\lambda(z')}{\mathrm{d}z'}\frac{\mathrm{d}z'}{(z-z')^{1/3}}. \tag{5}$$

Equation (5) is just the integral term introduced abruptly in Eq. (4). We remark that the total bunch energy loss is N times the single-particle one and therefore, as expected, the coherent emission shows an N^2 dependence on the total field intensity. It is also important to stress out that the CSR-induced energy loss does *not* depend on beam energy. Equation (5) suggests that the energy loss is enhanced by fast variations of $\lambda(z)$ (current spikes, fast rises of the current profile and current modulation), so that $\mathrm{d}U_\mathrm{e}/\mathrm{d}z$ along the bunch can be monotonic or an oscillating function of z, for example, depending on the shape of the bunch current profile.

With the help of Eq. (5), we find for a Gaussian line-charge distribution:

$$
\begin{aligned}
U_\mathrm{e} &\cong -0.028 \times Z_0 c e^2 N \frac{\theta R^{1/3}}{\sigma_z^{4/3}}, \\
\langle \delta_\mathrm{CSR} \rangle &\cong -0.35 \times r_\mathrm{e} \frac{N}{\gamma} \frac{\theta R^{1/3}}{\sigma_z^{4/3}}, \\
\sigma_{\delta,\mathrm{CSR}} &\cong 0.7 \times |\langle \delta_\mathrm{CSR} \rangle|
\end{aligned}
\tag{6}
$$

for the single-particle energy loss, relative mean energy loss and rms relative energy spread, respectively. $Z_0 = 120\pi\ \Omega$ is the vacuum impedance and r_e the classical electron radius. For typical numbers in FEL linac drivers, $U_\mathrm{e} \approx 10^{-5} - 10^{-4}$ and $\langle \delta_\mathrm{CSR} \rangle \approx 10^{-4} - 10^{-3}$.

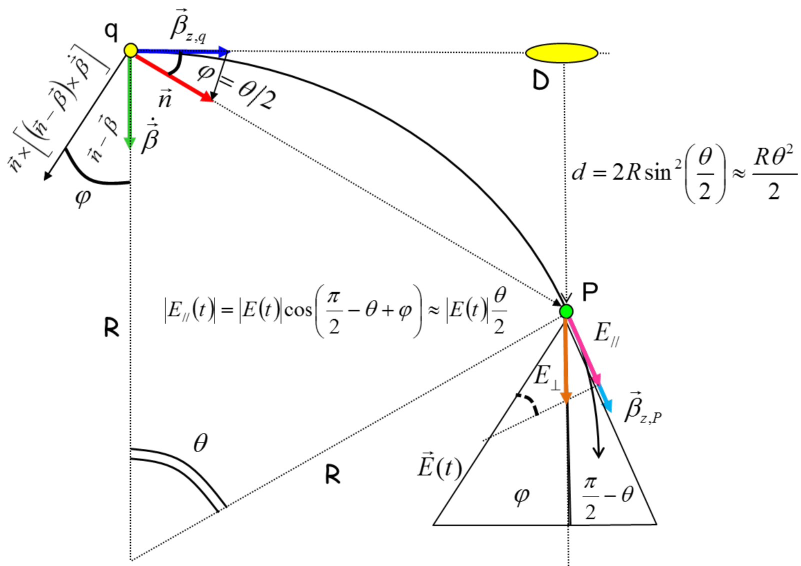


Fig. 5: Geometrical representation of CSR electric field in the 1D steady-state regime of emission in a dipole magnet. Symbols refer to Eq. (3). See context for details.

3 Transverse emittance growth

The growth of beam emittance in the bending plane is treated assuming the steady-state 1D CSR model discussed above. The transverse charge distribution in phase space (x, x') is characterized by its second-order momenta, which are associated to the so-called Twiss parameters: $\langle x_\beta^2 \rangle = \beta_x \varepsilon_x, \langle x'^2_\beta \rangle = \gamma_x \varepsilon_x, \langle x_\beta x'_\beta \rangle = -\alpha_x \varepsilon_x$; ε_x is the beam geometric emittance [13]. Particle motion is described as the linear superposition of betatron and dispersive coordinates [14]:

$$
\begin{aligned}
x(s) &= x_\beta(s) + R_{16}(s_0 \to s)\delta(s) \equiv x_\beta + \Delta x\ , \\
x'(s) &= x'_\beta(s) + R_{26}(s_0 \to s)\delta(s) \equiv x'_\beta + \Delta x'\ .
\end{aligned}
\tag{7}
$$

Here δ is the relative change in longitudinal momentum due to, e.g., absorption of CSR radiation in a dipole magnet. In the presence of a *single* energy perturbation ('kick') at location s, the perturbed beam emittance becomes

$$\begin{aligned}\varepsilon_x^2 &\equiv \left\langle x_\beta^2 \right\rangle \left\langle x'^2_\beta \right\rangle - \left\langle x_\beta x'_\beta \right\rangle^2 \\ &= \varepsilon_{x,0}^2 + \varepsilon_{x,0}\left(\beta_x \left\langle \Delta x'^2 \right\rangle + 2\alpha_x \left\langle \Delta x \Delta x' \right\rangle + \gamma_x \left\langle \Delta x^2 \right\rangle\right) + \left(\left\langle \Delta x^2 \right\rangle \left\langle \Delta x'^2 \right\rangle - \left\langle \Delta x \Delta x' \right\rangle^2\right) .\end{aligned} \tag{8}$$

If δ is due to *uncorrelated* events (e.g., ISR emission), the perturbed emittance results: $\varepsilon_x^2 \propto \left\langle \Delta x^2 \right\rangle_{\text{inc}} = \int_{s_0}^{s} \mathrm{d}s' R_{16}^2(s') \frac{\mathrm{d}\sigma_\delta^2(s')}{\mathrm{d}s'} \mathrm{d}s'$ (and an analogous integral of R_{26}). The integral is always positive, that is, the emittance growth is a non-reversible process. On the contrary, if δ is *correlated* with bunch length, the perturbed emittance is $\varepsilon_x^2 \propto \left\langle \Delta x^2 \right\rangle_{\text{CSR}} = \left[\int_{s_0}^{s} \mathrm{d}s' R_{16}(s') \frac{\mathrm{d}\sigma_\delta(z,s')}{\mathrm{d}s'} \mathrm{d}s'\right]^2$ (and analogously for R_{26}). In the latter case, a magnetic lattice can be designed in a way that the integral evaluated over the whole beam path is minimized or made null [15–17]. This situation applies to CSR emission, whose induced energy loss along a Gaussian bunch is shown in Fig. 6, left-hand plot.

We associate to each longitudinal bunch slice an average slice longitudinal momentum, and follow the slice centroid motion in (x, x') as predicted by Eq. (7). As a consequence of change of longitudinal momentum in a dispersive region, each slice will start moving on a different dispersive trajectory, and the beam *projected* emittance will increase according to Eq. (8) (see Fig. 6, centre plot). Moreover, since Δx and $\Delta x'$ are correlated along the bunch, we have that $\left\langle \Delta x^2 \right\rangle \left\langle \Delta x'^2 \right\rangle - \left\langle \Delta x \Delta x' \right\rangle^2 = 0$ and, given the single-kick approximation $\left\langle \Delta x^2 \right\rangle = \eta_x^2 \sigma_{\delta,\text{CSR}}^2, \left\langle \Delta x'^2 \right\rangle = \eta'^2_x \sigma_{\delta,\text{CSR}}^2$, with $\sigma_{\delta,\text{CSR}}$ the rms *relative* energy spread induced by CSR, we rewrite Eq. (8) as follows:

$$\varepsilon_x^2 = \varepsilon_{x,0}^2 + \varepsilon_{x,0}\left(\beta_x \left\langle \Delta x'^2 \right\rangle + 2\alpha_x \left\langle \Delta x \Delta x' \right\rangle + \gamma_x \left\langle \Delta x^2 \right\rangle\right) = \varepsilon_{x,0}^2 + \varepsilon_{x,0} H_x \sigma_{\delta,\text{CSR}}^2 . \tag{9}$$

Equation (9) introduces the optics function $H_x \equiv \left[\eta_x^2 + \left(\beta_x \eta_x' + \alpha_x \eta_x\right)^2\right] / \beta_x$, to be evaluated at the location of the CSR kick. It suggests that CSR-induced projected emittance growth can be minimized or even cancelled through a suitable design of H-function, for any given value of $\sigma_{\delta,\text{CSR}}$. It is worth noticing that, according to this picture, the slice emittance is not affected.

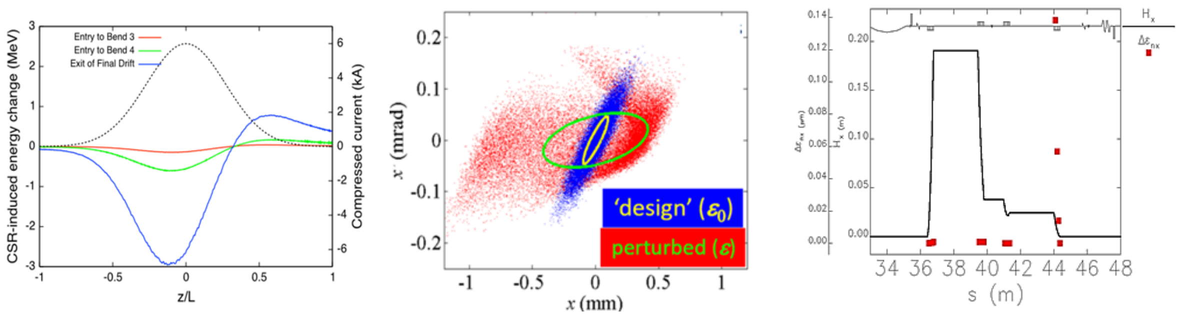


Fig. 6: (Left-hand side) CSR-induced mean slice energy difference for a Gaussian line-charge density, time compressed in a four-dipole symmetric chicane. The result is shown at the entry to third dipole (red), the entry to fourth dipole (green) and the exit of fourth dipole (blue). Published in Ref. [18]. Copyright of American Physical Society. (Middle) Charge distribution in the horizontal phase space of a beam matched to some design phase-space ellipse (blue dots), and for a beam mismatched due to CSR-induced emittance dilution (red dots). Ellipses represent second-order momenta. Courtesy of P. Emma. (Right-hand side) H-function and CSR-induced projected emittance growth along the first bunch compressor chicane of FERMI FEL. Published in Ref. [19]. Copyright of Elsevier. All plots are simulation results.

The single-kick picture in Eq. (9) usually applies to a four-dipole magnetic chicane adopted for a bunch-length compressor, and henceforth named a BC (see previous chapter). The CSR effect is strongest for a shortest bunch (see Eq. (6)) and, since the bunch approaches its final length already at the exit of the third magnet of a BC, this usually leads to a dominant CSR contribution from the fourth dipole, as shown in Fig. 6, left- and right-hand plots. Now, if the beam is forced to a horizontal waist in the second half of the chicane, and the dipoles' bending angle is small, it can be shown [19] that Eq. (9) reduces to $\varepsilon_x^2 \approx \varepsilon_{x,0}^2 + \varepsilon_{x,0}\left(\beta_x \theta^2 \sigma_{\delta,\mathrm{CSR}}^2\right)$. Thus, emittance growth can be minimized by small bending angle and small horizontal betatron function in proximity of the fourth dipole magnet [20].

The perturbed emittance at the exit of a multibend transport line, locally isochronous so that the beam has the same length at identical dipole magnets, can be expressed in the form $\varepsilon_x^2 = \varepsilon_{x,0}^2 + \varepsilon_{x,0} X\left(\alpha_x, \beta_x, \Delta\mu_x\right)_{s_\mathrm{f}} H_x\left(s_1\right) \sigma_{\delta,\mathrm{CSR}}^2$, with $H_x\left(s_1\right)$ at the location of the first CSR kick, and $X\left(\alpha_x, \beta_x, \Delta\mu_x\right)_{s_\mathrm{f}}$ is a function of Twiss parameters and relative betatron phase advance at the location of all other kicks. The energy kicks all being identical in modulus, the beam-line optics can be designed in a way that consecutive transverse kicks eventually cancel [21]. A most common and simple set-up is the one of identical Twiss parameters *and* dispersion functions at identical dipole magnets separated by π phase advance [16]. Other designs with non-symmetric optics parameters and phase advance different from π are in principle allowed. Similarly, emittance growth in a non-isochronous multibend line, such as an arc compressor, can be minimized with a proper optics design, where the 'optimal' Twiss functions now depend on the local compression factor [22, 23].

As an alternative to CSR-immune optics designs, Eq. (5) suggests that a line-charge distribution could be suitably produced at the entrance of a BC in order to generate a uniform energy loss by CSR along the bunch. In this case there would be no relative misalignment of bunch slices in phase space, and therefore no projected emittance growth [18].

4 Three-dimensional CSR model and shielding

A 3D CSR model is usually considered in numerical simulations whenever the Derbenev criterion largely fails. This may happen when a large relative energy spread or large dispersion function enlarges the beam size well above the betatron envelope, such as in the inner dipoles of a BC. Moreover, when the bunch is deflected, a longitudinal bunch slice does not stay perpendicular to the slice longitudinal momentum, but yaws with respect to it. In the former case, particles far from the beam axis, either belonging to the same slice or not, can sample a radial dependence of the CSR field. In the latter case, particles within the same slice can sample non-linearities of the longitudinal CSR field. In both cases, slice emittance growth may happen [1, 24]. Numerical predictions and experimental results suggest that these mechanisms may be playing a role already at relatively low compression factors, and affecting the slice normalized emittance at the level of ~0.1 μm rad [25, 26].

An additional complication to a CSR-dominated beam dynamics in a real lattice is provided by the shielding effect of the CSR field by the vacuum chamber [27]. In general, not all spectral components of CSR propagate in the chamber, and therefore the actual radiating energy is smaller than in a free-space environment. An analytical recipe for evaluating the shielding effect of an infinite parallel plate chamber of total height *h*, in a dipole magnet of curvature radius *R*, is [28]

$$\frac{\Delta E_\mathrm{shield}}{\Delta E_\mathrm{free}} \cong 4.2\left(\frac{n_\mathrm{th}}{n_\mathrm{c}}\right)^{\frac{5}{6}} \mathrm{e}^{-\frac{2n_\mathrm{th}}{n_\mathrm{c}}},\quad n_\mathrm{th} > n_\mathrm{c}\ . \tag{10}$$

Here $n_{\mathrm{th}} = \sqrt{\frac{2}{3}\left(\frac{\pi R}{h}\right)^3}$ is the threshold harmonic number for propagating radiation, and $n_{\mathrm{c}} = R/\sigma_z$ is the characteristic harmonic number for a Gaussian longitudinal density distribution whose standard deviation is σ_z. The meaning of n_c is that the spectral component of radiation with harmonic numbers beyond it is incoherent. A more immediate although ~~less~~ accurate recipe for evaluating the shielding effect says that CSR emission is suppressed if $\sigma_z \geq \sqrt{\frac{h^2 w}{\pi^2 R}}$, where w is the vacuum chamber width [29].

In practice, it is difficult to shield CSR completely because, for example, a beam pipe diameter of ≤ 2 mm would be needed for ultra-relativistic sub-picosecond-long bunches. At such small gaps, resistive-wall wake field could become intolerable [30].

5 Microbunching instability

Self-developing microstructures in the longitudinal phase space of electron bunches undergoing strong compression have been observed in several high-brightness FEL linac drivers. In accordance with experimental results, computer simulations of longitudinal space charge (LSC) force (longitudinal interparticle Coulomb interaction) [31] and CSR in bunch compressors show that a so-called microbunching instability (MBI) may significantly amplify small longitudinal density and energy modulations, and hence degrade the beam quality. In its simplest picture, MBI relies on the accumulation of energy modulation induced by initial non-uniformity of the charge distribution, such as due to electron beam intrinsic shot noise or non-uniformity of the photo-injector laser pulse, and on its successive transformation to an amplified density modulation through non-isochronous ($R_{56} \neq 0$) dispersive insertions, as shown in Fig. 7 [32]. Amplified density modulations will further drive energy modulations at higher amplitudes, and thereby a positive feedback for the instability is established. The process repeats at downstream stages of acceleration and compression and, in spite of relatively high energies, the beam may eventually show strong density and energy modulations. Especially disrupting for an FEL, the beam will show large slice energy spread, at scales equal to or longer than the (shortest) modulation wavelength. The model assumes that density modulations induce energy modulations at the same wavelength(s). Moreover, the effects of LSC and of R_{56} are fully decoupled: energy modulation only happens in drift (linac) sections, while density modulation only changes through BCs. Of course, as the bunch length is compressed, the initial modulation wavelength is compressed by the same factor. At this stage of description, CSR is not needed. In fact, it amplifies the LSC-induced MBI with an analogous mechanism established in the BC.

In order to estimate the spatial scale and beam energies at which LSC and therefore MBI becomes important, we consider a two-particle 1D beam model. The beam is assumed to be ultra-relativistic, with Lorentz mean energy factor γ. One particle is at bunch centre, and it represents the whole bunch charge Q. The other particle of charge q is at the very bunch head, and the two are separated by a distance l'_{b} in the co-moving frame of the bunch. The longitudinal electric field sampled by q is $E_z' = Q/\left(4\pi\varepsilon_0 l_{\mathrm{b}}'^2\right)$. In the laboratory frame, the two particles are separated by the Lorentz-contracted length $l_{\mathrm{b}} = l'_{\mathrm{b}}/\gamma$, and the electric field sampled by q is $E_z = E_z' = Q/\left(4\pi\varepsilon_0 \gamma^2 l_{\mathrm{b}}^2\right)$. Thus, the LSC field is stronger at lower energies and at shorter scale lengths. The model also suggests that (leading) particles gain energy, while others (trailing) lose it, namely, an energy modulation builds up. The work $\Delta U = qE_z L$ done by the LSC field over a length $L \sim 10$ m becomes comparable to typical beam energies spread values at beam mean energies < 10 MeV, and at bunch length scales of the order of 1 mm. At higher beam energies, the LSC effect can only be important at wavelengths much shorter than the bunch length, e.g., < 0.1 mm. In the following, we present a quantitative analytical treatment of MBI as induced by LSC, and compare it with other collective effects common to beam dynamics in FEL linac drivers.

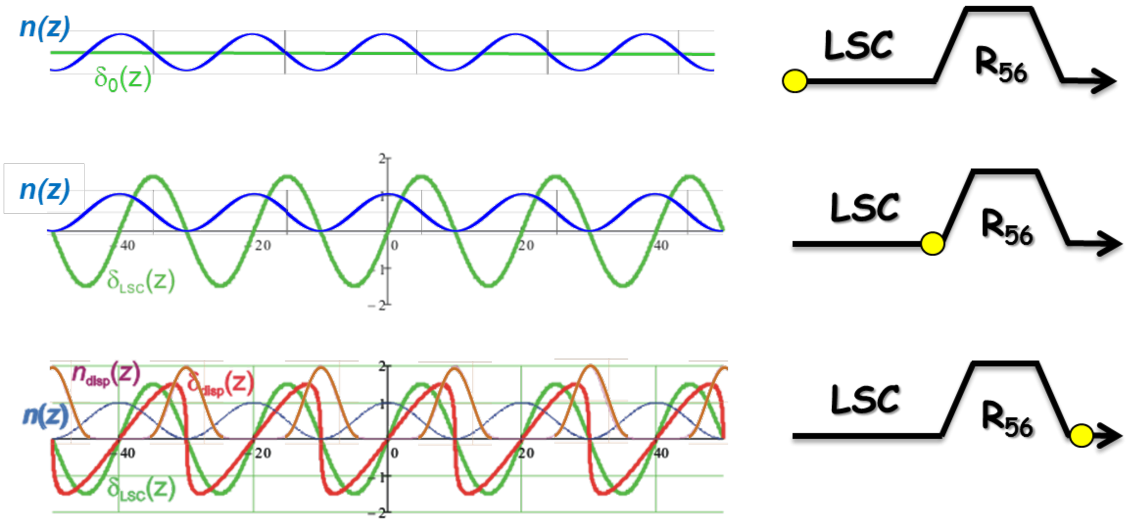


Fig. 7: Evolution (left-hand diagrams) of density (n) and energy modulation (δ) along the bunch (z), evaluated at different locations of an RF linac plus BC beam line (sketched on the right-hand side; beam is at the yellow dot). MBI is assumed to be only driven by LSC in the linac section, and by R_{56} in the BC. Upper: beam at the linac entrance has a small density modulation, and is uniform in energy. Middle: passing through the linac, the initial density modulation induces an energy modulation at the same wavelength via LSC field. Bottom: beam has passed through the BC. The longitudinal phase space is sheared due to the non-zero R_{56} in the BC ($\delta_{\rm disp}$, red curve), and the density modulation is amplified ($n_{\rm disp}$, orange curve). For a direct comparison, the initial density and energy modulations are superimposed.

6 Impedances

Accumulation of energy modulation in straight sections is driven by LSC and, at longer wavelengths, by linac geometric wake fields [30]. Since MBI is commonly studied in the frequency domain, an energy modulation at wavenumber $k = 2\pi/\lambda$ can be evaluated as the integral of the LSC impedance $Z_{\rm LSC}(k)$ over a drift length L, times the Fourier transform of the bunch peak current I [33]:

$$b(k) = \frac{1}{Nec}\int I(z)\mathrm{e}^{-\mathrm{i}kz}\mathrm{d}z\ , \qquad \Delta\gamma(k) = -\frac{4\pi I}{I_{\rm A}} b(k)\int_0^L \frac{Z(k,s)}{Z_0}\mathrm{d}s\ . \tag{11}$$

$b(k)$ is called the 'bunching factor', and its value is proportional to the density modulation amplitude *relative* to the average bunch current. For a line-charge distribution dominated by shot noise, the bunching in a bandwidth $\Delta\lambda$ can be written $b(\lambda) = \sqrt{\frac{2ec}{I\Delta\lambda}}$. The free-space LSC impedance per unit length, averaged over a Gaussian transverse distribution of rms sizes $\sigma_{x,y}$, is [34]

$$Z_{\rm LSC}(k) = \frac{\mathrm{i}Z_0}{\pi k r_{\rm b}^2}\left[1 - 2I_1(\xi)K_1(\xi)\right]\ , \tag{12}$$

with $\xi = kr_{\rm b}/\gamma$, I_1 and K_1 being modified Bessel functions of the first kind and $r_{\rm b} = 0.8735(\sigma_x + \sigma_y)$ is the equivalent radius of a transverse uniform distribution with area $\pi r_{\rm b}^2$ [35]. $Z_{\rm LSC}$ has a maximum for

$\xi \cong 1$. It tends to underestimate the effect of 3D fluctuations of the electric field, which happens at $\xi \geq 0.5$ [35], i.e., when the Lorentz-contracted wavelength of modulation becomes comparable to, or shorter, than the transverse beam size ('pancake-beam' scenario). In the strict 1D limit ('pencil-beam' scenario), $2I_1(\xi) \to \xi$ and, for $\xi \ll 1$, $I_1(\xi)K_1(\xi) \to \ln(\xi)$ [36]. Analytical approximations for Z_{LSC} in the presence of boundary conditions, for example as given by a perfectly conducting round vacuum chamber of given radius r_p, are available in the literature [37]. They rarely apply to cases of interest here because they become important at relatively long wavelengths, $\xi \leq r_\mathrm{b}/r_\mathrm{p}$.

In a way analogous to Eq. (12), we introduce an impedance per unit length for the CSR longitudinal electric field in a dipole of curvature radius R [38, 39], and the longitudinal geometric impedance of an RF structure of inner iris radius a [40]:

$$Z_{\mathrm{CSR}}(k) = \frac{Z_0 k^{1/3}}{\pi R^{2/3}}(0.41 + \mathrm{i}0.23) , \qquad Z_{\mathrm{RF}}(k) \approx \frac{\mathrm{i}Z_0}{\pi k a^2} . \tag{13}$$

The imaginary nature of an impedance implies a redistribution of the particles' longitudinal momentum inside the bunch, with no net energy loss, as contrary to a real impedance. In Eq. (13), the RF geometric impedance is associated to a 'wake function' in a periodic structure (this is proportional to a longitudinal electric field in the time domain) $w_{\mathrm{RF}}(s) = \frac{Z_0 c}{\pi a^2}\exp\left(-\sqrt{s/s_0}\right)$, with $s_0 \approx 1$ mm [30], and we considered the impedance behaviour at high frequencies. Figure 8 shows the spectral domain of each of the aforementioned impedances. The typical band of interest for the MBI is at initial wavelengths in the range 1–100 µm, which, after a compression factor in the range 10–100, fit characteristic scale lengths of UV and X-ray FEL dynamics, such as FEL co-operation length and undulator slippage length [2].

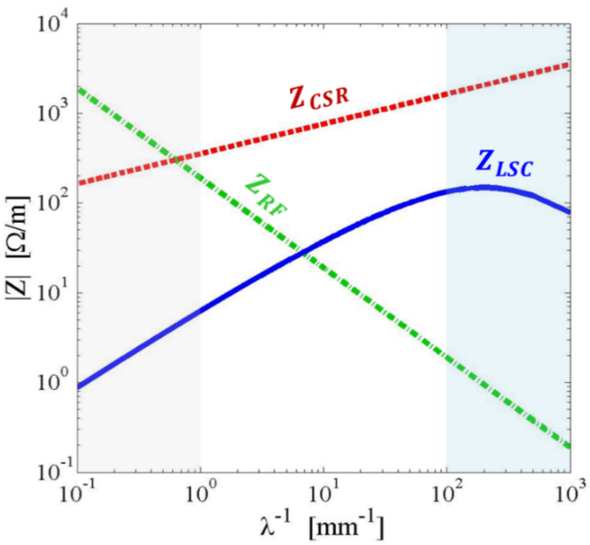


Fig. 8: Spectral behaviour of the modulus of RF, CSR and LSC impedance per unit length, see Eqs. (12) and (13). Z_{RF} is for an iris radius a = 10 mm. Z_{CSR} is calculated for R = 5 m. Z_{LSC} is calculated for a beam energy of 300 MeV and r_b = 400 µm. The shadow area on the left-hand side covers characteristic lengths of uncompressed electron bunches. The shadow area on the right-hand side highlights the typical band of interest for the MBI (wavelengths before compression).

7 Spectral gain

Modulation wavelengths of interest for FELs are usually much shorter than the electron-bunch length ('coasting beam' approximation). Moreover, it has physical sense to require that density modulation amplitudes be much smaller than the average current ('linear regime' of the instability). When both these assumptions hold, the amplitude of the density modulation at each wavelength grows *independently*, and the strength of MBI is quantified by a spectral gain $G(k_\mathrm{i}) \approx |b_\mathrm{f}(k_\mathrm{f})/b_\mathrm{i}(k_\mathrm{i})| \geq 0$ [32].

We now want to estimate the MBI gain in a relatively simple system, e.g. a straight section followed by a BC. Let us assume that the instability starts from a monochromatic sinusoidal modulation of the beam-current profile, $I(z_\mathrm{i}) = I_0\left[1 + A\sin(k_\mathrm{i}z_\mathrm{i})\right]$, with A the relative modulation amplitude (the same logic and formalism applies to the case of initial energy modulation). According to Eq. (11), an energy modulation $\Delta\gamma(z_\mathrm{i},k_\mathrm{i}) = 4\pi A\frac{|Z_\mathrm{LSC}(k_\mathrm{i})|}{Z_0}\frac{I_0}{I_\mathrm{A}}L\sin(kz_\mathrm{i}) \equiv \Delta\gamma_0\sin(kz_\mathrm{i})$ is accumulated over a distance L. As the beam travels with a correlated energy spread δ_c or, equivalently, a linear energy chirp h_i through the BC, the generic particle's longitudinal coordinate follows $z_\mathrm{f} = z_\mathrm{i} + R_{56}\delta_\mathrm{c} + R_{56}\,\Delta\gamma(z_\mathrm{i},k_\mathrm{i})/\gamma$ (see previous chapter).

We will make use of the differential form $\frac{\mathrm{d}z_\mathrm{f}}{\mathrm{d}z_\mathrm{i}} = 1 + h_\mathrm{i}R_{56} + R_{56}\frac{\Delta\gamma_0}{\gamma}\cos(k_\mathrm{i}z_\mathrm{i})$ to find the final line-charge density: $\rho_\mathrm{f} \equiv \frac{\mathrm{d}N}{\mathrm{d}z_\mathrm{f}} = \frac{\mathrm{d}N}{\mathrm{d}z_\mathrm{i}}\frac{\mathrm{d}z_\mathrm{i}}{\mathrm{d}z_\mathrm{f}} \cong C\rho_\mathrm{i}\left[1 - Ck_\mathrm{i}R_{56}\frac{\Delta\gamma_0}{\gamma}\cos(k_\mathrm{i}z_\mathrm{i})\right]$. This expression is valid at first order in $\Delta\gamma$ and having introduced the linear compression factor $C = (1 + h_\mathrm{i}R_{56})^{-1}$. We are now able to write down the expression for the MBI gain as a function of the initial modulation wavenumber:

$$G(k_i) \cong \left|\frac{\Delta\hat{\rho}_f/\hat{\rho}_f}{\Delta\hat{\rho}_i/\hat{\rho}_i}\right| = 4\pi Ck_i|R_{56}|\frac{I_0}{I_A}\frac{|Z_{LSC}(k_i)|}{Z_0}\frac{L}{\gamma}. \tag{14}$$

Owing to the fact that any real beam has a non-zero *uncorrelated relative* energy spread $\sigma_{\delta,\mathrm{u}}$, particles belonging to the same bunch slice will travel through the BC along different path lengths because of their energy difference. The MBI gain is partially suppressed by those additional particles' longitudinal slippage. The particles' motion being uncorrelated, this process is called energy or longitudinal Landau damping, and it tends to exponentially smear energy and density modulations at relatively short wavelengths. With $\sigma_{\delta,\mathrm{u}}$ defined at the *entrance* of the BC, the MBI gain after a linac plus BC section becomes [36]:

$$G(k_i) \cong \frac{4\pi}{Z_0}\frac{I_0}{I_A}Ck_i|R_{56}|\left|\int_0^L\frac{Z_{LSC}(k_i;s)}{\gamma(s)}ds\right|\exp\left[-\frac{1}{2}\left(Ck_iR_{56}\sigma_{\delta,u}\right)^2\right]. \tag{15}$$

One can see that for $Z_\mathrm{LSC}(k)$ approximately constant in amplitude over a wide range of k, the gain is peaked at the wavelength that satisfies $Ck_\mathrm{i}R_{56}\sigma_{\delta,\mathrm{u}} = 1$. $Z_\mathrm{LSC}(k)$ being a broadband impedance (see Fig. 8), this result is approximately true in general. Owing to the fact that the gain is exponentially suppressed by $\sigma_{\delta,\mathrm{u}}$ at short wavelengths, we find a natural cut-off of the MBI gain at $Ck_\mathrm{i}R_{56}\sigma_{\delta,\mathrm{u}} \geq 1$. If the dispersive line is isochronous, and MBI is only driven by LSC in the linac, we do not expect any gain, because particles' position inside the bunch would be the same at the entrance and at the exit of the BC.

It is worth pointing out that the gain in Eq. (15) is independent of initial modulation amplitudes. In other words, the final bunching factor is assumed to be dominated by the energy-to-density transformation that happens in the BC, and the contribution of the very initial bunching to the final bunching is neglected. This is the so-called 'high-gain' approximation. A representative LSC-induced spectral gain is shown in Fig. 9, left-hand plot, evaluated under all the approximations discussed so far.

If the MBI gain is relatively small, it may have physical sense to include the initial bunching into its expression [41]. Instead, if the linear regime of development of MBI is not satisfied (in our derivation of Eq. (14), that regime essentially corresponds to a Taylor expansion of ρ_{f} at first order in $\Delta\gamma$), an analytical prediction of the final gain could become a difficult task. In fact, a non-linear theory of MBI is still an open field of research [42]. For example, the linear regime may or may not apply to the case of weak compression, multistage compression schemes or to multibend lines. In a two-stage compression, the approximation of linear gain still holds through the second BC only if the bunching factor at the entrance of it is small, $\left|b_{\mathrm{f},1}\right| << 1$, as well as the induced energy modulation, $\left|C_1 C_2 k_{\mathrm{i}} R_{56,2} \Delta\gamma_2 / \gamma_{\mathrm{f}}\right| << 1$. The total gain of the beam line can in that case be estimated as the product of individual gains at the two BCs: $G_{\mathrm{tot}} \approx G_1 \times G_2$. Depending on beam and machine parameters, those conditions may not happen, and the total gain can largely exceed the product of individual gains. In general, the presence of numerous dipoles, such as in arcs or multibend transport lines, leads to multiple stages of amplification of the MBI, whose gain evaluation requires numerical methods.

An expression for the amplification of initial energy modulation amplitudes equivalent to Eq. (15) can be obtained as well. Here, we prefer to point out that, once the final energy modulation amplitude is computed (see Eq. (11)), an upper limit for the growth of beam *uncorrelated* energy spread due to MBI is given by the assumption that the whole modulation translates into energy spread. Such a limit is obtained by integrating the final energy modulation amplitude over all frequency components [43]:

$$\sigma_{\mathrm{E,MBI}}^2 \approx \frac{\left(m_{\mathrm{e}} c^2\right)^2}{2\pi n_z} \int_{-\infty}^{+\infty} \mathrm{d}k_{\mathrm{i}} \left| G\left(k_{\mathrm{i}}\right)^2 \Delta\gamma_{\mathrm{f}}\left(k_{\mathrm{i}}\right)^2 \right| , \tag{16}$$

with n_z the line-charge density in m^{-1}.

The CSR contribution to MBI in compressors of FEL linac drivers is typically weaker than the one from LSC, because the effect of the latter one is summed over much longer distances. As the lattice starts being dominated by dipole magnets, however, such as in recirculating accelerators, the contribution to the gain from Z_{CSR} may become comparable to, if not greater than, the one from Z_{LSC}. In general, the CSR effect reinforces, if not drives, MBI [44]. This is because the CSR field modulates particles in energy inside dipole magnets, and the dispersion function translates energy modulations into density ones. A positive feedback for the instability, as already depicted for the LSC field, is therefore established inside the dispersive insertion [8].

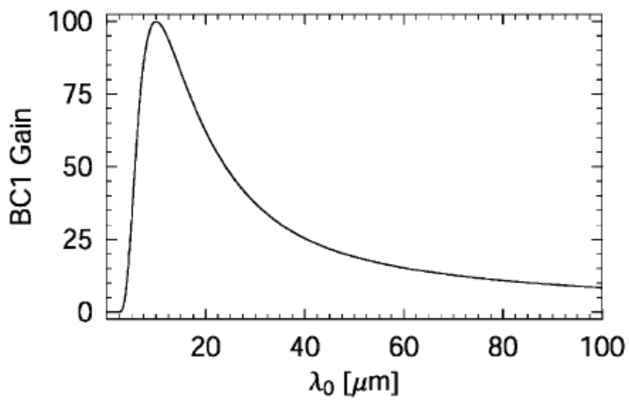
At the same time, since the CSR instability couples transverse and longitudinal motion of particles, we should also consider that particles belonging to the same bunch slice but moving with different betatron amplitudes, as in a beam with non-zero horizontal geometrical emittance ε_0, will follow different path lengths. Thus, beam emittance causes some damping of the CSR-induced MBI in a way analogous to energy-Landau damping induced by $\sigma_{\delta,\mathrm{u}}$. It can be shown [19] that in order for the rms path-length difference to generate smearing of MBI at a wavenumber k, beam emittance and optics functions at the entrance of the dipole of interest must satisfy

$$k\sqrt{\varepsilon_0 H} \geq 1\ . \tag{17}$$

The H-function can be related to R_{51} and R_{52} transport matrix terms of an arbitrary beam line. If we do so for a BC, and if some relevant CSR gain is generated in the third and fourth dipoles of it (we could expect that because the beam is shorter, and therefore the CSR field stronger, in the second half of the BC), transverse Landau damping has to be calculated by means of matrix terms that propagate particles' coordinates from the third dipole to the exit of the BC. We would therefore obtain an expression for the gain similar to that in Eq. (15), but now also dependent on Z_{CSR}, and including an additional transverse damping term describing suppression of the final bunching [45]:

$$\frac{b(k_{\mathrm{i}},s)}{b(k_{\mathrm{i}},0)} \propto \exp\left[-\frac{1}{2}\left(Ck_{\mathrm{i}}R_{56}\sigma_{\delta,\mathrm{u}}\right)^2\right]\exp\left\{-\frac{1}{2}C^2k_{\mathrm{i}}^2\left[\varepsilon_0\beta_0\left(R_{51}-\frac{\alpha_0}{\beta_0}R_{52}\right)^2-\frac{\varepsilon_0}{\beta_0}R_{52}^2\right]\right\}\ . \tag{18}$$

In Eq. (18), beam motion is assumed to start at coordinate s = 0, and the matrix terms have to be calculated at position s > 0; α_0, β_0 are Twiss parameters at s = 0. A representative 1D CSR-induced spectral gain is shown in Fig. 9 – left and right, including the effect of energy- and emittance-Landau damping. We stress that since R_{51} = R_{52} = 0 through a *whole* four-dipole chicane, there is no net transverse damping of MBI induced by an upstream Z_{LSC}, unless we consider a collective effect (Z_{CSR}) taking place *inside* the chicane.



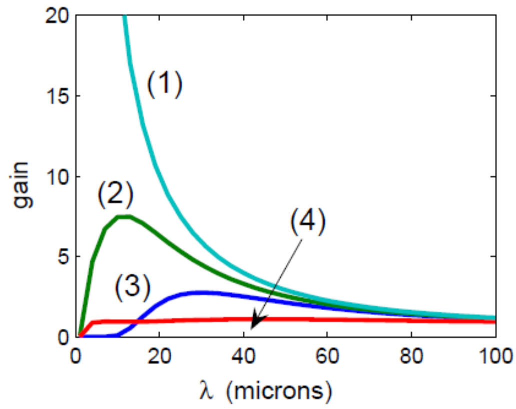


Fig. 9: (Left-hand side) MBI spectral gain induced by LSC after an RF linac section and a BC in the LCLS beam line, as a function of the initial modulation wavelength. Published in Ref. [36]. Copyright of American Physical Society. (Right-hand side) MBI spectral gain induced by CSR from a 6 kA peak current bunch in a BC at 5 GeV and R_{56} = –25 mm. Beam initial uncorrelated energy spread and normalized horizontal emittance are: (1) $\varepsilon_{n,x} = 1\ \mathrm{nm\ rad}$, $\sigma_{\delta,\mathrm{u}} = 2\times10^{-6}$; (2) $\varepsilon_{n,x} = 1\ \mu\mathrm{m\ rad}$, $\sigma_{\delta,\mathrm{u}} = 2\times10^{-6}$; (3) $\varepsilon_{n,x} = 1\ \mu\mathrm{m\ rad}$, $\sigma_{\delta,\mathrm{u}} = 2\times10^{-5}$; (4) $\varepsilon_{n,x} = 20\ \mu\mathrm{m\ rad}$, $\sigma_{\delta,\mathrm{u}} = 2\times10^{-6}$. Published in Ref. [46].

8 Laser heater

Equation (18) illustrates the two mechanisms discussed so far for suppression of MBI gain, i.e., energy- and emittance-Landau damping. Owing to the fact that, in FEL linac drivers, MBI is usually dominated by LSC, the main setting for gain suppression is to increase the beam initial uncorrelated relative energy spread. Typical energy spread values for high-brightness beams generated in RF photo-injectors are $\sigma_{\mathrm{E,u}}$ ≈ 1–3 keV [47], and the associated peak gain in long FEL drivers may be as high as 10^2–10^4. It was demonstrated in existing facilities that by increasing $\sigma_{\mathrm{E,u}}$ to 10–40 keV rms level at the beam energy of ~100 MeV, the total peak gain can be lowered by a factor of up to 10^2, and the FEL intensity is increased by a factor of up to 3 [48, 49]. Doing so, however, the final beam uncorrelated energy spread is increased too, by approximately $\sigma_{\delta,\mathrm{f}} \approx C\sqrt{\sigma_{\delta,0}^2 + \sigma_{\delta,\mathrm{LH}}^2}$, where C is the total compression factor, $\sigma_{\delta,0}$ the natural relative energy spread at the injector exit and $\sigma_{\delta,\mathrm{LH}}$ the relative energy spread added to the beam. For an FEL, we shall still require $\sigma_{\delta,\mathrm{f}} \leq 0.5\rho$ [2].

A 'laser heater' (LH) was proposed [50], and is now in operation or in design stages at several FEL facilities, in order to generate $\sigma_{\delta,\mathrm{LH}}$ in a controlled manner. Basically, a resonant interaction of electron bunches with an external infrared (IR) laser in a short undulator induces rapid energy modulation at the laser frequency. If the undulator is installed in the middle of a small chicane, the chicane geometry and the beam-line optics can be designed in a way that the IR modulation eventually washes out, and transforms into purely uncorrelated energy spread for Landau-damping purposes. Typically, the electron beam is forced to a waist in the undulator, so that the beam angular divergence is large. Then the transport through the *second* half of the chicane allows smearing of the IR modulation if $2\pi R_{52}\sigma_{x'} \geq \lambda_{\mathrm{IR}}$. Due to beam interaction with electromagnetic radiation in a dispersive section, beam emittance is also affected and its growth can be estimated at the level of $\frac{\Delta\varepsilon_x}{\varepsilon_x} \approx \frac{1}{2}\frac{\left(\eta_x\sigma_{\delta,\mathrm{LH}}\right)^2}{\beta_x\varepsilon_x}$, with optical parameters evaluated at the undulator location.

For a plane-polarized undulator with strength parameter K (see previous chapter) and length L_u, the relative energy spread induced by a LH is approximately [48]

$$\sigma_{\delta,\mathrm{LH}} \cong \frac{L_\mathrm{u}}{\sqrt{\left(\sigma_x^2 + \sigma_y^2\right)}}\sqrt{\frac{P_\mathrm{L}}{2P_0}}\frac{K[JJ]}{\gamma_0^2}\ , \tag{19}$$

where σ_x, σ_y are rms electron beam sizes at the undulator, $P_0 = m_\mathrm{e}c^3/r_\mathrm{e} \cong 8.7\,\mathrm{GW}$, $L_\mathrm{u} = N_\mathrm{u}\lambda_\mathrm{u}$ is the undulator length, given by the number of undulator periods times the period length, $[JJ] = J_0(\varsigma) - J_1(\varsigma)$ are Bessel functions of argument $\varsigma = K^2/\left(4+2K^2\right)$ and P_L is the required laser peak power. Equation (19) is valid in the limit that the laser Rayleigh length is much longer than the undulator: $\pi w_0^2/\lambda_{\mathrm{IR}} >> N_\mathrm{u}\,\lambda_\mathrm{u}$, with w_0 twice the laser rms transverse size at the undulator. This ensures that the laser cross-section does not vary substantially along the undulator due to radiation diffraction.

Figure 10 shows the slice energy spread of a beam compressed in a two-BC scheme in the FERMI linac, measured at the beam energy of 1.2 GeV, as a function of the energy spread induced by a LH at 0.1 GeV (approximately 150 m upstream) [49]. For large heating levels, MBI is suppressed and the final energy spread is linearly proportional to the initial value. The constant of proportionality reflects the total compression factor $C = 7$, by virtue of the preservation of the longitudinal emittance (see previous chapter). For small heating levels, MBI 'wins' and the final energy spread is dominated by the instability gain. In this case, the final energy spread has a non-linear dependence on the initial value. In between the two regions, an optimum for the FERMI FEL is found that allows us to mitigate MBI while achieving a minimum slice energy spread at the linac end. The measurement was found to be in agreement with numerical predictions [51].

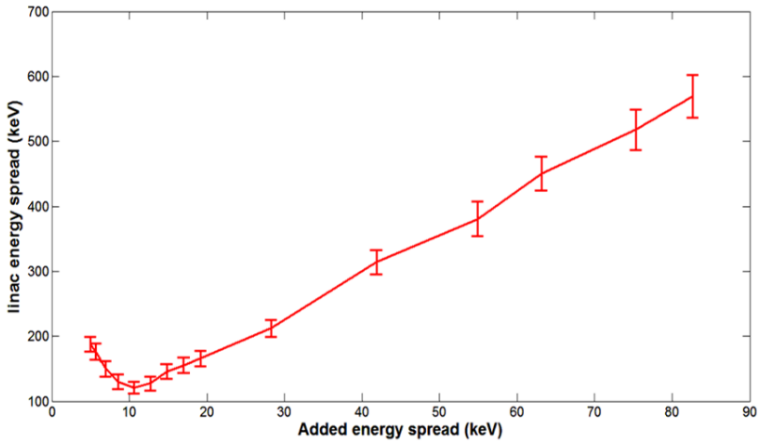


Fig. 10: Slice energy spread of a 500 pC charge bunch, compressed by a total factor of 7 in a two-BC scheme at the FERMI FEL, and measured at the energy of 1.2 GeV, as a function of the energy spread added by the LH at the energy of 0.1 GeV. Published in Ref. [49]. Copyright of American Physical Society.

9 Recent trends

Analytical modelling of MBI may become challenging when complicated beam lines are considered, and when some of the approximations depicted so far do not apply any more. Several codes are available for a numerical computation of the instability gain and of its effect on the beam energy spread, each code having its own weak and strong points. One of the major issues related to particle-in-cell codes is the one of avoiding numerical sampling noise generated by tracking (macro)particles, which are often treated in populations a few orders of magnitude smaller than the real number of beam particles. For MBI driven by shot noise, the final bunching $b_{\mathrm{f}} \propto G_{\mathrm{tot}} b_{\mathrm{i,sn}} \propto G_{\mathrm{tot}} \sqrt{\lambda/N}$ (see Section 6), with N the real number of electrons per modulation wavelength λ. By tracking a number of particles $N_{\mathrm{tr}} < N$, we tend to overestimate the total gain by a factor $\sqrt{N/N_{\mathrm{tr}}}$ [32]. Attempts to overcome this issue include filtering techniques, preparation of 'quiet' initial distributions, tracking the real number of particles, and development of algorithms intrinsically noise-free.

Present paths of research in physics of particle accelerators include practical ways to suppress the MBI, either alternative or complementary to a LH. For example, machine designs that minimize the MBI gain with a suitable compression scheme, and techniques to enhance either energy or transverse Landau damping without affecting the final beam quality. A review of such works can be found in Ref. [19]. Recently, some focus has been given to multibend lattice designs able to simultaneously minimize CSR-induced projected emittance growth and CSR-induced MBI [52].

Theoretical, numerical and experimental studies have pointed out the capability of a LH of improving the FEL intensity also as a function of transverse shaping of the LH photon pulse and, indirectly, of the electron beam energy distribution that results from the LH interaction [53]. The adoption of a LH is also opening the door to an optical shaping of the FEL pulse duration [54] and of its spectral content [55]. MBI *itself* has been used, on purpose and in a controlled manner, to enlarge the range of spectral features of a coherent light source [56, 57].

Acknowledgement

The author's present understanding of CSR effects and MBI is largely due to interactions with many colleagues, who are all acknowledged. M. Venturini and S. Spampinati are especially acknowledged for discussions on the analytical modelling of MBI, P. Smorenburg and I. Setija for elaborations on numerical modelling and support to Figs. 2, 3 and 7 of this chapter, and S. Milton for encouragements to facing such a fascinating topic.

References

[1] M. Dohlus, Proc. Int. Committee for Future Accelerator Beam Dynamics Newsletter No. 38, Eds. I.S. Ko and W. Chou (2005).

[2] P. Schmuser, M. Dohlus and J. Rossbach, in *Ultraviolet and Soft X-ray Free-Electron Lasers*, Eds. A.-F. Chiba *et al.* (Springer, Berlin, 2008).

[3] Ya.S. Derbenev, J. Rossbach, E.L. Saldin and V.D. Shiltsev, TESLA-FEL 95-05, DESY, Hamburg, Germany (1995).

[4] Hofmann, Characteristics of synchrotron radiation, CAS Course on Synchrotron Radiation and Free-electron Lasers, Chester, UK (1989), pp. 115–141. DOI 10.5170/CERN-1990-003.115.

[5] K.L.F. Bane *et al.*, *Phys. Rev. ST Accel. Beams* **12** (2009) 030704. https://doi.org/10.1103/PhysRevSTAB.12.030704

[6] C.C. Hall *et al.*, *Phys. Rev. ST Accel. Beams* **18** (2015) 030706. https://doi.org/10.1103/PhysRevSTAB.18.030706

[7] M. Borland, APS Technical Note LS-207 (2000).

[8] M. Borland, *Phys. Rev. ST Accel. Beams* **4** (2001) 070701. https://doi.org/10.1103/PhysRevSTAB.4.070701

[9] E. Saldin *et al.*, *Nucl. Instrum. Phys. Res. A* **398** (1997) 373. https://doi.org/10.1016/S0168-9002(97)00822-X

[10] P. Emma and G. Stupakov, Proc. 8th European Accelerator Conf., WEPRI029, 2002, p. 1479.

[11] H. Wiedemann, School on Synchrotron Radiation at ICTP, Trieste, Italy (2002).

[12] L. Rivkin, CAS Course on Synchrotron Radiation Basics, Granada, Spain (2012).

[13] S.Y. Lee, *Accelerator Physics*, 2nd ed. (World Scientific, 2004), pp. 47–52 and 61–67. https://doi.org/10.1142/5761

[14] K.L. Brown, *Adv. Part. Phys.* **1** (1968) 71 and SLAC-75 Rev. 4 (1982).

[15] P. Emma and R. Brinkmann, SLAC-PUB-7554 (1997).

[16] D. Douglas, JLAB-TN-98-012 (1998).

[17] M. Venturini, *Nucl. Instrum. Methods Phys. Res. A* **794** (2015) 109. https://doi.org/10.1016/j.nima.2015.05.019

[18] C. Mitchell *et al.*, *Phys. Rev. ST Accel. Beams* **16** (2013) 060703. https://doi.org/10.1103/PhysRevSTAB.16.060703

[19] S. Di Mitri and M. Cornacchia, *Phys. Rep.* **539** (2014) 1. https://doi.org/10.1016/j.physrep.2014.01.005

[20] M. Dohlus, P. Emma and T. Limberg, in Ref. [1].

[21] S. Di Mitri, M. Cornacchia and S. Spampinati, *Phys. Rev. Lett.* **110** (2013) 014801. https://doi.org/10.1103/PhysRevLett.110.014801

[22] S. Di Mitri and M. Cornacchia, *Eur. Phys. Lett.* **109** (2015) 62002. https://doi.org/10.1209/0295-5075/109/62002

[23] S. Di Mitri, *Nucl. Instrum. Methods Phys. Res. A* **806** (2016) 184. https://doi.org/10.1016/j.nima.2015.10.015

[24] A. Novokhatski, SLAC-PUB-14893 (2012). http://slac.stanford.edu/pubs/slacpubs/14750/slac-pub-14893.pdf

[25] S. Di Mitri *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **608** (2009) 19. https://doi.org/10.1016/j.nima.2009.06.028

[26] S. Bettoni *et al.*, *Phys. Rev. ST Accel. Beams* **19** (2016) 034402. https://doi.org/10.1103/PhysRevAccelBeams.19.034402

[27] V. Yakimenko, M. Fedurin, V. Litvinenko, A. Fedotov, D. Kayran and P. Muggli, *Phys. Rev. Lett.* **109** (2012) 164802. https://doi.org/10.1103/PhysRevLett.109.164802

[28] R. Li, C.L. Bohn and J.J. Bisognano, Proc. 1997 Particle Accelerator Conf., Vancouver, BC, Canada, 1997, p. 1644.

[29] R. Warnock and P. Morton, *Part. Accel.* **25** (1990) 113.

[30] K.L.F. Bane, *Int. J. Mod. Phys. A* **22** (2007) 3736. https://doi.org/10.1142/S0217751X07037391

[31] J. Rosenzweig *et al.*, DESY Report No. TESLA-FEL-96-15 (1996).

[32] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, *Nucl. Instrum. Phys. Res. A* **490** (2002) 1. https://doi.org/10.1016/S0168-9002(02)00905-1

[33] E.L. Saldin, E.A. Schneidmiller and M.V. Yurkov, DESY Report No. TESLA-FEL-2003-02 (2003).

[34] J. Qiang *et al.*, *Phys. Rev. ST Accel. Beams* **12** (2009) 100702. https://doi.org/10.1103/PhysRevSTAB.12.100702

[35] M. Venturini, *Phys. Rev. ST Accel. Beams* **11** (2008) 034401. https://doi.org/10.1103/PhysRevSTAB.11.034401

[36] Z. Huang *et al.*, *Phys. Rev. ST Accel. Beams* **7** (2004) 074401. https://doi.org/10.1103/PhysRevSTAB.7.074401

[37] Y. Li and L. Wang, *Nucl. Instrum. Phys. Res. A* **769** (2015) 44. https://doi.org/10.1016/j.nima.2014.09.038

[38] J. Murphy, S. Krinsky, and R. Gluckstern, inProc. 1995 Particle Accelerator Conf., 1995, p. 2980.

[39] S. Heifets, S. Krinsky and G. Stupakov, *Phys. Rev. ST Accel. Beams* **5** (2002) 064401. https://doi.org/10.1103/PhysRevSTAB.5.064401

[40] R. Gluckstern, *Phys. Rev. D* **39** (1989) 2780. https://doi.org/10.1103/PhysRevD.39.2780

[41] S. Di Mitri, M. Cornacchia, S. Spampinati and S. Milton, *Phys. Rev. ST Accel. Beams* **13**(1) (2010) 010702. https://doi.org/10.1103/PhysRevSTAB.13.010702

[42] M. Venturini, *Nucl. Instrum. Methods Phys. Res. A* **599** (2009) 140. https://doi.org/10.1016/j.nima.2008.11.004

[43] D. Ratner *et al.*, *Phys. Rev. ST Accel. Beams* **18** (2015) 030704. https://doi.org/10.1103/PhysRevSTAB.18.030704

[44] M. Borland *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **483** (2002) 268. https://doi.org/10.1016/S0168-9002(02)00325-X

[45] Z. Huang and K.-J. Kim, *Phys. Rev. ST Accel. Beams* **5** (2002) 074401. https://doi.org/10.1103/PhysRevSTAB.5.074401

[46] P. Piot, in Proc. 2004 Linear Accelerator Conf., WE104, Lübeck, Germany, 2004.

[47] M. Hüning and H. Schlarb, in Proc. 2003 Particle Accelerator Conf., WUPAB017, Portland, OR, USA, 2003.

[48] Z. Huang *et al.*, *Phys. Rev. ST Accel. Beams* **13** (2010) 020703. https://doi.org/10.1103/PhysRevSTAB.13.020703

[49] S. Spampinati *et al.*, *Phys. Rev. ST Accel. Beams* **17** (2014) 120705. https://doi.org/10.1103/PhysRevSTAB.17.120705

[50] E.L. Saldin, E.A. Schneidmiller andM. Yurkov, *Nucl. Instrum. Methods Phys. Res. A* **528** (2004) 355. https://doi.org/10.1016/j.nima.2004.04.067

[51] M. Venturini, *Phys. Rev. ST Accel. Beams* **10** (2007) 104401. https://doi.org/10.1103/PhysRevSTAB.10.104401

[52] C.-Y. Tsai *et al.*, in Proc. 37th Int. Free Electron Laser Conf., MOP087, Daejeon, Korea, 2015.

[53] E. Ferrari *et al.*, *Phys. Rev. Lett.* **112** (2014) 114802. https://doi.org/10.1103/PhysRevLett.112.114802

[54] Marinelli *et al.*, *Phys. Rev. Lett.* **116** (2016) 254801. https://doi.org/10.1103/PhysRevLett.116.254801

[55] E. Roussel *et al.*, *Phys. Rev. Lett.* **115** (2015) 214801. https://doi.org/10.1103/PhysRevLett.115.214801

[56] E.A. Schneidmiller and M.V. Yurkov, *Phys. Rev. ST Accel. Beams* **13** (2010) 110701. https://doi.org/10.1103/PhysRevSTAB.13.110701

[57] Marinelli *et al.*, *Phys. Rev. Lett.* **110** (2013) 264802. https://doi.org/10.1103/PhysRevLett.110.264802

Bibliography

I. S. Ko and W. Chou, Eds., ICFA Beam Dynamics Newsletter No. 38 (2005).

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Electron Sources and Injection Systems

E. Chiadroni
National Laboratory of Frascati—INFN, Frascati, Italy

Abstract
High-brightness photo-injectors are mandatory for several applications, for instance plasma-based accelerators, linear colliders, novel radiation sources, such as free-electron lasers, terahertz radiation sources, and inverse Compton scattering sources; all require the production, acceleration, and manipulation of high-brightness electron beams.

Keywords
Brightness; photo-emission; photo-injectors.

1 Introduction

The physics and technology concerning the generation, manipulation, and transport of high-energy, high-quality electron beams are of crucial importance in different fields of science, e.g., for the R&D of future-generation light sources and novel plasma-based accelerators. Indeed, free-electron lasers (FELs) [1–4], energy-recovery linacs [5], light sources, inverse Compton scattering sources [6], and plasma-based accelerators [7] all demand high-brightness electron beams. The figure of merit of this electron source is the normalized brightness in 6D phase space. There is no unique definition of brightness, as the literature shows [8–13]; I choose in this regard the definition reported in Ref. [10] extended to the 6D phase space:

$$B_{\mathrm{6D}} = \frac{2I_{\mathrm{p}}}{\frac{\Delta\gamma}{\gamma}\varepsilon_{\mathrm{n}}^2} \,. \tag{1}$$

Efforts are made to maximize the brightness, thus minimizing the transverse projected emittance (ε_{n} is of the order of millimetre milliradians) and the energy spread ($\Delta\gamma/\gamma \sim 0.1\%$), or increasing the peak current (I_{p} of the order of kiloamps), which means shortening the electron bunch down to femtosecond-scale durations. In particular, in both energy-recovery linacs [5] and FELs [1–4], high time-resolution user experiments require extremely short X-ray pulses (shorter than femtoseconds), imposing the need for small and linear longitudinal emittances to allow for proper compression along the linac. Indeed, the exponential gain of self-amplified spontaneous emission FELs [14], as shown in Section 3, depends on the peak current and not on the charge, at the price of a reduced number of photons per pulse, with the advantage that a reduction of beam charge allows better control of the beam quality.

A great development of FEL user facilities has been driven by the advent of radio-frequency (RF) based photo-injectors, contributing to an improvement of the normalized transverse emittance of at least one order of magnitude [15].

In a photo-injector, electrons are emitted by a photocathode, located inside an RF cavity, illuminated by a laser pulse, so that the time structure of the electron beam can be controlled and shaped on a picosecond or subpicosecond time-scale via the laser pulse. This feature is one of the main advantages of laser-based RF injectors. Indeed, the space charge can be controlled by reducing the beam charge density, especially in the cathode region, where the beam energy is low, by working with larger laser spot sizes; space charge can also be controlled by increasing the bunch length, which, however, contributes to increasing the longitudinal emittance, requiring compression methods. In addition, to preserve the brightness, the emitted electrons must be rapidly accelerated to relativistic energies, thus damping the space-charge forces, which scale as $1/\gamma^2$, and resulting in a partial mitigation of the emittance growth. At this regard, RF fields allow for large, from ≈40 to ≈130 MV/ m, electric fields at the cathode surface;

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.401

therefore, a laser-assisted RF gun is the preferable choice because of the high peak field, $\sim 10^2$ MV/ m, and the possibility of shaping the 6D electron-beam phase spaces by acting on the laser pulse distributions, both transverse and longitudinal. In addition, the generation of longitudinally modulated electron bunches directly at the cathode by means of a *comb-like* laser profile [16] is advantageous for the successful development of new classes of application, e.g., plasma wakefield accelerators [17], high-power narrow-band coherent terahertz sources [18], and two-colour FEL radiation [19–21].

In this report, I will overview electron injectors, highlighting the physics of the emission process and focusing on the beam dynamics in normal conducting RF guns. An extensive and detailed study can be found in Ref. [22]. In addition, I will introduce emittance compensation techniques to limit emittance growth at the end of the injector, integrated in a RF-based longitudinal compression method.

2 Definition of brightness

In 1939, Von Borries and Ruska, who were awarded the Nobel Prize in 1986 for the invention of the electronic microscope [23], introduced the concept of beam brightness as

$$B_{\mathrm{microscope}} = \frac{I}{A\Omega} = \frac{Ne}{\pi r^2 \pi \alpha^2 \Delta t} \approx \text{constant} , \tag{2}$$

where $Ne/\Delta t$ is the electron current density escaping from the cathode surface with area $A = \pi r^2$, $\Omega = \pi\alpha^2$ is the solid angle in which electrons are emitted, and Δt is the time interval. This quantity is practically constant in the microscope column: therefore, the smaller the spot size, the larger the divergence. The brightness is extremely important because it defines the quality of the source and sets the kind of experiments that can be done. Indeed, for imaging applications, the larger the number of electrons, the better the contrast; spatial resolution and coherence are enhanced by a small area and a collimated beam, i.e., small angles α, while temporal resolution is improved by short pulses. The definition in Eq. (2) still holds nowadays in the field of electron microscopy, with peak values of $B_{\mathrm{microscope}}$ up to 10^{13} A/ m^2/ sr.

In analogy with the electron microscope brightness, it is convenient to define the 6D electron-beam brightness to compare and describe electron sources, as the number of electrons per unit volume V_{6D} occupied by the beam in 6D phase space, i.e., transverse (x, p_x, y, p_y), longitudinal (z, p_z), and proportional to the product of the three normalized emittances:

$$B_{\mathrm{6D}} = \frac{Ne}{V_{\mathrm{6D}}} \propto \frac{Ne}{\varepsilon_{\mathrm{n}x}\varepsilon_{\mathrm{n}y}\varepsilon_{\mathrm{n}z}} . \tag{3}$$

Liouville's theorem states that the phase space volume, bounded by a closed, arbitrary surface in the phase space, is constant, provided that only conservative forces act on the particles. As long as the particle dynamics in the beamline elements (transport optics, accelerating sections, etc.) can be described by Hamiltonian functions, that is, neither particle–particle collisions nor stochastic processes are considered, the phase space density will stay constant throughout the accelerator.

If we write the longitudinal emittance explicitly, in terms of energy spread, σ_γ, and bunch duration, σ_t, then the physical meaning of the 6D brightness is clear,

$$B_{\mathrm{6D}} \propto \frac{Ne}{\varepsilon_{\mathrm{n}x}\varepsilon_{\mathrm{n}y}\sigma_t\sigma_\gamma} : \tag{4}$$

a large number of quasi-monochromatic electrons, concentrated in very short bunches, with small transverse size and divergence, which means small transverse emittance. For a fixed charge/bunch this translates to preserving the transverse emittance and increasing the final current by reducing the bunch length. In numbers, for typical electron-beam parameters ($N \approx 10^9$, $\sigma_\gamma \approx 10^{-3}$, $\varepsilon_{\mathrm{n}} \approx 1$ mm mrad, $\sigma_t < 1$ ps), the 6D brightness is of the order of 10^{15} A/ m^2.

The maximum brightness theoretically achievable by an electron beam is set by the Heisenberg uncertainty principle with one electron in each elementary quantum h^3 unit of phase space volume:

$$B_{\text{quantum}} = \frac{2e}{h^3}(m_0c^2)^3 = \frac{2e}{\lambda_\text{c}^3}\,, \tag{5}$$

with h the Planck constant, e the elementary charge, m_0 the particle mass, and c the speed of light; λ_c is the Compton wavelength, which for electrons is 2.426 pm. Replacing numbers in Eq. (5), $B_{\text{quantum}} \approx 10^{25}$ A/ m^2, being 10 orders of magnitude larger than typical 6D brightnesses in photo-injectors.

The maximum brightness practically achievable is limited by the degeneracy parameter δ, which represents the number of particles per elementary volume of the phase space: $B = \delta B_{\text{quantum}}$. In principle, assuming the photo-emission process by metal photocathodes, we expect a peak brightness approaching the quantum limit, since the degeneracy factor inside a metal is $\sim$1. However, in practice, we have 10 orders of magnitude less. Electron emission mechanisms and Coulomb interactions play a crucial role in this severe reduction of beam brightness [24].

The brightness generated at the electron source represents the ultimate value. Possible sources of emittance growth are:

– non-linear space-charge forces;
– non-linear forces from electromagnetic components, e.g., due to wakefields;
– synchrotron radiation emission (in magnetic compressors).

Sources of emittance degradation need to be kept under control, keeping in mind that the final beam quality is set by the linac and ultimately by its injector and electron source; therefore, a careful definition and specific requirements for both electron sources and injection systems are mandatory.

3 Applications of high-brightness electron beams

The great improvement of the characteristics of electron injectors started in the 1960s with thermionic guns [25]. Thermionic RF guns employ cathodes that must be heated to allow emission of electrons. Commonly used cathode materials are LaB_6, CeB_6, and BaO, with typical operating temperatures of the order of 1000 °C. The disadvantages of thermionic guns lie in the fact that emission occurs throughout the RF accelerating phase and during every RF period, resulting in a beam with large momentum and time spread. However, the high stability is the main advantage.

In 1985, the need for fast and precise control of the electron pulse shape led to the first use of photocathode RF guns because of the impressive reduction in transverse emittance, more than a factor of ten [15], promoted by the ability to shape drive laser pulses and rapidly accelerate electrons from rest to relativistic energies.

Starting from the first working prototype of an RF gun [26, 27], RF photo-injectors are nowadays routinely exploited as electron sources for FEL user facilities [1–4] and multidisciplinary test facilities, such as SPARC_LAB [28]. Radio-frequency photo-injectors are also fundamental for the successful development of plasma-based accelerators where external injection schemes are considered, i.e., particle-beam-driven [17] and laser-driven [29] plasma wakefield accelerators, since the ultimate beam brightness and its stability and reproducibility are strongly influenced by the RF-generated electron beam.

The main supporter for the development of high-brightness injectors is represented by X-ray free-electron lasers [30]. In particular, for FEL applications, the 5D brightness is often used to compare electron sources:

$$B_{\text{5D}} = \frac{2I_\text{p}}{\varepsilon_{\text{n}x}\varepsilon_{\text{n}y}} = \frac{2I_\text{p}}{(\beta\gamma)^2\varepsilon_x\varepsilon_y}\,, \tag{6}$$

and is the relativistic analogue of the microscopic brightness.

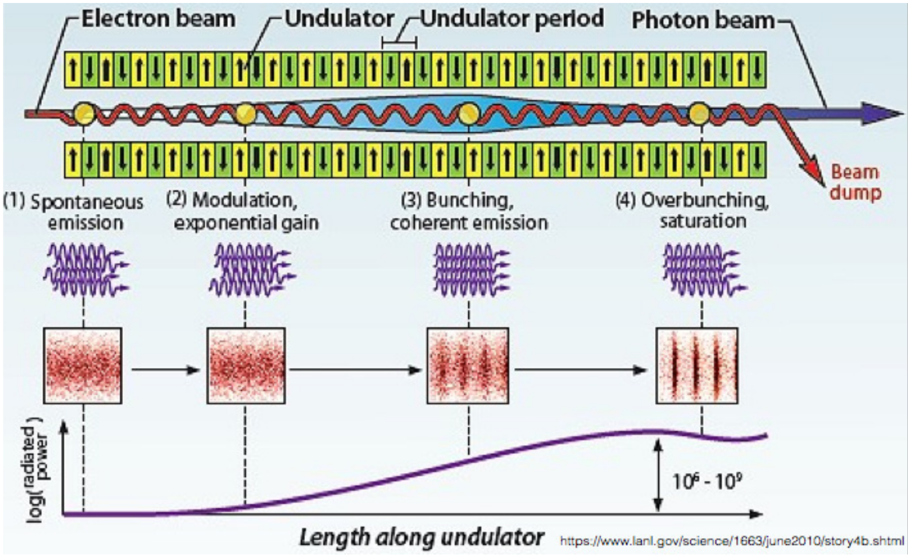


Fig. 1: Top: propagation of electron beam in undulator module. Middle: evolution of electron bunching along undulator. Bottom: exponential increase of FEL radiation power along the undulator.

In an electron beam entering a magnetic undulator, electrons start to oscillate with period λ_{u} and emit spontaneous radiation as synchrotron radiation because of the oscillating trajectory. They start to lose energy in favour of the radiation and since the trajectory in the undulator is energy-dependent, while propagating they start to bunch, which means that they start to modulate themselves longitudinally, as shown in Fig. 1.

Emission begins to be coherent and the radiation wavelength is set by the resonance condition:

$$\lambda_{\mathrm{r}} = \frac{\lambda_{\mathrm{u}}}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\theta^2\right) , \tag{7}$$

where γ is the electron-beam Lorentz factor, $K = (eB_0\lambda_{\mathrm{u}})/(2\pi m_0 c)$ is the undulator parameter, which defines the oscillation amplitude, and θ is the emitted angle with respect to the axis. B_0 is the maximum amplitude of the magnetic field and m_0 is the electron mass, thus K is a measure of the on-axis undulator magnetic field once the undulator period is fixed.

Because of the longitudinal modulation of the electrons in slices, the radiated power along the undulator increases exponentially as a function of gain length, which, in 1D FEL theory, is defined as

$$L_{\mathrm{g}} = \frac{\lambda_{\mathrm{u}}}{4\pi\rho\sqrt{3}} , \tag{8}$$

assuming a mono-energetic beam and neglecting space-charge forces. L_{g} represents the length necessary for the radiation field to be increased by e, Napier's constant; therefore it is an indication of how long the undulator chain should be in order to reach saturation. Beginning from shot noise, it typically takes a self-amplified spontaneous emission FEL about 20 L_{g} to reach saturation. The gain length is related to the beam brightness through the Pierce parameter, ρ,

$$\rho = \frac{1}{2\gamma}\left[\frac{I_{\mathrm{p}}}{I_{\mathrm{A}}}\left(\frac{\lambda_{\mathrm{u}} K[JJ]}{\sqrt{8}\pi\sigma_x}\right)^2\right]^{1/3} , \tag{9}$$

which scales with the peak current as $\rho \propto I_{\mathrm{p}}^{1/3}$, resulting in $L_{\mathrm{g}} \propto B_{\mathrm{5D}}^{-1/3}$; $I_{\mathrm{A}} = 4\pi\varepsilon_0 \frac{m_0 c^3}{e} \approx 17$ kA is the Álfven current for an electron, $[JJ]$ is the Bessel function coupling factor depending on the undulator,[1]

[1] $[JJ]$=1 for helical undulators and $[JJ] = J_0\left(\frac{K^2}{4+2K^2}\right) - J_1\left(\frac{K^2}{4+2K^2}\right)$ for planar ones.

and σ_x is the r.m.s. electron transverse beam size in the undulator. I_A represents a limit on the amount of charge that can be transported. Physically, the beam self-magnetic field becomes sufficiently strong to stop the propagation of electrons, whose trajectories reverse in direction, so that most of the electrons are reflected back [31, 32].

The Pierce parameter, also called the universal FEL parameter, gives an indication of the gain of the FEL and the speed of the bunching process; typically for kiloamp-current beams, $\rho \sim 10^{-3}$ to 10^{-4} in the VUV/X-ray regime. In practice, $\rho \sim 1/N_\mathrm{u}$, where N_u is the number of undulator periods electrons must travel to increase the FEL power by 2e times. Indeed, the saturation power of the self-amplified spontaneous emission FEL radiation is $P_\mathrm{sat} = P_\mathrm{in}\mathrm{e}^{z/L_\mathrm{g}}$, with P_in the initial beam power; $P_\mathrm{in} = \rho I_\mathrm{p} E_\mathrm{beam}$.

In general, because of the finite energy spread and non-negligible space-charge effects, the exponential gain might weaken, resulting in an increased undulator length to preserve the FEL gain and aim at laser saturation. Therefore, the electron-beam brightness has to be extremely high in order to obtain a relatively short FEL gain length.

In addition, the brightness plays an important role on the efficiency of the FEL process and on its spectral characteristics. Indeed, to ensure coupling between the electron beam and the radiation it emits, the matching condition for transverse emittance has to be satisfied ($\varepsilon_\mathrm{n} < \gamma\lambda_\mathrm{r}/4\pi$). Concerning the energy spread, the matching condition requires $\sigma_\delta < \rho \approx 10^{-3}$, which defines the radiation bandwidth $\Delta\omega/\omega \approx 1/N_\mathrm{u} \approx \rho$.

Beyond X-ray FELs, high-brightness electron beams are needed for several other applications: for instance, high-power FELs, either for soft X-rays or IR, energy-recovery linacs, and linac-based terahertz sources all demand average currents of the order of milliamps, repetition rates of the order of megahertz, and low emittance. In addition, terahertz sources require ultrashort electron bunches, of the order of 10–100 fs, to extend the frequency spectrum up to several terahertz. Concerning the development of radiation sources, high-brightness electron beams are also fundamental in the success of γ-ray sources, such as inverse Thomson or Compton scattering [6]. A relativistic electron beam collides head-on with an ultrafast laser to produce scattered photons upshifted in energy in a $1/\gamma$ forward cone angle. The high-charge beam increases the X-ray yield, while short bunches allow for ultrafast X-rays. Both small emittance, which enables focusing to a micrometre-scale spot size, and low energy spread contribute to narrow the radiation bandwidth [33].

High-brightness photo-injectors are fundamental for the successful development of plasma-based accelerators. In particular, in particle-beam-driven plasma wakefield accelerators, the high-gradient wakefield is driven by an intense, high-energy charged-particle beam, named the driver beam, as it passes through the plasma. The space-charge forces of the electron bunch blow out the plasma electrons, which rush back in and overshoot, setting up a plasma density oscillation at a frequency $\omega_\mathrm{p} = \sqrt{(n_0 e^2)/(\varepsilon_0 m_0)}$, which depends on the plasma density n_0. A second, appropriately phased accelerating beam, named the witness beam, and containing fewer particles than the drive beam, is then accelerated by the wake. The energy transfer from the drive bunch to the plasma is optimized by maximizing the transformer ratio $R = |E_{+,\mathrm{max}}/E_{-,\mathrm{max}}|$, defined as the ratio between the maximum accelerating field behind the drive bunch, $E_{+,\mathrm{max}}$, where the particles of the witness bunch can be placed, and the maximum decelerating field within the drive bunch, $E_{-,\mathrm{max}}$. R quantifies the energy gain of a witness bunch placed at the accelerating phase. To accelerate high-brightness electron beams, both driver and witness bunches must be focused to a size of a few micrometres, as requested by the transverse matching condition at the plasma entrance, i.e.,

$$\sigma_\perp^\mathrm{matching} = \left(\frac{2}{\gamma}\right)^{1/4} \sqrt{\frac{\varepsilon_x}{k_\mathrm{p}}}\,, \tag{10}$$

with $k_\mathrm{p} = 2\pi/\lambda_\mathrm{p}$ the plasma wave number, λ_p the plasma wavelength, and ε_x the transverse emittance; therefore, the lower the emittance the smaller the beam transverse size. In addition, the great flexibil-

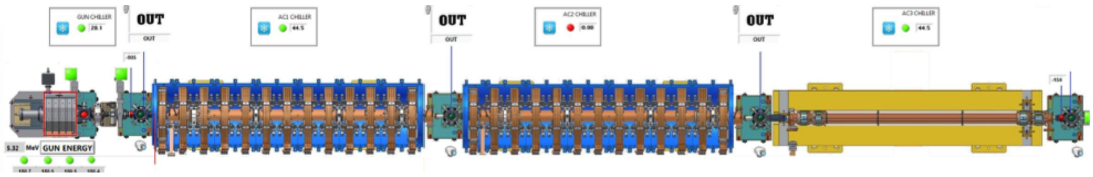


Fig. 2: Layout of SPARC_LAB photo-injector: a copper cathode, illuminated by UV laser pulses and embedded in a 1.6-cell standing wave S-band RF gun, generates a 5.6 MeV electron beam. A four-coil solenoid magnet focuses and matches the beam into three S-band travelling wave accelerating structures to boost the energy to 180 MeV. The first two accelerating structures are embedded by multicoil solenoid magnets to provide additional focusing when the first section is used as a RF compressor in the so-called velocity-bunching regime.

ity of photo-injectors to shape and manipulate the longitudinal phase space allows for resonant plasma wakefield acceleration driven by multibunch trains to increase the transformer ratio [34].

4 Photo-injector theory

An electron injector is the overall system from the electron source, the cathode, up to the place where electrons have energies where space-charge forces, which scale as $1/\gamma^2$, can be considered negligible. Therefore, the beam evolution is no longer space-charge dominated.

There are, at present, three broadly identifiable cathode technologies for electron-beam sources and two main accelerator technologies used to perform the initial acceleration of the beam from those cathodes, i.e., normal conducting (LCLS, FERMI, SPARC_LAB) and superconducting (FLASH, XFEL). Most injector systems use photocathodes; one exception is the SACLA XFEL [35, 36], which successfully uses a thermionic cathode.

In this review, I will focus on the experience at SPARC_LAB which, being an R&D facility on high-brightness photo-injectors, is paradigmatic for describing electron-beam generation, manipulation, and acceleration. I recommend Ref. [22] for a more complete treatment of electron sources and injection systems and Refs. [30, 37] for FEL physics and direct experience at user facilities.

The typical layout of an electron photo-injector is reported in Fig. 2. A photo-injector consists of a laser-generated source, followed by an electron-beam optical system, which preserves and matches the beam into a high-energy accelerator. A photocathode, embedded in a RF gun, releases picosecond or subpicosecond electron bunches when irradiated by laser pulses of given wavelength, depending on the cathode material. The high electric fields produced by the RF gun are necessary both to extract the high current and to minimize the effects of space charge on emittance growth. Since the RF gun acts as a strong defocusing lens, a solenoid magnet is needed to focus the divergent beam, preventing particle losses, and to properly match the electron beam to the accelerating cavities, minimizing emittance growth at the end of the accelerating chain (i.e., the linac exit).

A typical photocathode RF system depicts a $1\frac{1}{2}$-cell with a cathode embedded in the half cell. The cathode might be either metal or semiconductor depending mainly on applications: in the low-charge regime, the ultimate brightness performance of the linac is set by the cathode intrinsic emittance, while for high-repetition-rate photon sources, high-quantum-efficiency photocathodes are required.

The elements that constitute a photo-injector are highlighted in the following sub-sections.

4.1 Photocathode emission

The emission process determines the fundamental lower limit of the beam emittance, called the intrinsic emittance, which represents the minimum emittance achievable and depends on the emission mechanism, i.e.:

– thermionic emission [25];

– field emission [38];
– photo-electric emission [39].

The ideal cathode should have low intrinsic emittance, since this limits the maximum achievable brightness; high quantum efficiency, to reduce the laser load; long lifetime, uniform emission, since hot spots on the cathode and surface roughness contribute to emittance growth [40]; and fast response, which allows transverse and longitudinal electron-beam manipulation by properly shaping the laser pulse. At present, there is no cathode that meets all these criteria. Semiconductor photocathodes, for instance, are the best choice for high quantum efficiency and low emittance, while metallic photocathodes are chosen because of their fast temporal response, long operational lifetime, and high vacuum compatibility. An extensive and detailed study is reported in Ref. [22].

To understand the mechanism of electron emission from a cathode material, it is important to identify the fields and the potentials at the cathode surface and close to it. The electric potential energy as a function of the distance, x, from the cathode is given as

$$e\Phi = e\Phi_{\text{work}} - \frac{e^2}{16\pi\varepsilon_0 x} - eE_0 x \,, \tag{11}$$

and is the sum of the work function energy, $e\Phi_{\text{work}}$, which represents the work necessary to separate a charge from its image within the material and depends on the material, the image charge potential, $-e^2/(16\pi\varepsilon_0 x)$, and the potential corresponding to a constant electric field normal to the surface, E_0. Electrons with energies greater than the work function can escape the barrier, while those with lower energies can tunnel through it. In the case of thermionic and photo-electric emission, electrons are excited above the barrier to escape, while in the field emission the barrier height is lowered by the external field to encourage tunnelling. The reduction of the barrier by the applied field is called the Schottky effect and plays a central role in all emission processes, especially field emission [41].

Photo-electric emission is well described by Spicer's three-step model.

1. Photon energy absorption by the electron:
   (a) the optical skin depth depends on photon wavelength (∼14 nm for UV light on copper); reflectivity, and absorption as the photons travel into the cathode.
2. Electron transport to the surface through:
   (a) electron–electron scattering, dominant for metals;
   (b) electron–phonon scattering, dominant for semiconductors;
   (c) angular cone of escaping electrons.
3. Electron escape through the barrier:
   (a) Schottky effect and abrupt change in electron angle across the metal–vacuum interface; classical escape over the barrier due to the applied field.

Combining the three steps, the quantum efficiency can be expressed in terms of the probabilities of these processes occurring:

$$\mathrm{QE}(\omega) = [1 - R(\omega)] F_{e-e}(\omega) \frac{\int_{E_\mathrm{F}+\Phi_{\text{eff}}-\hbar\omega}^{E_\mathrm{F}} \mathrm{d}E \int_{\sqrt{\frac{E_\mathrm{F}+\Phi_{\text{eff}}}{E+\hbar\omega}}}^{1} \mathrm{d}(\cos\vartheta) \int_0^{2\pi} \mathrm{d}\phi}{\int_{E_\mathrm{F}-\hbar\omega}^{E_\mathrm{F}} \mathrm{d}E \int_{-1}^{1} \mathrm{d}(\cos\vartheta) \int_0^{2\pi} \mathrm{d}\phi} \,. \tag{12}$$

The first factor represents the probability of a photon being absorbed by the metal, which depends on the optical reflectivity: $R(\omega) \sim 40\%$ for metals and $R(\omega) \sim 10\%$ for semiconductors; the second factor, $F_{e-e}(\omega)$, is the probability that an electron will reach the surface without scattering with another electron, which for metals is 20%; the third factor is the probability that an electron will be excited into a state with sufficient perpendicular momentum to escape the material, over the barrier:

– occupied states with enough energy to escape, $\sim$0.04;
– electrons with angle within the max angle for escape, $\sim$0.01;
– azimuthally isotropic emission, $\sim$1.

The total energy of an electron inside the cathode after absorption of the photon is $E + \hbar\omega$, therefore the total momentum inside and outside is

$$p_{\text{total,in}} = \sqrt{2m_0(E + \hbar\omega)}\,, \qquad p_{\text{total,out}} = \sqrt{2m_0(E + \hbar\omega - \Phi_{\text{eff}} - E_{\text{F}})}\,, \tag{13}$$

where E_{F} is the Fermi level. The electron within the material, approaching the boundary surface, needs to have sufficient longitudinal momentum to escape, i.e.,

$$\sqrt{2m_0(E + \hbar\omega)}\cos\vartheta_{\text{in}} \geq \sqrt{2m_0(\Phi_{\text{eff}} + E_{\text{F}})}\,, \tag{14}$$

where $\cos\vartheta_{\text{max,in}} = \sqrt{\Phi_{\text{eff}} + E_{\text{F}}/E + \hbar\omega}$. In analogy to photons, electrons refract as they transit the cathode interface, as a consequence of the boundary condition requiring conservation of transverse momentum across the metal–vacuum transition, $p_{x,\text{in}} = p_{x,\text{out}}$ and $p_{\text{total,in}}\sin\vartheta_{\text{in}} = p_{\text{total,out}}\sin\vartheta_{\text{out}}$.

Combining the probabilities, it is possible to evaluate the nominal quantum efficiency for a copper photocathode, i.e., QE $\sim 5 \times 10^{-5}$ [42]. As a rule of thumb, the quantum efficiency is given by QE $= N_{\text{e}}/N_{\text{ph}} = (h\nu[\text{eV}]/E_{\text{laser}}[\text{J}])Q[\text{C}]$; therefore, to extract 1 nC of charge from a copper photocathode, a UV (266 nm) laser energy of the order of 100 µJ is required.

The linear dependence between the number of emitted electrons and the incident number of photons holds until a critical intensity of the laser pulse is reached. Indeed, for higher laser intensities, electrons will be emitted even for photon energies below the work function. In this case, $n > 1$ photons have to be absorbed at the same time in order to promote an electron to an unbound state with enough kinetic energy to escape the material, the scaling of the emitted charge being the nth power of the laser intensity. Taking advantage of the multiphoton emission process, an IR-wavelength laser incident on a metal cathode might also be used for electron-beam generation, in particular, when ultrashort pulses are required [43].

Depending on the application, the cathode choice must be driven not only by the quantum efficiency, but also by the intrinsic emittance, which represents the ultimate achievable emittance.

The intrinsic emittance is a feature of the emission process. To derive it, the definition of r.m.s. emittance, $\varepsilon_{\text{n},x} = \beta\gamma\sqrt{\langle x^2\rangle\langle x'^2\rangle - \langle xx'\rangle^2}$, can be used, observing that the correlation term, $\langle xx'\rangle^2$, is null out of the cathode. Therefore,

$$\varepsilon_{\text{n},x} = \sigma_x \frac{\sqrt{\langle p_x^2\rangle}}{m_0 c} \tag{15}$$

is a function of the laser pulse spot size, σ_x, and the transverse momentum, p_x, as determined by the emission process.

By calculating the IInd-order transverse momentum of the electron distribution function, we get the photo-electric normalized emittance as

$$\varepsilon_{\text{n},x}^{\text{intrinsic}} = \sigma_x \sqrt{\frac{\hbar\omega - \Phi_{\text{eff}}}{3m_0c^2}}\,. \tag{16}$$

Therefore, to minimize the intrinsic emittance, one has to reduce the energy difference, $\Delta = \hbar\omega - \Phi_{\text{eff}}$, between the energy of the incident photons and the effective work function of the material. However, since the quantum efficiency is proportional to Δ^2, a low intrinsic emittance also results in a low quantum efficiency.

4.1.1 Space-charge limit emittance

As emitted electrons come out from the cathode surface, they create their own electric field and start to fill the entire region, which can be represented as two parallel plates, i.e., cathode and anode, placed at a distance d, with an applied field or bias potential, V. In the beam tail, the beam self-field is opposed to the applied field and increases with the extracted charge. The effective potential is then distorted, creating asymmetries in the tail and setting a maximum extractable current in the steady-state regime. The maximum current density extractable is given, by solving the Poisson equation, by the Child–Langmuir law [44], which expresses how the steady-state current varies with both the gap distance, d, and the bias potential of the parallel plates, V, that schematize the RF gun:

$$j_{\mathrm{CL,1D}} = \frac{4\varepsilon_0}{9}\sqrt{\frac{2e}{m_0}}\frac{V^{3/2}}{d^2} \, . \tag{17}$$

This formula has been derived by assuming an infinitely wide beam in the transverse dimensions (i.e., 1D approximation) that completely fills the accelerating gap. However, in state-of-art photo-injectors, the initial electron-beam pulse length is always much smaller than the accelerating gap; in addition, the laser spot size on the cathode is usually smaller than millimetres in diameter to decrease the cathode emittance contribution, therefore the 1D Child–Langmuir formula is no longer valid. Indeed, to describe a real case, it is convenient to introduce the aspect ratio, i.e., the ratio between the beam radius and its length, to define pancake-like ($\gg$1) and cigar-like ($\ll$1) beams. In the case of pancake-like beams, the maximum surface charge density is set by the cathode extraction field, while in cigar-like beams only a small part of the beam contributes to the space-charge field and a higher charge can be extracted. The space-charge limit is reached when the space-charge field equals the applied, external field, E_0, and electron emission saturates. At the space-charge limit, the emitted charge saturates and the emission becomes constant. In the RF gun, the signature for the observation of the space-charge limit is the non-linear dependence of the charge on the laser energy.

The space-charge limit sets a minimum value for the beam emittance, once the applied field at the cathode, E_a, and the required charge, Q_{bunch}, are known. Indeed, for a cylindrical uniformly filled beam with radius R, the r.m.s. size is $\sigma_x = R/2 = \sqrt{(Q_{\mathrm{bunch}})/(4\pi\varepsilon_0 E_{\mathrm{a}})}$; therefore, substituting it in the normalized intrinsic emittance for photo-electric emission (Eq. (16)), we get

$$\varepsilon^{\mathrm{SCL}} = \sqrt{\frac{Q_{\mathrm{bunch}}(\hbar\omega - \Phi_{\mathrm{eff}})}{4\pi\varepsilon_0 m_0 c^2 E_{\mathrm{a}}}} \, . \tag{18}$$

The dependencies of the space-charge limit emittance on both the applied field and the charge are reported in Fig. 3(a) and (b), respectively. The space-charge limit emittance sets a maximum transverse brightness, which in the case of pancake-like beams does not depend on the charge, but only on the applied RF field and on the emission process:

$$B_\perp = \frac{N}{\varepsilon_{\mathrm{n}x}\varepsilon_{\mathrm{n}y}} \qquad \to \qquad B_\perp^{\mathrm{max}} = 4\pi\varepsilon_0 \frac{E_{\mathrm{a}}}{e}\frac{m_0 c^2}{\hbar\omega - \Phi_{\mathrm{eff}}} \, . \tag{19}$$

4.2 RF gun

Once the electrons are out of the cathode surface, they have to be promptly accelerated to keep under control the emittance growth driven by space-charge forces. For this purpose, the relevant modes are those with large longitudinal electric fields E_z, since the energy transfer occurs when the particle's velocity is parallel to the electric field:

$$\frac{\mathrm{d}U}{\mathrm{d}t} = Q\vec{v}\cdot\vec{E} \, . \tag{20}$$

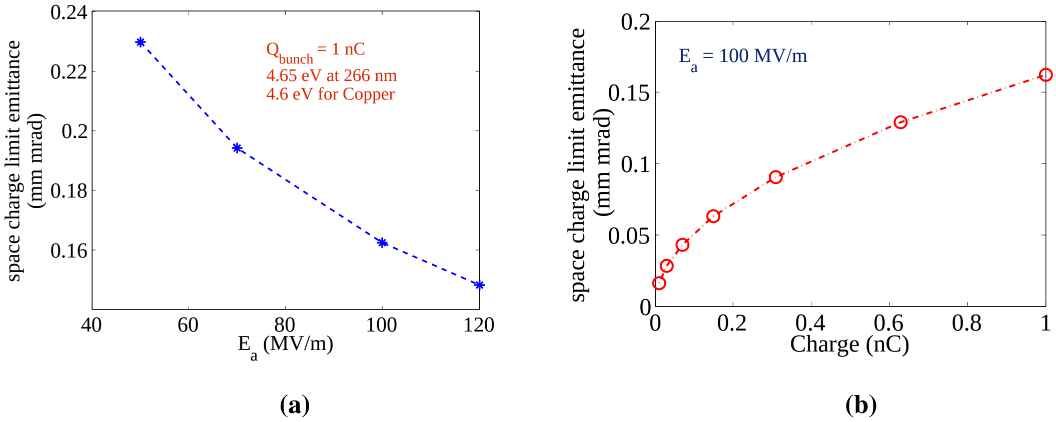


Fig. 3: Space-charge limit emittance as function of a) the applied field at and b) the charge for a Cu cathode (Φ_{eff}=4.6 eV), illuminated by a UV laser pulse ($\hbar\omega = 4.66$ eV at 266 nm).

These modes are the transverse magnetic or TM$_{m,n,p}$ modes, since $B_z = 0$ and m, n, and p denote the rotational symmetry, the radial dependence, and the longitudinal mode of the cavity, respectively. In particular, the m-mode number represents the azimuthal angle, defining the ϑ-dependence or rotational symmetry of the fields. For all RF guns, since a beam with rotational symmetry is desired, $m = 0$. The p-mode number denotes the longitudinal mode of cavity, contributing to the RF emittance; for this reason, the full cell length for most RF guns is $\lambda/2$ and $p = 1$.

Considering a pillbox geometry for the RF gun cavity, the longitudinal electric field is

$$E_z^{mnp}(r,z) = E_0 J_m(k_{mn}r)\cos(m\vartheta)\cos\left(\frac{2p\pi z}{\lambda}\right)\mathrm{e}^{\mathrm{i}\omega z/c}\,, \tag{21}$$

with $J_m(k_{mn}r)$ the mth-order Bessel function, k_{mn} the nth zero of the mth-order Bessel function. Considering the π-mode for a $1^{1/2}$-cell standing wave RF structure, therefore, $m = 0$, $n = 0$, $p = 1$, and the gun longitudinal field is $E_z = E_0\cos(kz)\sin(\omega t + \phi_0)$, presenting the maximum field on the cathode surface, i.e., at $z = 0$ ($k = \omega/c$, with ω the RF angular frequency, and ϕ_0 the RF phase at which electrons leave the cathode surface and start to be accelerated). The radial field E_r and the azimuthal field B_ϑ can be derived by solving the Maxwell equations:

$$E_r = \frac{kr}{2}E_0\sin(kz)\sin(\omega t + \phi_0) = -\frac{r}{2}\frac{\partial}{\partial z}E_z\,, \tag{22}$$

$$B_\vartheta = c\frac{kr}{2}E_0\cos(kz)\cos(\omega t + \phi_0) = \frac{r}{2c}\frac{\partial}{\partial t}E_z\,, \tag{23}$$

where $k = 2\pi/\lambda$ is the RF wave number.

The force acting on a particle in the RF gun can be derived by combining E_r and B_ϑ:

$$F = e(E_r - \beta c B_\vartheta)\,. \tag{24}$$

The radial momentum kick is obtained by integrating the radial force impulse over the position at the last iris:

$$\Delta p_r = e\int E_r\frac{\mathrm{d}z}{\beta c} = -\frac{e}{2}\int\frac{r}{\beta c}\frac{\partial E_z}{\partial z}\mathrm{d}z\,. \tag{25}$$

Assuming that the RF field is a constant step function over the gun length, the change in radial momentum is

$$\Delta p_r = -\frac{eE_0}{m_0c^2}r\sin\phi, \tag{26}$$

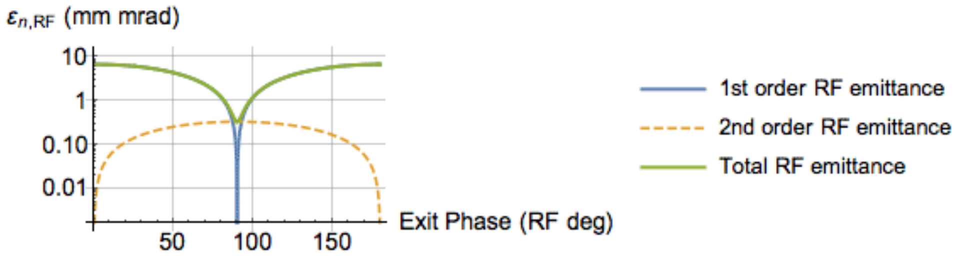


Fig. 4: RF emittance as a function of the exit phase for a 100 MV/ m gun with a 1 mm r.m.s. size beam and a Gaussian longitudinal distribution of 4 RF degrees r.m.s. at the exit iris, corresponding to a 10 ps FWHM bunch length. The total emittance (green solid line) is the quadratic sum of the first-order (blue solid line) and second-order (orange dashed line) emittances.

which depends on the applied field and the phase with respect to the wave, therefore, on the relative position of the particle within the bunch. Moving from cylindrical to Cartesian co-ordinates, we obtain the change in transverse momentum at the exit of the iris in terms of a kick angle x':

$$\Delta p_x = \beta\gamma x' = -\frac{eE_0}{2m_0c^2} x \sin\phi \qquad \rightarrow \qquad x' = -\frac{eE_0}{2\beta\gamma m_0 c^2} x \sin\phi. \tag{27}$$

The RF gun can then be schematized as a defocusing lens: the electron exiting the gun undergoes a transverse kick from the exit iris, which can be written in terms of a focal length, f_{RF}:

$$x' = \frac{x}{f_{\mathrm{RF}}} \qquad \rightarrow \qquad f_{\mathrm{RF}} = -\frac{2\beta\gamma m_0 c^2}{eE_0 \sin\phi} \,. \tag{28}$$

The focal strength is phase dependent, therefore electrons at different longitudinal positions, or slices, along the bunch, arriving at different phases at the gun exit, experience different kicks. In numbers (a SPARC_LAB case), an S-band standing wave RF gun operating at $E_0 = 110$ MV/ m, $\phi = 30°$, and an electron energy of 5 MeV, has a focal length $f_{\mathrm{RF}} = -18$ cm, where the minus sign denotes the defocusing effect of the gun exit iris. The different angular kicks received by the different slices along the bunch contribute to an increase in the overall projected emittance, which can be estimated by deriving the r.m.s. divergence from the variation of the angular dispersion with the exit phase, as

$$\Delta x' = -\frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{1}{f_{\mathrm{RF}}}\right)\Delta x \Delta\phi \,, \tag{29}$$

and averaging over the particle distribution we get

$$\sigma_{x'} = \frac{eE_0 \cos\phi}{2\gamma m_0 c^2}\sigma_x \sigma_\phi \,. \tag{30}$$

The impact of the phase dependence on the projected emittance is then summarized as

$$\varepsilon_{\mathrm{n}}^{\mathrm{RF}} = \frac{eE_0}{2m_0c^2}\sigma_x^2\sigma_\phi \sqrt{\cos^2\phi + \frac{\sigma_\phi^2}{2}\sin^2\phi} \,, \tag{31}$$

as shown in Fig. 4. At 90°, corresponding to the crest of the RF field, the phase space is strongly correlated; this means that all the slices lie on the same line, resulting in a minimum in the total RF emittance; conversely, far from the crest, each slice in the bunch explores more different fields, though contributing with a different slice emittance, resulting in a finite projected RF emittance.

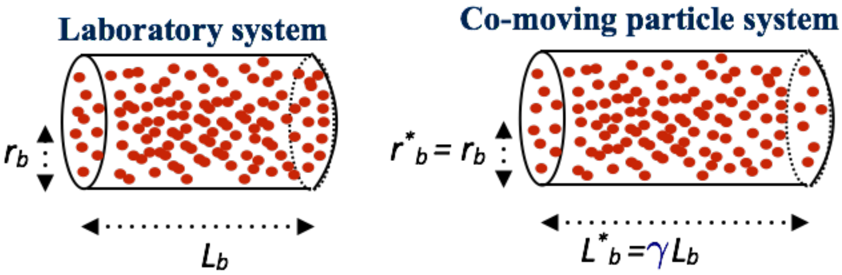


Fig. 5: Cylindrical bunch distribution in both the laboratory and the co-moving particle frame

4.2.1 Space-charge derivation

Space-charge forces influence the beam dynamics and are one of the main performance limitations in high-brightness photo-injectors.

To describe the origin of space-charge forces, let us consider the case of highly relativistic bunches (Fig. 5). In the laboratory system, let us assume N relativistic electrons uniformly distributed in a cylinder with radius r_b and length L_b; in a co-moving particle co-ordinate system, electrons are at rest and the field is a pure Coulomb field. When transforming from the laboratory frame to the co-moving one, the number of electrons and the radius remain invariant, while the length is expanded by a factor γ, the Lorentz factor. Therefore, since $\gamma \gg 1$, the approximation of an infinitely long cylindrical distribution is valid and, by applying the Gauss theorem, the electric field is retrieved having only a radial component:

$$E_r^*(r) = -\frac{Ne}{2\pi\varepsilon_0 L_\mathrm{b}^*}\frac{r}{r_\mathrm{b}^2}\,, \qquad r \leq r_\mathrm{b}\,, \tag{32}$$

$$E_r^*(r) = -\frac{Ne}{2\pi\varepsilon_0 L_\mathrm{b}^*}\frac{1}{r}\,, \qquad r \geq r_\mathrm{b}\,. \tag{33}$$

Transforming back to the laboratory frame, the radial component of the electric field yields a radial electric field and an azimuthal magnetic field [45]:

$$E_r(r) = \gamma E_r^*(r) = -\frac{Ne}{2\pi\varepsilon_0 L_\mathrm{b}}\frac{r}{r_\mathrm{b}^2}\,, \tag{34}$$

$$B_\phi(r) = \frac{v}{c^2}E_r(r)\,, \qquad r \leq r_\mathrm{b}\,. \tag{35}$$

The force a test particle inside the bunch experiences due to E_r and B_ϕ is given by the Lorentz force, $\vec{F} = -e(\vec{E} + \vec{v} \times \vec{B})$; therefore,

$$F_r(r) = \frac{Ne^2}{2\pi\varepsilon_0 L_\mathrm{b}}\frac{r}{r_\mathrm{b}^2}\left(1 - \frac{v^2}{c^2}\right) = \frac{Ne^2}{2\pi\varepsilon_0 L_\mathrm{b}}\frac{r}{r_\mathrm{b}^2}\frac{1}{\gamma^2}\,, \tag{36}$$

which has a linear dependence on r and is proportional to γ^{-2}. The overall force points outwards and is then a defocusing force, which vanishes for $\gamma \to \infty$.

The repulsive space-charge forces represent an unavoidable issue, which must be compensated to avoid a strong increase in emittance. For cylindrical bunch transverse distributions, the total space-charge force depends linearly on the displacement r, as shown in Eq. (36). The linear dependence on the particle displacement from the axis produces a quasi-laminar propagation of the beam, since there is a full correlation between the particles' position and their transverse angle; the particle motion follows a laminar flow, therefore, particle trajectories do not cross each other. Therefore, applying a magnetic lens, such as a solenoid, whose field increases linearly with r, the internal forces are counterbalanced and the beam is focused, contributing to preserve the emittance.

In the case of Gaussian transverse beam distribution, the radial force does not depend linearly on r, as

$$F_r(r) = \frac{Ne^2}{2\pi\varepsilon_0 L_\mathrm{b} r}\left(1 - \mathrm{e}^{-\frac{r^2}{2\sigma^2}}\right)\frac{1}{\gamma^2} : \tag{37}$$

it increases almost linearly from the axis ($r = 0$) up to 0.8σ of the Gaussian distribution, then it reverses its trend and starts decreasing for $r > 1.8\sigma$, since the charge density is lower than around the axis. In this case, space-charge forces cannot be fully compensated. A flat, uniform beam distribution is also advantageous in the longitudinal direction, because slices in the core or in the tails would have different charge density otherwise, experiencing a different defocusing effect due to space-charge along the bunch, while the same external focusing force is applied along the bunch.

4.3 Emittance compensation

In the 1980s, Carlsten [46] first explained how to reduce the emittance increase induced by space-charge in a RF gun. He suggested using a solenoid, placed at the exit of the RF gun, to control the emittance oscillations in the downstream drift.

The beam can be virtually divided, in the longitudinal direction, by thin slices, each having its own transverse phase space distribution. Assuming that each slices are independent and do not interact with each other, the projected emittance is the phase space ellipse, which encloses all the slice phase space. At the cathode, all the slices have low angular divergence and they can be considered almost aligned, resulting in a projected emittance which is close to the intrinsic one. As the beam leaves the cathode surface, the slices undergo different angular kicks, depending on their peak current. The ensemble of phase spaces of each slice forms a kind of fan, resulting in a larger enclosing ellipse and, therefore, in an increased projected emittance. A solenoid, whose focusing field is linear, can be used to give to each slice a kick of the same sign. Indeed, the beam entering the end radial field of the solenoid receives a transverse kick and rotates inward or outward, depending on the solenoid polarity. The particle is then closer in when passing through the end radial field at the opposite end of the solenoid and, since it is further in, the kick is smaller. After the solenoid, the beam drifts a distance with the slices all converging at a beam waist, placed at the entrance of the high-energy linac. To prevent additional space-charge emittance increase in subsequent accelerating sections, the final emittance minimum has to be reached at high beam energy so that space-charge forces are sufficiently damped. Indeed, the linac boosts the beam energy from the space-charge-dominated regime to the emittance-dominated one, freezing the aligned slices. Therefore, a solenoidal field is needed to focus the beam and match it into a high-gradient booster to damp emittance oscillations.

In the space-charge-dominated regime, i.e., when the space-charge collective force is largely dominant over the emittance pressure, mismatches between the space-charge correlated forces and the external RF focusing gradient produce slice envelope oscillations that cause normalized emittance oscillations, also referred to as plasma oscillations.[2] It has been shown [48] that to conveniently damp emittance oscillations the beam has to be injected into the booster with a laminar envelope waist and the booster accelerating gradient has to be properly matched to the beam size, energy, and peak current, according to

$$\gamma' = \frac{2}{\sigma}\sqrt{\frac{I_\mathrm{p}}{2I_\mathrm{A}\gamma}}\,, \tag{38}$$

with $\gamma' \approx 2E_\mathrm{acc}$, where E_acc is the accelerating field. The matching condition guarantees emittance oscillation damping, preserving beam laminarity during acceleration. The final value of the emittance, however, is strongly dependent on the phase of the plasma oscillation at the entrance of the booster, whose typical behaviour is shown in Fig. 6, as computed by numerical simulations with PARMELA code [49] for different initial electron pulse shapes, i.e., flat top with different rise time from a pure cylindrical

[2]For further detailed studies refer to Ref. [47].

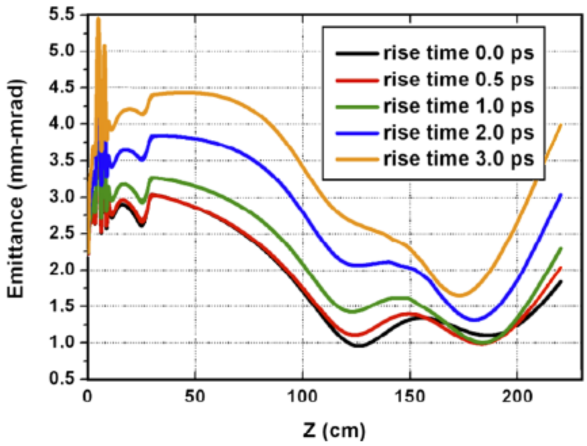


Fig. 6: Normalized r.m.s. emittance oscillations in the drift downstream the RF gun as computed by PARMELA, for different initial electron pulse rise times. Gun length, 15 cm, solenoid length 20 cm centred at z = 20 cm [52].

bunch (0 ps rise time) to a quasi-Gaussian distribution (3 ps rise time). The emittance minimum decreases for shorter rise times because of the reduced non-linear transverse space-charge effects in cylindrical-like bunch charge distributions [50]. In addition, an emittance oscillation appears in the drift downstream of the RF gun, showing a double emittance minimum [51]. The relative emittance maximum disappears at longer rise times and becomes a 'knee' in a quasi-Gaussian distribution (yellow curve in Fig. 6). Emittance oscillations of this kind are produced by a beating between head and tail plasma frequencies caused by correlated chromatic effects in the solenoid. The working point matching condition, suitable for damping emittance oscillations, requires the emittance to be a local maximum and the envelope to be a waist at the entrance to the booster. In this way, the second emittance minimum can be shifted at higher energies and frozen at the smallest value, taking advantage of the additional emittance compensation occurring in the booster. The waist size is related to the strength of the RF fields and the peak current: RF focusing aligns the slices, resulting in a smaller projected emittance, and acceleration damps the emittance oscillations.

Experimental evidence of emittance oscillations in the drift before the booster has been obtained at the SPARC high-brightness photo-injector [52] by using an emittance meter [53] to measure the evolution of the beam transverse phase space in the drift downstream of the RF gun. The behaviour of both projected normalized emittance and envelope along the longitudinal co-ordinate z is reported in Fig. 7. The projected emittance evolution along z shows the expected double minimum oscillation, which is crucial in achieving minimum emittance in high-brightness photo-injectors.

4.4 Longitudinal compression

Space-charge effects at low energy prevent the generation of short, subpicosecond, electron bunches, with a significant amount of charge (tens of picocoulombs) directly from the electron source, leading to emittance degradation and bunch elongation within a few centimetres downstream of the cathode. Bunch compression is therefore necessary to shorten the electron pulse, achieving a high peak current, of the order of kiloamps. Either magnetic or RF-based compression methods can be used for this purpose.

In magnetic compressors, a bunch with a time-energy correlation (or chirp) is driven along an energy-dependent path length by a dispersive, non-isochronous beam transport section [54]. While this scheme has been proved successful in increasing the beam current at high energies, the emittance increase due to coherent synchrotron radiation in bending magnets can be dramatic. As an alternative, the compression scheme used at SPARC_LAB exploits the interaction with the electromagnetic fields of an accelerating cavity. Based on RF compression, it uses rectilinear trajectories, avoiding the degradation

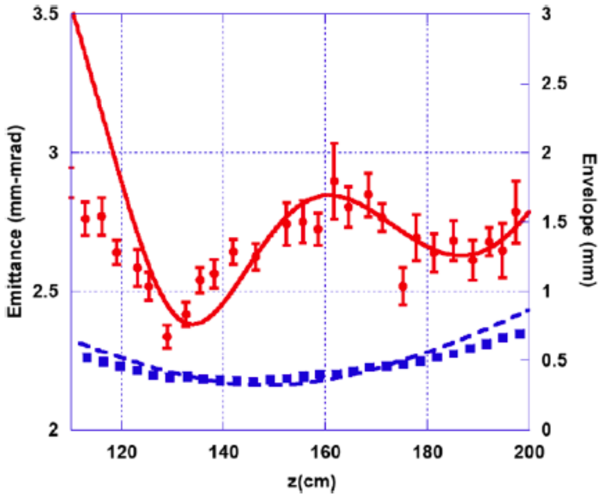


Fig. 7: Normalized emittance and envelope (r.m.s. values) evolution from the cathode up to the beam line end, as computed by PARMELA (red and blue curves), compared with measurements (red dots and blue squares) taken in the emittance meter range (up to 2 m from the cathode). Beam measured parameters: 500 pC, 5 ps FWHM (quasi-flat top pulse shape), 1.5 ps rise time, 5 MeV.

due to coherent synchrotron radiation suffered by the beam going through bending trajectories. In addition, working at relatively low energy [55], of the order of megaelectronvolts, it can be integrated into the emittance compensation process [46] illustrated in the previous section. This scheme is commonly known as *velocity bunching* [56].

The longitudinal phase space rotation in the velocity-bunching process is based on a correlated time-velocity chirp in the electron bunch, causing electrons in the bunch tail to be faster than electrons in the bunch head. The correlated chirp induces a longitudinal phase space rotation in the travelling RF wave potential (longitudinal focusing), accelerating the beam inside a long multicell RF structure, as depicted in Fig. 8 (right panel).[3] Thus, simultaneously, an of-crest energy chirp is applied to the injected beam. Subrelativistic electrons injected into a travelling wave cavity at zero crossing field phase move more slowly than the RF wave. Thus, the beam slips back to phases, towards $-90°$, as shown in Fig. 8 (left-hand panel), where the field is accelerating, and it is chirped and compressed: compression and acceleration take place at the same time within the same accelerating section, i.e., the first one following the RF gun. The velocity-bunching technique is characterized by longitudinal and transverse phase space distortions, leading to asymmetric current profiles and higher final projected emittances, which can, however, be minimized by keeping the transverse beam size under control through solenoidal magnetic fields in the region where the bunch is undergoing compression and the electron density is increasing [57]. For this reason, the typical layout for a high-brightness RF photo-injector, shown in Fig. 2, presents solenoid coils embedding the first two accelerating sections. As shown in Fig. 9, the effect on emittance compensation produced by the solenoids is clearly visible in the simulation (right plot: curve c).

5 Conclusions

High-brightness electron beams can be achieved in RF photo-injectors by means of RF guns, equipped with laser-driven photocathodes, followed by booster sections. An emittance compensation scheme [46] based on a focusing solenoid at the exit of the RF gun can be used in photo-injectors to control emittance increase due to space-charge effects. In addition, by properly matching the transverse phase space of the electron beam with the downstream accelerating sections (booster), it is possible to control the transverse

[3] General Particle Tracer is a 3D code to study charged-particle dynamics in electromagnetic fields.

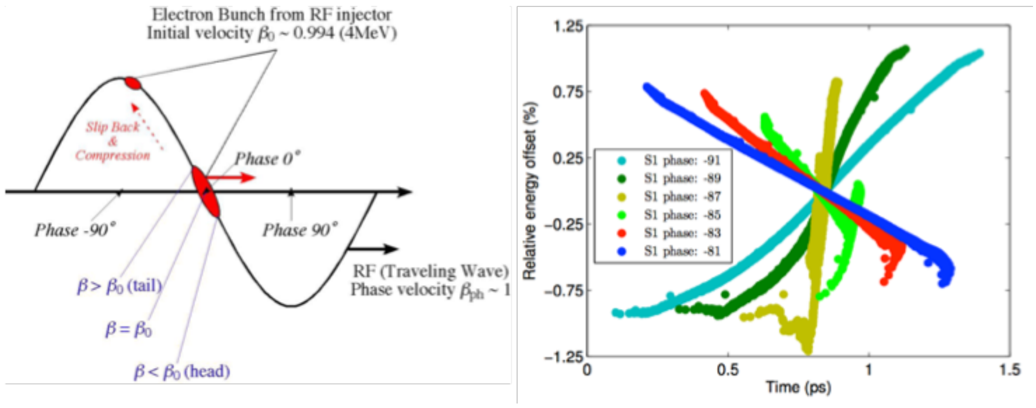


Fig. 8: Left-hand side: longitudinal electric field in the first travelling wave section as a function of RF phase, showing slippage of the beam, injected at the 0 crossing phase, towards $-90°$. Right-hand side: General Particle Tracer simulation of the longitudinal phase space. The maximum bunch compression, for this case, occurs at around $-87°$ (olive green curve).

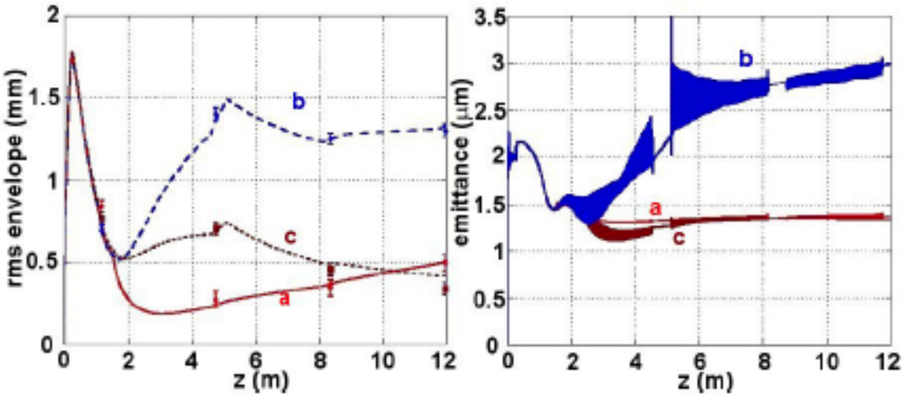


Fig. 9: Left-hand side: measured envelopes and PARMELA simulations. Right-hand side: emittance evolution along the linac (PARMELA simulations); (a) no compression, (b) compression with solenoids off along the first travelling wave section, (c) same compression with solenoids set to 450 G [57].

emittance oscillations during the acceleration. Under conditions of invariant envelope and proper phasing of space-charge oscillations [52], the final emittance is almost compensated down to the intrinsic emittance value given by cathode emission, with an expected emittance scaling such as $\varepsilon_{\mathrm{n}} \sim \sigma_{\mathrm{cathode}} \sim \sqrt{Q}$, where $\sigma_{\mathrm{cathode}}$ is the hitting laser spot size on the photocathode, and Q is the extracted electron charge. A compression stage can occur to shorten the beam length so to achieve the required large peak current. The so-called velocity-bunching method [55] has opened up a new possibility of compressing the beam inside an RF structure and, if integrated in the emittance compensation process [48, 56], can provide the desired bunch current values with the advantage of compactness of the machine and absence of the coherent synchrotron radiation effects present in a magnetic compressor [58–60]. It is interesting to note that a shortened beam length also enables the energy spread dilution due to RF curvature degradation to be contained; indeed the energy spread depends on the bunch length and the accelerating frequency as $\Delta\gamma/\gamma \approx 2((\pi f_{\mathrm{RF}}\sigma_z)/c)^2$, where an on-crest operation, in full relativistic conditions, has been considered.

The electron source is one of the key components, since the brightness generated at the electron source represents the ultimate achievable value. The final application must drive the choice of both electron source and injector, since stability and reliability are common issues. Photo-injectors have the advantage of producing a variety of bunch trains that can be tailored to the needs of a particular machine by properly shaping the laser pulses. Photocathodes (with respect to thermionic sources) can easily produce peak current densities of the order of kiloamps per centimetre squared, and high current densities at the cathode are needed to minimize the transverse emittance. Conversely, DC photocathode electron guns are the best solution for high-average power beams.

Acknowledgements

The material for this lecture has been liberally taken from talks, papers, lectures, proceedings, and notes from a large number of people, whom I acknowledge here. References are listed in the text: S. Di Mitri, D.H. Dowell, D. Filippetto, P. Musumeci, F. Sannibale. In addition, I wish to thank M. Ferrario, A. Cianchi, and A. Mostacci for helpful discussions, suggestions, and revision.

References

[1] C. Bostedt *et al.*, *Rev. Mod. Phys.* **88** (2016) 015007, https://doi.org/10.1103/RevModPhys.88.015007.

[2] V. Ayvazyan *et al.*, *Phys. Rev. Lett.* **88** (2002) 104802, https://doi.org/10.1103/PhysRevLett.88.104802.

[3] W. Ackermann *et al.*, *Nat. Photonics* **1** (2007) 336, https://doi.org/10.1038/nphoton.2007.76.

[4] C.J. Bocchetta *et al.*, Conceptual design report for the FERMI@Elettra project, ST/F-TN-07/12 (2007).

[5] M. Tigner, *Il Nuovo Cimento* **37** (1965) 1228, https://doi.org/10.1007/BF02773204.

[6] `http://www.eli-np.ro/documents/ELI-NP-WhiteBook.pdf`, last accessed December 06th 2017.

[7] P.A. Walker *et al.*, *J. Phys. Conf. Ser.* **874** (2017) 012029, https://doi.org/10.1088/1742-6596/874/1/012029.

[8] C. Lejeune and J. Aubert, *Adv. Electron. Electron Phys., Suppl. A* **13** (1980) 159.

[9] A.W. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 2013), p. 255, https://doi.org/10.1142/8543.

[10] C.A. Brau, in *The Physics and Applications of High Brightness Electron Beams*, Eds. J. Rosenzweig *et al.* (World Scientific, Singapore, 2003), p. 20, https://doi.org/10.1142/9789812705235_0002.

[11] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley-VCH, Weinheim, 2008), p. 61, https://doi.org/10.1002/9783527622047.

[12] Shyh-Yuan Lee, *Accelerator Physics* (World Scientific, Singapore, 2011), p. 419.

[13] J. Clarke, *The Science and Technology of Undulators and Wigglers* (Oxford University Press, Oxford), p. 73.

[14] R. Bonifacio *et al.*, *Opt. Commun.* **50** (1984) 373, https://doi.org/10.1016/0030-4018(84)90105-6.

[15] J.S. Fraser *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **250** (1986) 71, https://doi.org/10.1016/0168-9002(86)90862-4.

[16] F. Villa *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **829** (2016) 446, https://doi.org/10.1016/j.nima.2016.01.010.

[17] P. Chen *et al.*, *Phys. Rev. Lett.* **54** (1985) 693, https://doi.org/10.1103/PhysRevLett.54.693.

[18] F. Giorgianni *et al.*, *Appl. Sci.* **6** (2016) 2, https://doi.org/10.3390/app6020056.

[19] V. Petrillo *et al.*, *Phys. Rev. Lett.* **111** (2013) 114802, https://doi.org/10.1103/PhysRevLett.111.114802.

[20] A.A. Lutman *et al.*, *Phys. Rev. Lett.* **110** (2013) 134801, https://doi.org/10.1103/PhysRevLett.110.134801.

[21] G. De Ninno *et al.*, *Phys. Rev. Lett.* **110** (2013) 064801, https://doi.org/10.1103/PhysRevLett.110.064801.

[22] W. Anders *et al.*, in *An Engineering Guide to Photo-injectors*, Eds. D.H. Dowell and T. Rao, p. 271, CreateSpace Independent Publishing Platform (April 2, 2013).

[23] B. Von Borries and E. Rushka, *Z. Tech. Phys.* **20** (1939) 225.

[24] O. J. Luiten *et al.*, Ultracold electron sources, Proc. 46th Workshop of the INFN Eloisatron Project, Erice, 2005.

[25] C. Herring and M.H. Nichols, *Rev. Mod. Phys.* **21** (1949) 185, https://doi.org/10.1103/RevModPhys.21.185.

[26] P.G. O'Shea *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **318** (1992) 52, https://doi.org/10.1016/0168-9002(92)91023-3.

[27] K.C.D. Chan *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **318** (1992) 148, https://doi.org/10.1016/0168-9002(92)91041-7.

[28] M. Ferrario *et al.*, *Nucl. Instrum. Methods Phys. Res. B* **309** (2013) 183, https://doi.org/10.1016/j.nimb.2013.03.049.

[29] A. R. Rossi *et al.*, *Physics Procedia* **52** (2014) 90, https://doi.org/10.1016/j.phpro.2014.06.014.

[30] P. Schmüser *et al.*, *Ultraviolet and Soft X-ray Free Electron Lasers* (Springer, Berlin, 2008).

[31] H. Alfvén, *Phys. Rev.* **55** (1939) 425, https://doi.org/10.1103/PhysRev.55.425.

[32] J.D. Lawson, *J. Electron. Control* **5** (1958) 146, https://doi.org/10.1080/00207215808953898.

[33] C. Curatolo *et al.*, *Phys. Rev. Accel. Beams* **20** (2017) 080701, https://doi.org/10.1103/PhysRevAccelBeams.20.080701.

[34] I. Blumenfeld *et al.*, *Phys. Rev. ST Accel. Beams* **13** (2010) 111301, https://doi.org/10.1103/PhysRevSTAB.13.111301.

[35] T. Ishikawa *et al.*, *Nat. Photonics* **6** (2012) 540, https://doi.org/10.1038/nphoton.2012.141.

[36] T. Asaka *et al.*, *Phys. Rev. Accel. Beams* **20** (2017) 080702, https://doi.org/10.1103/PhysRevAccelBeams.20.080702.

[37] S. Di Mitri and M. Cornacchia, *Phys. Rep.* **539** (2014) 1, https://doi.org/10.1016/j.physrep.2014.01.005.

[38] E.L. Murphy and R.H. Good, *Phys. Rev.* **102** (1956) 1464, https://doi.org/10.1103/PhysRev.102.1464.

[39] L.A. Dubridge, *Phys. Rev.* **43** (1933) 0727, https://doi.org/10.1103/PhysRev.43.727.

[40] V. Fusco *et al.*, Spatial autocorrelation for transverse beam quality characterization, Proc. EPAC08, TUPC027, Genova, Italy (2008).

[41] L.W. Nordheim, *Proc. R. Soc. Lond. A* **121** (1928) 626, https://doi.org/10.1098/rspa.1928.0222.

[42] D.H. Dowell *et al.*, *Phys. Rev. ST Accel. Beams* **9** (2006) 063502, https://doi.org/10.1103/PhysRevSTAB.9.063502.

[43] P. Musumeci *et al.*, *Phys. Rev. Lett.* **104** (2010) 084801, https://doi.org/10.1103/PhysRevLett.104.084801.

[44] D. Filippetto *et al.*, *Phys. Rev. ST Accel. Beams* **17** (2014) 024201, https://doi.org/10.1103/PhysRevSTAB.17.024201.

[45] J.D. Jackson *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999).

[46] B.E. Carlsten, *Nucl. Instrum. Methods Phys. Res. A* **285** (1989) 313, https://doi.org/10.1016/0168-9002(89)90472-5.

[47] M. Ferrario, Space charge mitigation, Proc. CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 2016.

[48] L. Serafini and J.B. Rosenzweig, *Phys. Rev. E* **55** (1997) 7565,

https://doi.org/10.1103/PhysRevE.55.7565.

[49] L.M. Young, LANL Report LA-UR-96-1835.

[50] J. Yang *et al.*, *J. Appl. Phys.* **92** (2002) 1608, https://doi.org/10.1063/1.1487457.

[51] M. Ferrario *et al.*, SLAC, Report SLAC-PUB-8400.

[52] M. Ferrario *et al.*, *Phys. Rev. Lett.* **99** (2007) 234801, https://doi.org/10.1103/PhysRevLett.99.234801.

[53] A. Cianchi *et al.*, *Phys. Rev. ST Accel. Beams* **11** (2008) 032801, https://doi.org/10.1103/PhysRevSTAB.11.032801.

[54] S. Di Mitri, Bunch length compressors, Proc. CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 2016.

[55] B. Aune and R.H. Miller, Report SLAC-PUB 2393, 1979.

[56] L. Serafini and M. Ferrario, *AIP Conf. Proc.* **581** (2001) 87, https://doi.org/10.1063/1.1401564.

[57] M. Ferrario *et al.*, *Phys. Rev. Lett.* **104** (2010) 054801, https://doi.org/10.1103/PhysRevLett.104.054801.

[58] P. Emma, Accelerator physics challenges of X-ray FEL SASE sources, Proc. EPAC2002, (Paris, France, 2002).

[59] L.I. Schiff, *Rev. Sci. Instr.* **17** (1946) 6, https://doi.org/10.1063/1.1770395.

[60] M. Dohlus and T. Limberg, Emittance growth due to wake fields on curved bunch trajectories, XVIII International Free Electron Laser Conf., ‘Rome, Italy, 1996.

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Energy Efficiency

E. Jensen
CERN, Geneva, Switzerland

Abstract
Particle accelerators are precious tools, not only for high-energy and nuclear physics, but also for many other areas of science, medicine, and security. They do, however, have a substantial energy consumption, which makes it important to design them to make best use of the consumed energy. Maximizing the energy efficiency of an accelerator means minimizing not only the energy consumption and the environmental impact, but also allowing for smaller installations. The concepts and technologies developed to improve the energy efficiency of particle accelerators are not limited to accelerators, but will have a significant impact on energy systems generally. These developments include smart power grids and short-term energy storage devices, better energy conversion efficiencies for subsystems, efficient cryogenic systems, and, ultimately, the recovery of unused energy in more valuable forms than low-temperature heat. Energy recovery linacs are the prime example for good efficiency, since they accelerate, use, and decelerate a continuous beam, using the energy recovered during deceleration for acceleration.

Keywords
Energy conversion; exergy; efficiency; beam loading; Sankey diagram.

1 Introduction

1.1 Importance of energy consumption and efficient energy use

Humankind consumes about 160 PWh per year (or, on average, 18 TW or 2.4 kW per capita) for transportation, industry, nutrition, and comfort. More than 80% of this energy stems from fossil fuels (petrol, coal, natural gas), which are not abundant. Awareness of the scarcity of these natural resources arose in the 1960s (Club of Rome, 1968) and 1970s ([1], oil crisis 1973). Environmental concerns have increased since, relating global warming to greenhouse gas emissions and weather phenomena like El Niño to the incineration of fossil fuels. The 1997 Kyoto Protocol and the 2015 Paris Agreement document worldwide concern and underline the political will of many nations to reduce global warming.

Nuclear fusion and fission promise abundant resources, but have raised other concerns, either of public acceptance or economic feasibility. Renewable resources (wind, solar, hydro, geothermal) are not yet exploited to fully cover demand.

However, as illustrated using the example of the USA shown in Fig. 1, a large fraction of consumed energy is 'wasted' or 'rejected' rather than used for its purpose, and the term 'energy efficiency' is used here to denote the ratio between the useful output of an energy conversion system and the input. The goal of maximizing energy efficiency is synonymous to minimizing the amount of energy required to provide a requested service or product. Design towards higher energy efficiency will thus not only save energy but also allow a smaller installation, both for the supply of energy and for the disposal of rejected energy (typically cooling).

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.421

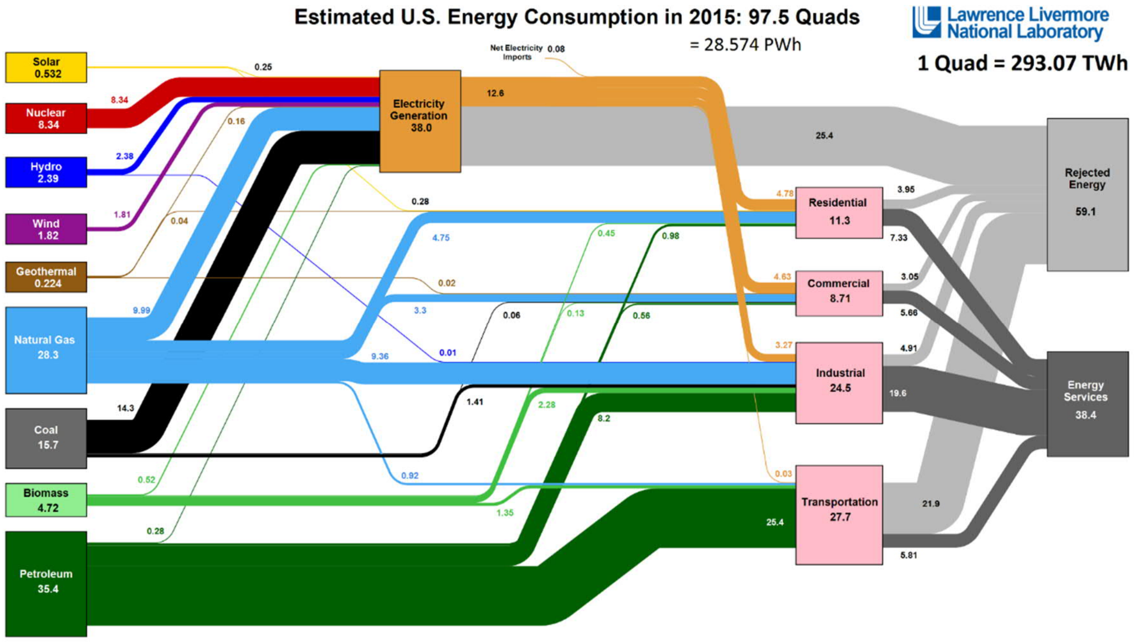


Fig. 1: Sankey diagram illustrating the different sources, forms, and uses of energy, using USA data for 2015 as an example [2].

1.2 Orders of magnitude

The total daily energy consumption of humankind is about 440 TWh or 60 kWh per capita (in Western Europe, 104 kWh). If we look at orders of magnitude for energy provision, storage, and use, we find the following.

1.2.1 1 kWh

This order of magnitude is the daily work delivered by a Tour-de-France race cyclist or one run of laundry in a household washing machine. A household freezer uses about 1 kWh per day, a laundry tumbler will require about 2 to 4 kWh for one run, an average Western European household uses 16 kWh per day per capita. The energy required to heat 1 m^3 of water by 1 K is 1.16 kWh. A window-mounted air conditioner unit running all day uses about 20 kWh. A mass of 37 t in 10 m height has a potential energy of 1 kWh in the Earth's gravitational field; a fully charged conventional lead car battery stores about 1 kWh. A Tesla Model S Li-ion battery holds up to 100 kWh for a reach of up to 540 km [3].

1.2.2 10 MWh

The order of magnitude 10 MWh corresponds to the electricity produced by a wind power station in one average day. Per year, an average household in Western Europe uses about 6 MWh electricity per capita, CERN's Linac4 consumes 10 MWh/day, CERN's PS-Booster or the Swiss Light Source require about 80 MWh/day. 1 MWh of electricity costs of the order of €100 (range €30 to €300). The energy stored in one ITER superconducting toroidal field coil is 11.5 MWh, about four times the energy stored in all 1232 LHC main magnets together at top energy.

1.2.3 10 GWh

The electrical energy produced per day by an average size nuclear reactor block is about 24 GWh. The 'Solar Star' photovoltaic power station in Southern California produces 4.5 GWh per day on average [4], the 'Alta Wind Energy Center' wind farm, with 320 wind turbines, also in California, about 7.3 GWh [5]. CERN's total daily electricity consumption on average is about 3.5 GWh; about half of this is needed for the LHC alone.

As indicated in Fig. 2, some proposals for post-LHC colliders will require RF powers alone of the order of 100 MW, or grid powers of hundreds of megawatts (translating to tens of gigawatt hours per day). It is clear that these large consumptions make the optimization of energy conversion efficiencies compulsory.

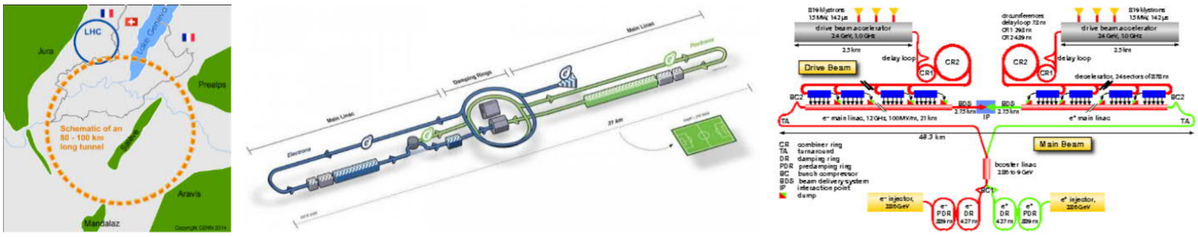


Fig. 2: Some major high-energy physics collider proposals for the post-LHC era. Left-hand side: FCC-ee circular collider, 90–350 GeV, 0.8 GHz continuous wave, total RF power: 110 MW. Centre: ILC (International Linear Collider), 0.5 TeV, 1.3 GHz, total (average) RF power: 88 MW. Right-hand side: CLIC (Compact Linear Collider),3 TeV, 1 GHz, total (average) RF power: 180 MW.

Regarding these examples, it is clear that large particle accelerators are in the size range that matters, both for the acceptance of accelerator projects by the public and because the technology developed to improve efficiency in particle accelerators may well be relevant for other areas of science and technology and thus may have a large societal impact.

Energy storage for this order of magnitude is difficult—the energy stored in all German hydroelectric energy storage plants together is of the order of 40 GWh.

1.2.4 1 TWh

France presently runs 19 nuclear power plants (58 reactor blocks) with a total power of 66 GW or 1.6 TWh per day. Germany uses about 1.6 TWh of electricity per day—humankind about 53 TWh. Energy storage does not really scale to this order of magnitude—the total energy that can be stored in all existing hydroelectric plants probably sums up to about 1 TWh. The annual electricity consumption of CERN is 1.2 TWh, that of the canton Geneva is 2.9 TWh; the whole of Switzerland uses 60 TWh, Germany 584 TWh.

The total solar energy (sunshine) on the Earth's surface is 3,000,000 TWh = 3 EWh every day.

1.3 Definition of energy efficiency

It is clear that it is important to define 'useful' in order to define energy efficiency. If, for example, we consider the purpose of heating an apartment, very clearly the purpose is to keep a certain comfortable temperature, which requires some heating if the outside temperature is lower. Large energy efficiency here certainly has to do with good thermal insulation—perfect insulation would not require any heating at all! One easily finds that all heating power is eventually required only to heat the environment through the apartment's imperfect thermal insulation, and good efficiency directly translates to avoided consumption of primary energy.

If we are looking at a subsystem that converts or transports energy, the definition of 'useful' is again relatively straightforward—the losses here are conversion or transmission losses and the efficiency very clearly is

$$\eta = \frac{\text{useful}}{\text{consumed}} = 1 - \frac{\text{lost}}{\text{consumed}} = \frac{1}{1+\text{lost}/\text{useful}} \ .$$

Also, note that the consumed energy is given by

$$\frac{\text{lost}}{1-\eta} = \frac{\text{useful}}{\eta} \ ,$$

and so both lost and consumed energy are reduced when η is maximized (indicated by the height of the blue boxes in Fig. 3, clearly smaller on the right-hand side). This, of course, means that the size of the installation and the cooling and ventilation systems can be designed to be smaller.

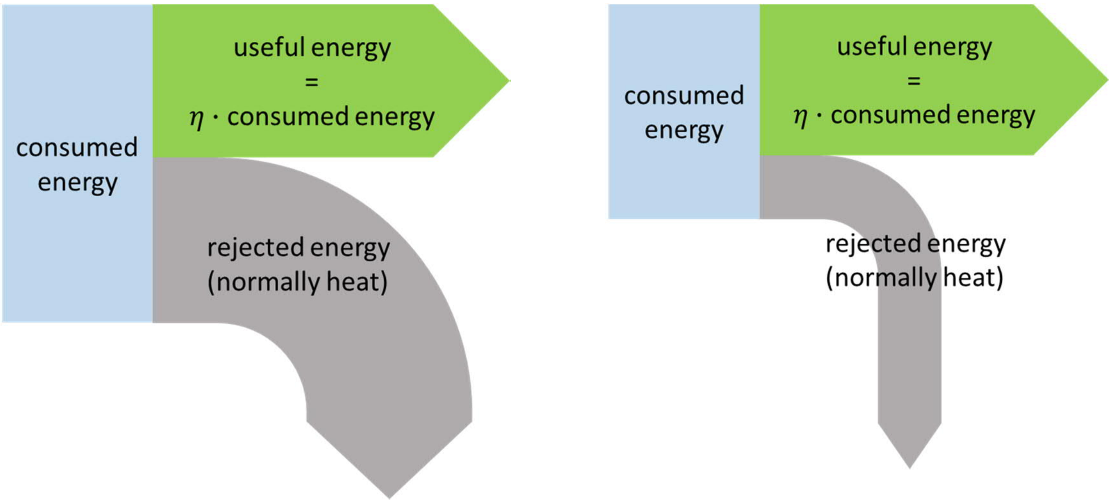


Fig. 3: Sankey diagram of the definition of efficiency as a ratio of useful energy to consumed energy. Note that the illustration of small efficiency (left-hand side) and large efficiency (right-hand side) is scaled to show identical 'useful energy' for the two cases (green), and that a larger efficiency also means a smaller installation for both the power supply (blue) and the cooling and ventilation (grey).

The fact that larger efficiency leads to less consumed energy allows us to coin the term 'avoided consumption', which can, in fact, be considered as a resource gained by increasing efficiency. Looking again at the example of Fig. 1, comparing the 'rejected energy' with the resources on the left, one realizes that this new resource can be of significant size.

2 Power flow in an accelerator

It is useful to look at all components of a particle accelerator that contribute to the energy conversion efficiency to identify at which point improvements pay off the most. Typically, one starts from the electricity grid at high voltage (400 kV) or medium voltage (18 kV) with large transformers. These feed so-called power converters that feed magnets, cryogenic installations, RF power generators and amplifiers, particle detectors, computers, and ancillary systems.

In the example given in Fig. 4, the RF system, magnet system, and scientific instruments share the consumed energy in almost equal parts, but this should not be generalized. A linac will certainly use more energy for RF than for magnets; the LHC uses the largest part for the cryogenic system to allow using only very little for the superconducting magnets themselves. In all cases, however, the energy of the beam is an important intermediate quantity that allows efficiencies of the accelerator subsystems to be quantified. In subsequent sections, we will look at these subsystems' efficiencies in more detail.

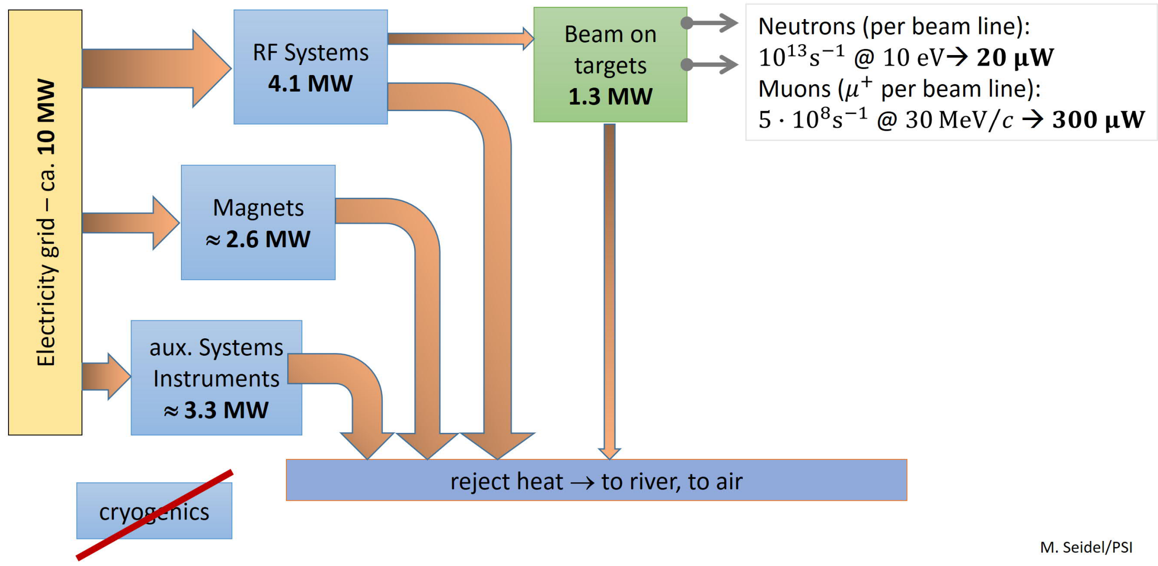
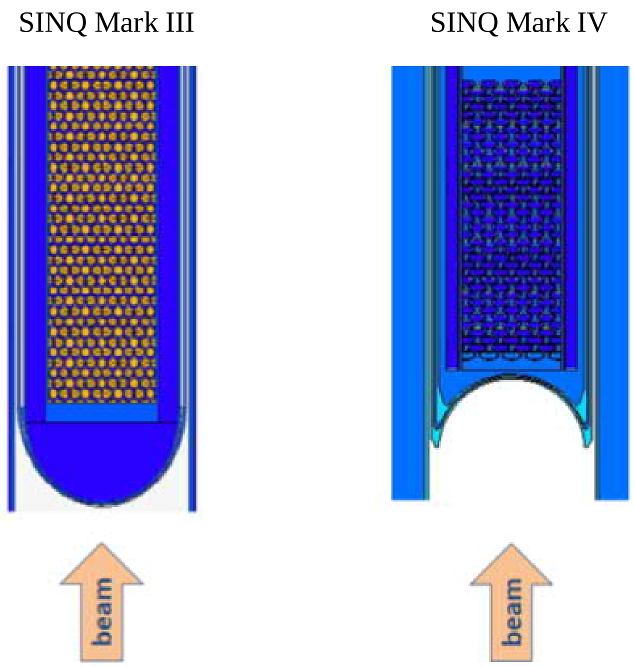


Fig. 4: Typical power flow in an accelerator (example PSI [6])

Conversely, it is equally important to assure that optimum use is made of the beam energy! The luminosity upgrade project of the LHC at CERN, 'HL-LHC', is an example aimed at increasing the luminosity (the number of physics events) by a factor of ten, while the beam intensities are increased by (only) a factor of two. This is achieved by squeezing the two proton beams even further (decreasing β^* from nominally 65 cm to below 20 cm). To avoid spurious collisions that would then result from head-on collisions requires, however, increasing the crossing angle, which can only be achieved by enlarging the aperture of the final focus magnets. A larger crossing angle with reduced β^* would, however, lead to incomplete overlap of the colliding bunches and thus a geometric luminosity reduction. Crab cavities will enable correction for this geometric aberration, allowing again for the same luminosity as in head-on collisions.

Another example for optimizing the use of the beam energy is presented in Fig. 5, which indicates a series of steps that led to the more efficient use of the incoming proton beam in a spallation neutron target, increasing its neutron yield by a remarkable factor of 1.42 [7]. This was achieved by using zirconium cladding, arranging the target rods to be closer together, the usage of lead reflectors, and reshaping the calotte for the incoming beam from convex to concave. Note that a yield increase by a factor of 1.42 is equivalent to a reduction of the used energy by the same factor!



Improvements

Zr cladding instead of steel	+12%
Compaction of rod bundle	+5%
Pb reflectors	+10%
Inverted entrance calotte	+10%
Total gain	+42.3%

Fig. 5: Improvements of a spallation neutron target for better neutron yield [7]

It is often difficult to quantify what exactly the 'best use of the beam energy' is, since it is, a priori, not known how much energy consumption is required for a physics discovery, but certainly one can make comparative studies, as in the two examples above, at least if the physical processes involved are understood.

In some cases, where a large beam energy is required continuously but the beam structure is not 'destroyed' in the process (for example, if electrons create light in a wiggler or undulator, or if electrons interact in an electron–ion collider), the 'best use' of the beam energy could be its recovery, for example, in an energy recovery linac (see Section 8 below).

3 Is everything lost?

Looking at the typical numbers in the previous examples it seems that—apart maybe from a few microwatts in neutrons or muons—virtually all energy is converted to waste heat. The obvious question arising is, 'Can we recover the heat in a more valuable form of energy?'

According to the second law of thermodynamics, one can distinguish forms of energy of different 'qualities': heat, for example, is a form of energy, which has value only in the presence of a temperature difference. Otherwise, in so-called thermodynamic equilibrium, it cannot be converted into another form of energy—in particular, it cannot do work. Other forms of energy can do work directly (potential, electric, to some degree kinetic) and thus have a higher value. Energy that can do work is often also referred to as *exergy* [8].

Heat in the presence of a temperature difference can do some work, and in spite of the former statement on heat as a form of energy of lower 'quality', it is practically involved in most energy conversion processes, from the steam engine via most power plants to nuclear reactors. The efficiency of the heat engine, the process of converting heat to work, is, however, strictly limited to the Carnot efficiency [9], which is given as $\eta_{\max} = 1 - T_\mathrm{C}/T_\mathrm{H}$, where T_C and T_H are the cold and hot temperature in kelvins, respectively. This is illustrated on the left-hand side of Fig. 6 in a Sankey diagram, while the Carnot efficiency of a heat engine operating in an ambient temperature of 20 °C is shown on the right-hand side. It is clear that that this efficiency becomes zero with vanishing temperature difference and reaches only 50% when $T_H = 2\,T_C$, which means that, with an ambient temperature of $T_C = 20\,°C = 293\,K$, the efficiency reaches only 25.4% for $T_H = 100\,°C = 393\,K$. Substantial efficiencies can be reached only at much larger ΔT.

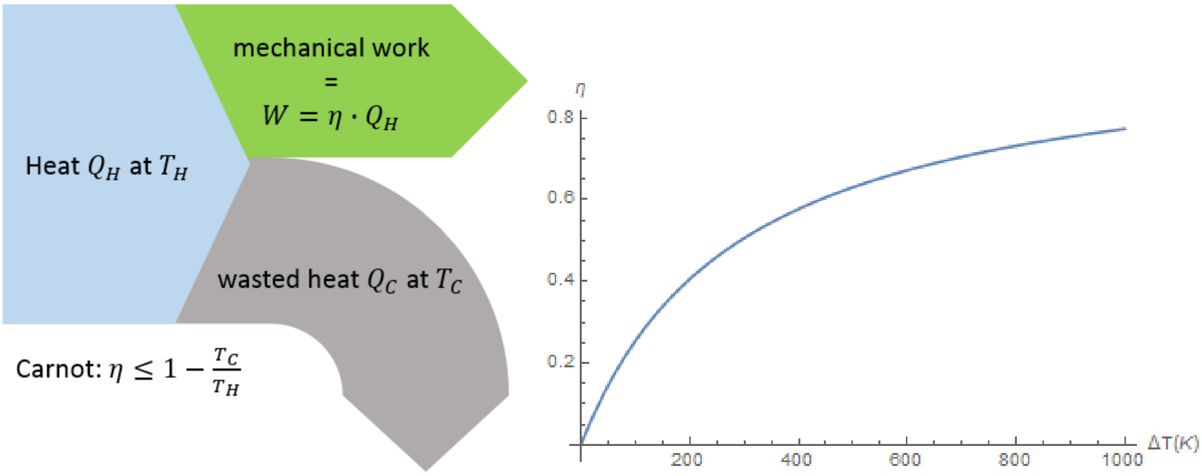


Fig. 6: Left-hand side: Carnot efficiency as the limit for the conversion of heat to work. Right-hand side: Carnot efficiency of a heat engine as a function of ΔT for $T_C = 20\,°C$.

From this consideration, it should be clear that the recovery of waste heat to exergy is not efficient and works better at high temperature. Please consider that the Carnot efficiency is the theoretical limit, while measured efficiencies are always lower because of friction and other imperfections.

Before directly rejecting waste heat into the environment, one may of course consider injecting it into a heating plant, but this should be considered only for the unavoidable remaining waste heat after all other processes have been optimized to minimize consumption.

4 Optimizing magnets

In principle, no energy is consumed to generate a DC magnetic field—a prominent example for the use of permanent magnets in accelerators is the Fermilab Recycler storage ring [10]. Permanent magnets, however, are limited in both field strength (neodymium magnets reach ≈ 1.4 T) and homogeneous field region. The downside of permanent magnets is, of course, the lack of tunability. An example showing how permanent magnets can also be made tuneable is illustrated in Fig. 7, which shows a permanent quadrupole studied for CLIC, which is tunable in a range from 15 T/m to 60 T/m by displacing parts of the magnetic circuit mechanically by a stroke of 64 mm.

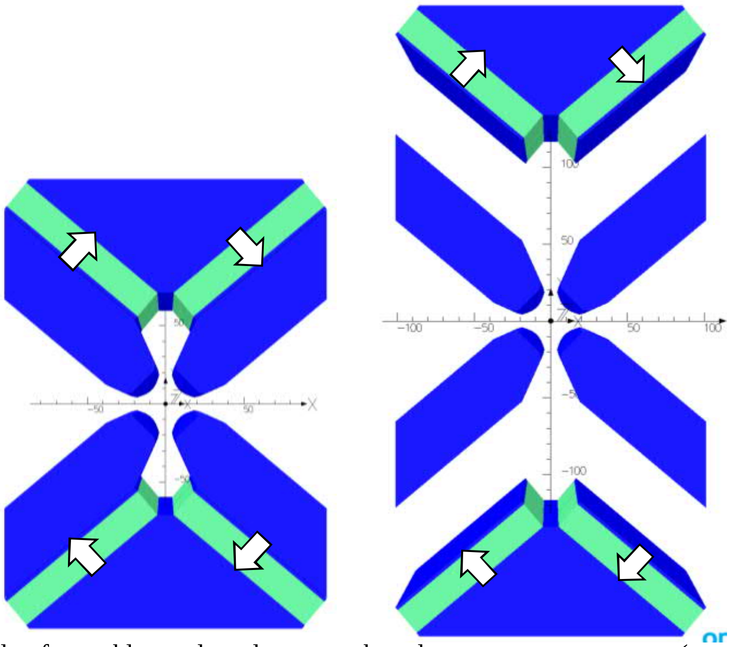


Fig. 7: Principle of tuneable quadrupole magnet based on permanent magnets (green) [11]

For the LHC, the main dipole magnets use NbTi superconductors up to 11.6 kA operated at 1.9 K to reach fields of the order of 8.33 T to guide the 7 TeV proton beams on the LHC curvature [12]. Thanks to superconductivity, the ohmic losses in the magnet coils are zero but, of course, energy is consumed to establish and maintain the cryogenic conditions, treated in Section 7 below.

5 RF power generation

Figure 8 shows the Sankey diagram for the FCC-ee as an example with optimistic but not untypical numbers for the energy conversion efficiencies involved. It should be noted that (i) the most significant conversion efficiency in the example is the generation of RF power and (ii) a large amount of energy is required to operate the cryogenic plant required to keep the cavities superconducting, which will be treated in Section 7 below.

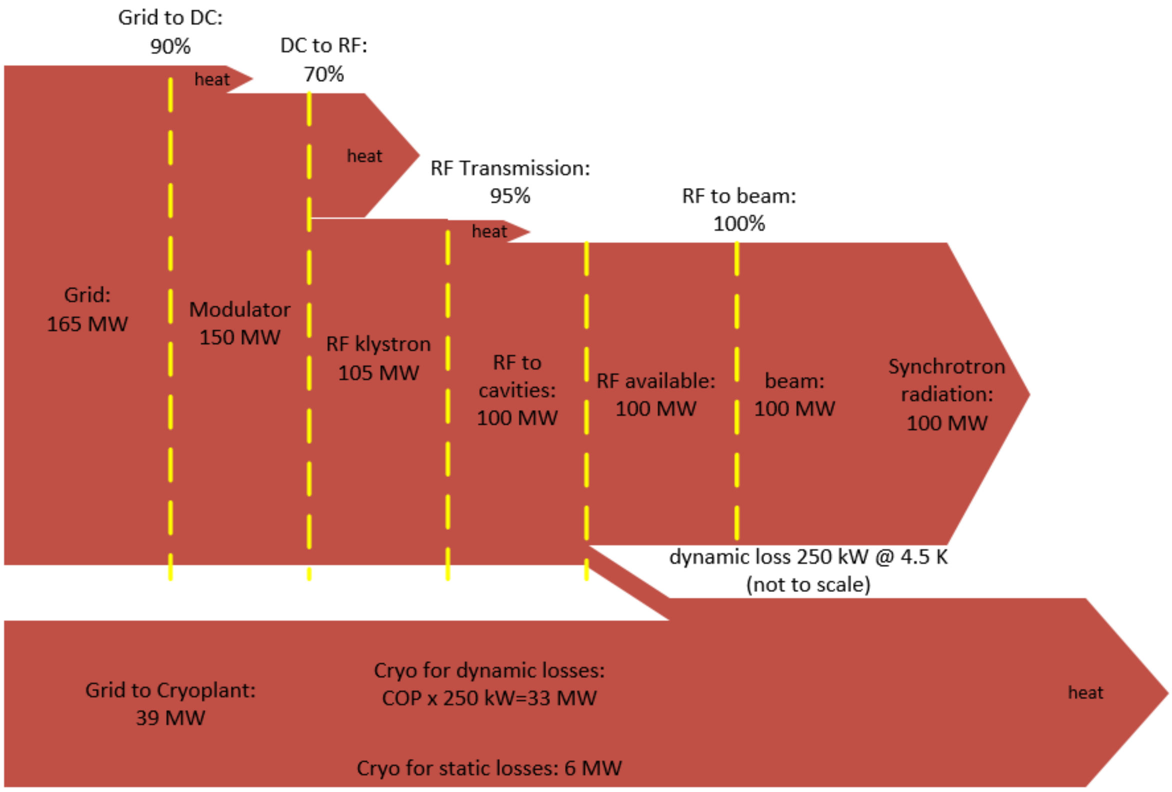


Fig. 8: Sankey diagram indicating energy conversion efficiencies for the FCC-ee circular collider RF systems and cavities. The generation of RF from DC is a significant contributor to the overall efficiency.

Active elements used for RF power generation are summarized in Table 1. Note that the frequencies, powers, and efficiencies given are merely meant to indicate typical orders of magnitude, while actual cases may significantly deviate. In particular, for solid-state power amplifiers, the attainable efficiency depends strongly on the frequency range and on the technology used to fabricate the active elements. The abbreviation IOT stands for ‘inductive output tube’, which is a hybrid between a tetrode and a klystron.

Table 1: Typical frequency, average power, and efficiency ranges of active elements for RF power generation

	Tetrodes	**IOTs**	**Klystrons**	**Solid-state power amplifier**	**Magnetrons**
f range	DC to 0.4 GHz	0.2–0.5 GHz	0.3–20 GHz	DC to 20 GHz	GHz range
P range	1 MW	1 MW	1.5 MW	1 kW	<1 MW
Typical η	85–90% (class C)	70%	50%	60%	90%
Remarks	Broadcast technology, widely discontinued		New ideas promise significant increase	Requires P combination of thousands	Oscillator, not amplifier!

5.1 The impact of improving the RF power generation efficiency

I will use the example of the FCC-ee study, sketched in Fig. 8, to illustrate the impact that an improvement of the klystron efficiency from 70% to 80% would have. Let us assume that the indicated 105 MW are given; in this case, the required power from the grid would decrease from 165 MW to 144 MW. All electrical installation on the primary side could thus be dimensioned 12.5% smaller. With an assumed annual operation period of 5000 h, the annual energy consumption would reduce by 105 GWh. With an assumed electricity cost of €50/MWh, this would reduce the annual electricity bill by €5.2 million, savings that could co-finance the R&D. The waste heat rejected to the environment would be reduced by 35% from 60 MW to 35 MW. This would also mean that the cooling and ventilation systems could be designed to be 35% smaller.

All of these savings are significant and make it worthwhile to invest in ideas that could lead to increased RF power generation efficiency and the development of better klystrons.

5.2 New concepts to increase klystron efficiencies

Klystrons are widely used in particle accelerators, in both pulsed and continuous wave operation. Conventional klystrons reach efficiencies of up to 70% in saturation and typically around 50% in standard operation. In a standard klystron, a DC voltage of typically 20–200 kV accelerates a continuous electron through a vacuum tube. A small RF voltage applied to the passing beam in an input cavity leads initially to a velocity modulation of the beam, which, after a drift region, leads to an intensity modulation (bunching). This density modulation gives rise to RF components of the electron beam, which can be extracted with an output cavity. In this simplest form of a klystron, consisting of just input cavity and output cavity, a maximum efficiency (extracted output power divided by DC voltage and beam current) of 58% can be reached. Additional passive cavities between the input cavity and the output cavity are often used; they are tuned slightly off the operating frequency or near its harmonics and help in the bunching process to reach larger efficiencies; this is how today's conventional klystrons reach approximately 70%.

Space charge forces limit the efficiency of a high-power klystron. Since short bunches are required for high efficiency and space charge forces counteract the formation of short bunches, this led to the invention of multi-beam klystrons, where the total beam current is divided into n smaller beams, reducing the space charge by a factor of n.

New ideas appeared in 2013, which fundamentally questioned the established methods of maximizing klystron efficiencies. One of these ideas is the core oscillation method, where the space charge is used to help the bunching rather than obstructing it [13]. Figure 9 tries to illustrate the core oscillation principle (bottom) by comparing it with conventional bunching (top). The plots on the left-hand side show the particle phase distribution (vertical axis) while travelling through the klystron, passing the different cavities (vertical grey lines). The particle distribution is sketched using two different colours to distinguish between those particles that arrive in the output cavity in a useful RF phase to contribute to power generation (light brown shading) and those that arrive in the wrong phase, i.e. particles that will actually be accelerated in the output cavity and thus reduce efficiency (grey shading). In conventional klystrons, the number of electrons arriving in the wrong phase is substantial—this leads to the limitation of efficiency.

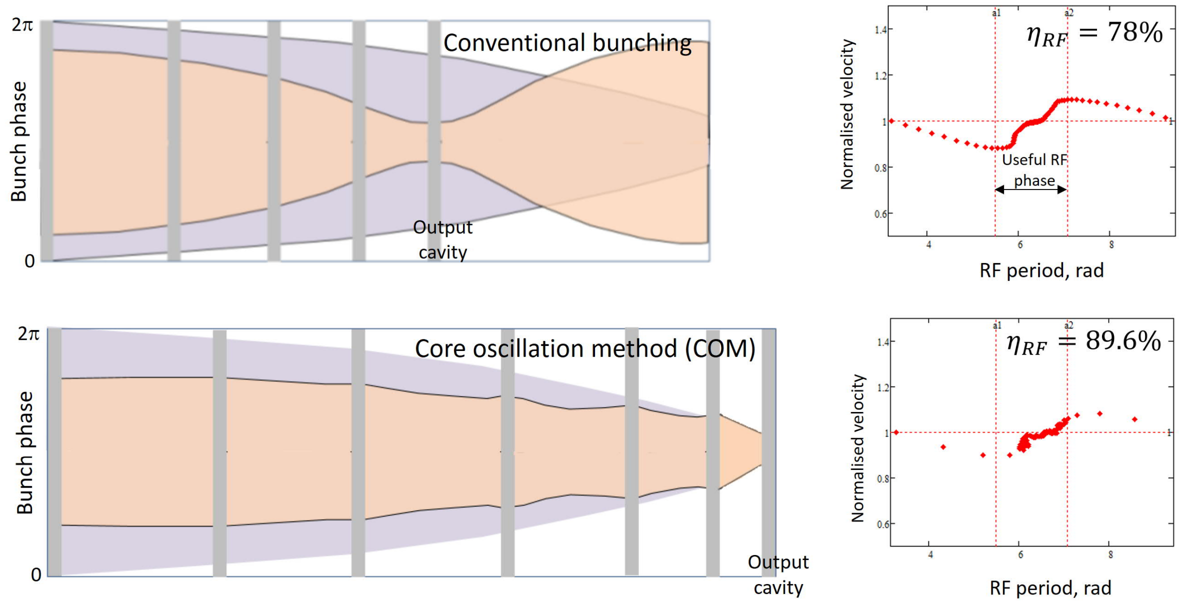


Fig. 9: Conventional bunching (top) compared with core oscillation method (bottom) to reach higher klystron efficiency.

For the core oscillation method, one allows the space charge to drive the forming bunches apart again and puts the next cavity further downstream. Bunches form again, are driven apart again by space charge and redressed again further downstream—this can be repeated several times, leading to 'core oscillation'. While this core oscillation takes place, the electrons that would otherwise arrive in the wrong phase in the output cavity see smooth focusing forces that allow them to drift also into the core of the bunch. This leads to a much more favourable electron distribution in the output cavity and thus significantly larger efficiency.

The disadvantage of the core oscillation method is the need for a much longer device, but other methods are the focus of recent studies to combine high efficiency with reduced length.

6 Conversion of RF power to beam power

Once the RF power is generated and transmitted to the cavity, where the interaction with the charged particle beam takes place, how much of this power can actually be transferred to the beam? How can one optimize this transfer of power to the beam?

In the early days of accelerators, the quantity optimized in the cavity design was the shunt impedance. This makes sense, since the shunt impedance R_{shunt} is defined as $|V_{\mathrm{acc}}|^2 = R_{\mathrm{shunt}} \cdot P$, i.e. maximizing R_{shunt} will result in the largest accelerating voltage given the input power. R_{shunt} can be expressed as the product of a factor R/Q, which depends only on the cavity geometry, and the quality factor Q, which is the number of RF periods, during which the stored energy decreases by a factor 535.5 ($= \mathrm{e}^{2\pi}$) if no RF power is replenished. A typical order of magnitude for the factor R/Q is 200 Ω; a typical order of magnitude for Q is 10^3 to 10^5 for normal-conducting copper cavities and 10^9 to 10^{11} for superconducting niobium cavities.

Note that the power P in this equation is the power lost in R_{shunt} and not the power transferred to the beam!

For the transfer of energy from the RF to the beam, however, optimizing the voltage by optimizing R_{shunt} neglects the effect of the beam current, which, if not small, is important for the energy transfer to the beam and thus for the conversion of RF power to beam power. Referring to Fig. 10, with the simplifying assumption that the cavity is in tune, such that L and C exactly compensate each other and can be neglected. In the absence of the beam ($I_\mathrm{B} = 0$), the generator current through R_shunt will create the accelerating voltage V_acc, as desired.

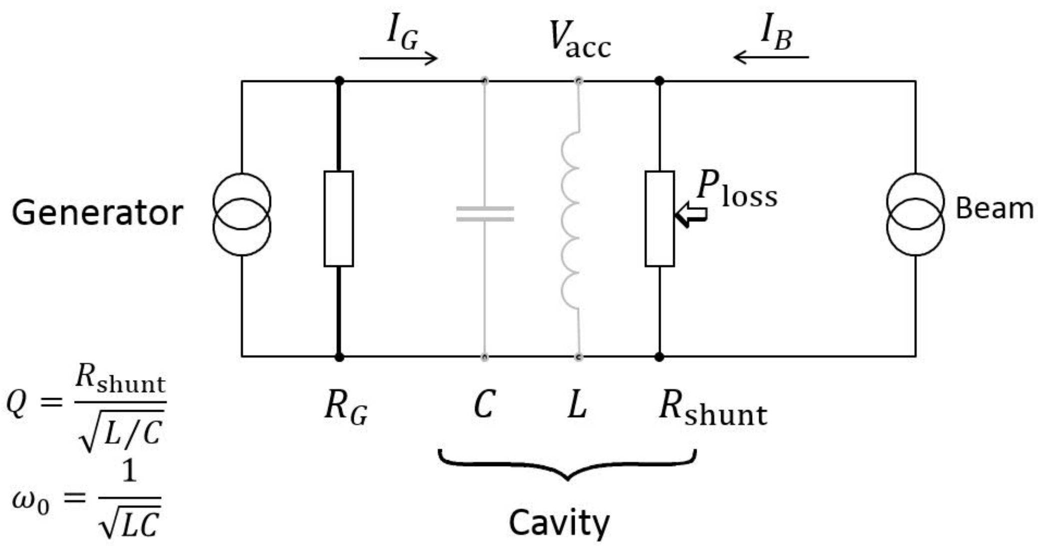


Fig. 10: Equivalent circuit for a single mode in an accelerating cavity

In the presence of the beam ($I_\mathrm{B} \neq 0$), the sum of I_G and I_B flows through R_shunt. When accelerating, with the arrow direction as chosen in Fig. 10, the beam current will, however, be in antiphase to the accelerating voltage ($I_\mathrm{B} < 0$) and thus counteract I_G, which means that the total current through R_shunt and thus V_acc will be reduced. This effect is generally referred to as 'beam loading'.

In a more general formulation, the currents and voltages in the equivalent circuit have arbitrary phases, best described as complex quantities, which also enables studying the cavity off tune and any beam phase. With the arrow direction chosen in Fig. 10, the power transferred to the beam is given as $-\Re\{V_\mathrm{acc} \cdot I_\mathrm{B}^*\}$, where $\Re$ denotes the 'real part' and the asterisk denotes 'complex conjugate'. This is the quantity to be maximized for best efficiency in the transfer of RF energy to beam energy. One can state that, to have a large energy transfer efficiency, the beam loading should be large.

If R_shunt is very large, as is the case for superconducting cavities, it can be neglected entirely in the equivalent circuit, and now the task of optimizing energy transfer is to match R_G to the equivalent beam impedance, $-V_\mathrm{acc}/I_\mathrm{B}$. If this matching is possible, the energy transfer efficiency can reach 100% in this limiting case.

Full beam loading can also be reached in normal-conducting accelerating structures, and was, in fact, reached in CTF3 in the frame of the CLIC study [14]. For the acceleration of the drive beam in a normal-conducting 3 GHz travelling-wave structure, an RF to beam energy transfer efficiency of 95.6% could be experimentally verified. On the downside, maximizing the efficiency reduces the accelerating voltage from the value that is reachable in the unloaded case. For full beam loading, this reduction is roughly a factor of two, as indicated in Fig. 11. This reduction is of no concern for a circular accelerator, but very important for linacs used in linear colliders. In the CLIC concept, full beam loading is used for the drive beam, but only 20% beam loading is chosen to accelerate the main beam, which is a trade-off between efficiency (36%) and a reduced accelerating gradient compared with the unloaded case (90%).

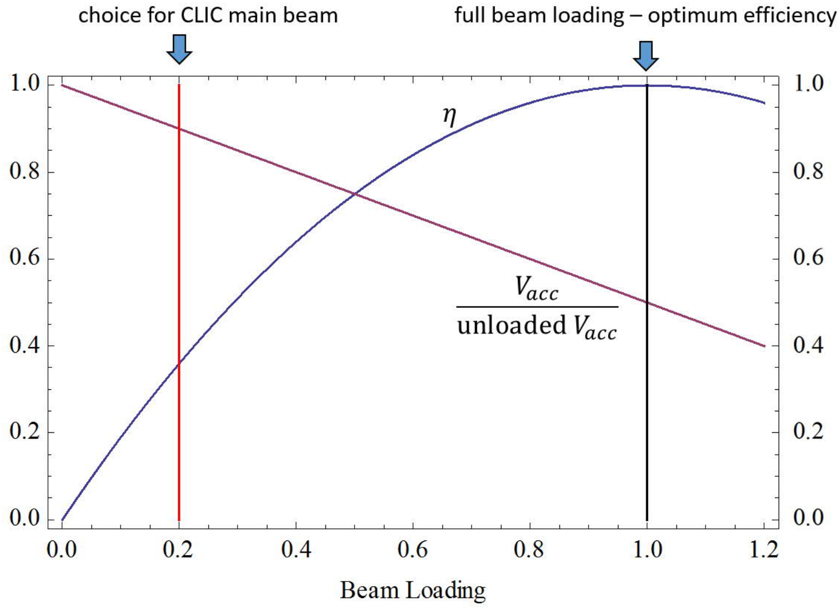


Fig. 11: RF to beam energy transfer efficiency as a function of degree of beam loading. For full beam loading, only about 50% of the unloaded gradient is reached; for partial beam loading, gradient and efficiency can be traded off.

7 Cryogenic system efficiency

The concept of the Carnot efficiency introduced in Section 0 above, limiting the efficiency of a heat engine, is also valid for other engines that transfer energy between heat and ordered motion, such as heat pumps or refrigerators.

Figure 12 illustrates a refrigerator, where mechanical work is done to extract heat at a low temperature—the basic principle of a cryogenic system. At the same time, this illustrates a heat pump, for which Q_H at T_H is the desired heating, extracting energy Q_C at the lower temperature T_C with the help of the mechanical work of the heat pump.

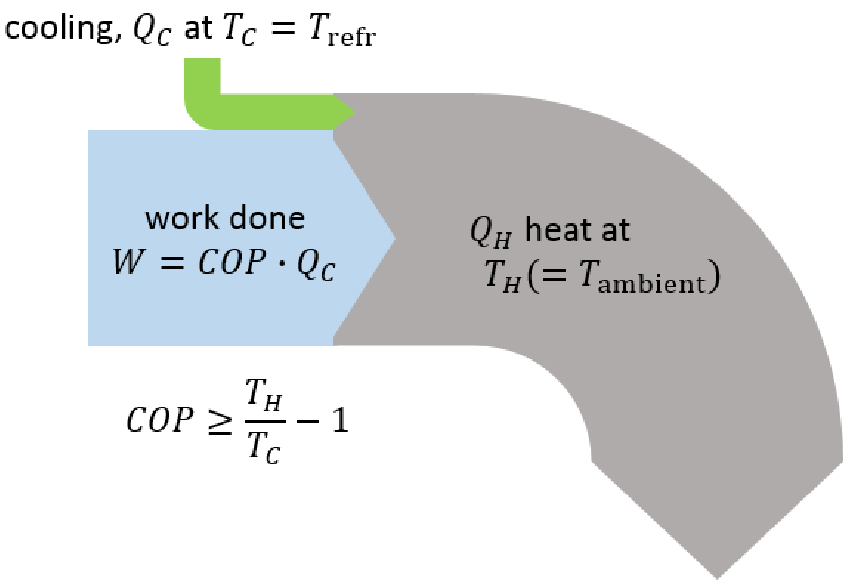


Fig. 12: Reversed Carnot cycle, illustrating the energy flow in a cryogenic system—the goal is to extract heat Q_c at T_c, which requires work $W = \mathrm{COP} \cdot Q_C$.

The established term used to describe the performance of cryogenic plants or heat pumps is the 'coefficient of performance', abbreviated COP. If we use our definition of efficiency as the ratio of 'useful' energy to 'used' energy, and define for a cryogenic plant the cooling power at low temperature, Q_C as the 'useful' energy, we could write $\mathrm{COP} = \eta^{-1}$. For a heat pump, where Q_H is the 'useful' energy, the efficiency is sometimes defined as $\eta = Q_H/W$, neglecting the used incoming heat Q_H, which leads to efficiencies well over 100%, limited by Carnot's principle to $T_C/\Delta T$.

For real technical installations, the COP obtained is much larger than the Carnot limit (Fig. 13). Based on consolidated experience with modern and large CERN cryogenic plants, the technically achieved COP is about 230 at 4.5 K and 930 at 1.8 K, much larger than the Carnot limits of 64.1 and 161.8, respectively (assuming an ambient temperature of 293 K).

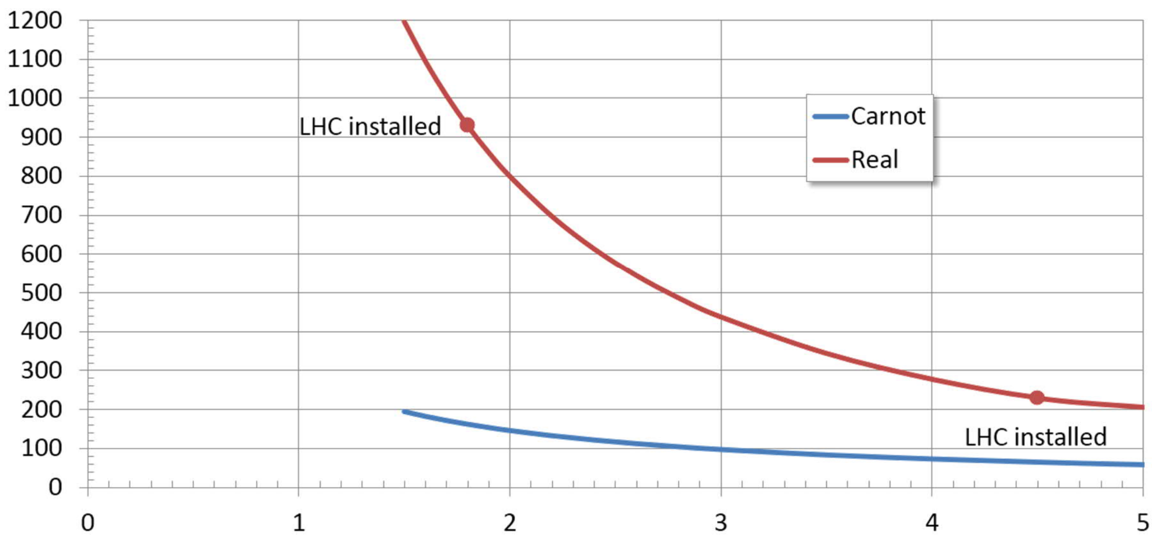


Fig. 13: COP obtained in large cryogenic systems as a function of temperature (P. Lebrun, private communication)

7.1 Optimum operating temperature for a superconducting RF system

When optimizing a superconducting RF system, as, for example, the one used as an example in Fig. 8, it is important to optimize the whole system and not just the subsystems. While superconducting cavities reach lower losses when operated at lower temperatures (see Fig. 14, left-hand side), it is clear that cooling to obtain those lower losses costs more energy, owing to the increased COP. Combining these two effects leads to an optimum operating temperature at which the energy consumption for the cryogenic system is minimized (see Fig. 14, right-hand side) [15].

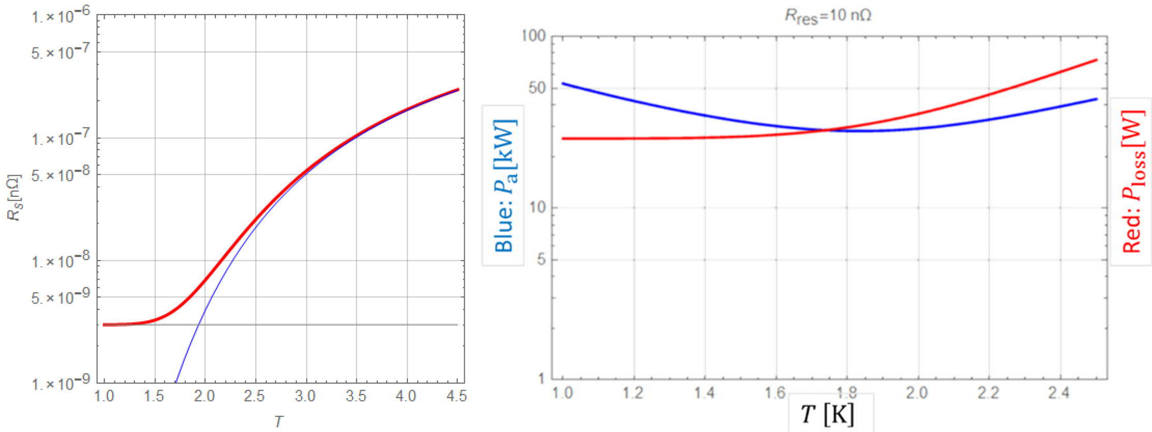


Fig. 14: Left-hand side: surface resistance in a Nb superconducting cavity as a function of temperature, according to the BCS model (blue) and real (red). Right-hand side: with the T dependence of both R_s and COP, an optimum operating temperature results. Red: cavity wall losses in watts. Blue: required cooling power at ambient temperature in kilowatts.

8 Recovering the beam energy

8.1 Beam energies and beam powers in synchrotrons and linacs

When we talk about 'energy recovery', we clearly mean the recovery of energy in a valuable form, preferably as exergy.

There is a substantial difference between synchrotrons and linacs concerning the beam power and beam energy: while a synchrotron is filled with a particle beam just once, which is then often kept for a relatively long period, linacs are constantly fed with a fresh beam which normally has only a single passage through it. In the following, we will illustrate this difference and its impact on the possibility and viability of beam energy recovery based on the examples of the LHC and CLIC.

The beam energy stored in the LHC at nominal parameters (2808 bunches of 1.15×10^{11} protons at 7 TeV) is 362 MJ or 100 kWh (small). This energy is distributed over practically the whole ring or 90 μs. This means that, when the beam is dumped, a 90 μs long pulse with a peak power of 4 TW will impact on the dump once. The recovery of this energy (just once) does not seem worth the effort.

Now consider a CLIC linac as an example: it is designed to operate at a repetition frequency of 50 Hz with pulses of 312 bunches of 3.72×10^9 electrons, accelerated to 1.5 TeV, resulting in 278 kJ per pulse (more than a factor 1000 below the stored energy in the LHC!). But the repetition every 20 ms leads to an average beam power of 14 MW, which, of course, would be very interesting to recover (if it were not used to collide).

Looking at these beam powers and beam energies it is clear that—contrary to a synchrotron—it looks interesting to recover the (large average) beam power of a linac.

8.2 The principle of beam energy recovery

Looking again at the equivalent circuit in Fig. 10, note that the beam on the right-hand side is modelled as an ideal current source, which, with our choice for the arrow direction, consumes power for acceleration if $\Re\{I_{\rm B}\} < 0$. However, the equivalent circuit is equally valid for $\Re\{I_{\rm B}\} > 0$, which would describe the case that the beam delivers power rather than consuming it. This can, in fact, be reached simply by placing the bunches in the decelerating rather than the accelerating phase. Fig. 15 illustrates the principle, combining two equal accelerating structures in series, with the second phased such that the particles are decelerated. Neglecting losses, exactly the same RF power would be generated from the beam in the second accelerating structure that is used in the first structure for acceleration. The beam energy, which reaches a maximum after acceleration in the first structure, would be decelerated to exactly the injection energy after passage through the second.

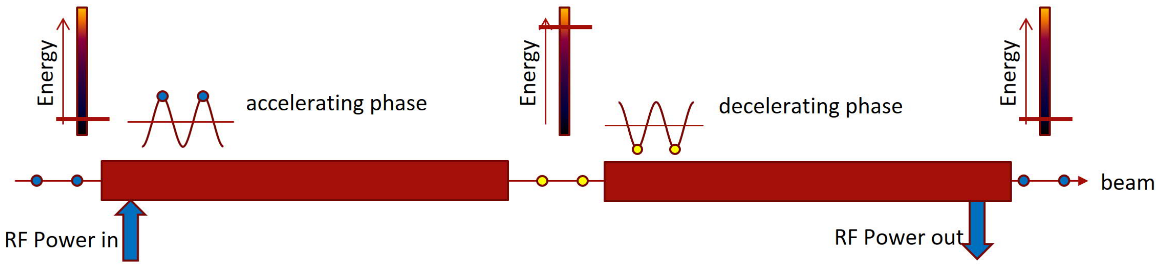


Fig. 15: Two equal accelerating structures with the same bunched beam passing through them. Depending on the phase of the beam relative to the RF in the cavity, the beam will be accelerated or decelerated.

This principle can be extended by adding arcs and thus feeding the same beam through the same accelerating structure again. This is the principle of an energy recovery linac, invented by M. Tigner in 1965 [16] and illustrated in Fig. 16.

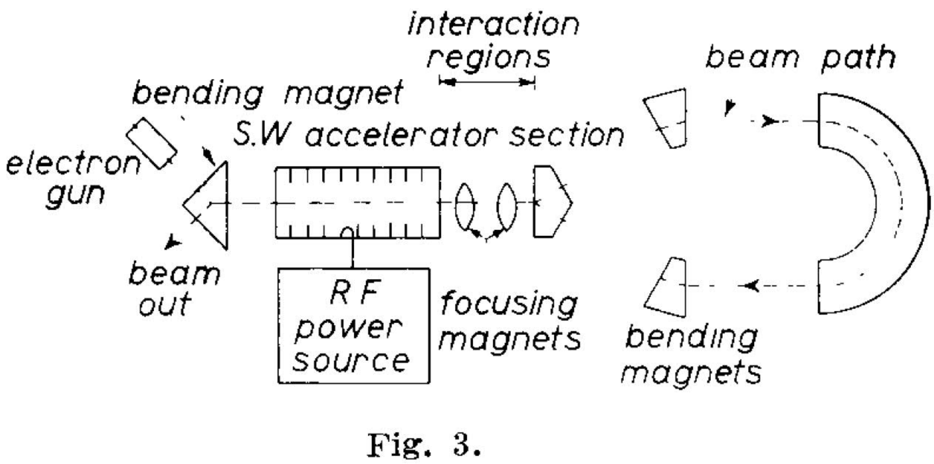


Fig. 16: The illustration in the original paper by Tigner [16], where the spent beam is passing through the same cavity again for deceleration to recover the beam energy (reproduced with permission from the publisher).

References

[1] D.H. Meadows et al. The Limits to Growth (Universe Books, New York, 1972).

[2] https://flowcharts.llnl.gov/, last accessed January 23, 2018. Credits to Lawrence Livermore National Laboratory and the Department of Energy

[3] https://en.wikipedia.org/wiki/Tesla_Model_S, last accessed January 23, 2018

[4] https://us.sunpower.com/utility-scale-solar-power-plants/, last accessed January 23, 2018

[5] https://en.wikipedia.org/wiki/Alta_Wind_Energy_Center, last accessed January 23, 2018

[6] M. Seidel *et al.*, Improving the energy efficiency of accelerator facilities, Proc. IPAC 2015, Richmond, VA, 2015 (JACoW, Geneva, Switzerland) , http://accelconf.web.cern.ch/AccelConf/IPAC2015/papers/wexc3.pdf, last accessed January 23, 2018

[7] M. Wohlmuther, Spallation target developments, Thorium Energy Conference 2013 (ThEC13), Geneva, 2013 (CERN, Geneva, Switzerland) https://indico.cern.ch/event/222140, last accessed January 23, 2018

[8] http://www.exergy.se/ftp/exergetics.pdf, last accessed January 23, 2018

[9] N.L.S. Carnot, *Réflexions sur la puissance motrice du feu* (Bachelier, Paris, 1824). https://fr.wikisource.org/wiki/R%C3%A9flexions_sur_la_puissance_motrice_du_feu, last accessed January 23, 2018

[10] J.T. Volk, *J Instrum* **6** (2011) T08003.

[11] B.J.A. Shepherd *et al.*, *J Instrum* **9** (2014) T11006.

[12] L. Rossi, The LHC superconducting magnets, Proc. PAC2003, Portland, OR (IEEE, Piscataway, NJ) http://accelconf.web.cern.ch/accelconf/p03/papers/toab001.pdf, last accessed January 23, 2018

[13] A. Baikov *et al. IEEE Trans Electron Devices* **ED-62** (2015) 3406.

[14] R. Corsini *et al.*, First full beam loading operation with the CTF3 linac, EPAC 2004, Lausanne, 2004 (JACoW, Geneva, Switzerland) http://accelconf.web.cern.ch/AccelConf/e04/PAPERS/MOOCH02.PDF, last accessed January 23, 2018, published by JACoW

[15] F. Marhauser, Cost rationales for an SRF proton linac, Proc. IPAC2014, Dresden, 2014 (JACoW, Geneva, Switzerland) http://accelconf.web.cern.ch/AccelConf/IPAC2014/papers/thpme053.pdf, last accessed January 23, 2018

[16] M. Tigner, *Il Nuovo Cimento*, **37** (1965) 1228.

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Beam Dynamics of Energy Recovery Linacs

A. Jankowiak, M. Abo-Bakr, and A. Matveenko
Helmholtz-Zentrum Berlin, Germany

Abstract
Energy recovery linacs (ERLs) combine the advantages of the two major accelerator types used at present: the high currents and low energy consumption of storage rings and the rapidly accelerated, high-brilliance beams of linear accelerators (linacs). As concepts from both accelerator types are adopted, similar beam physics challenges need to be overcome in the design of an ERL. In the first part of this article we describe the main requirements and layout options for an ERL's beam optics, pointing out the individual demands of the various subparts of the ERL: injector, merger and splitter, linac section, and recirculator. In the second part, collective high-current effects are introduced with a focus on space charge, coherent synchrotron radiation, the microbunching instability, and beam break-up. A comprehensive list of references is given for readers interested in further details of the topics introduced.

Keywords
ERL; beam dynamics; optics; collective effects.

1 Introduction

Energy recovery linacs (ERLs) can generate high-energy electron beams of huge virtual power and high density, and thus it is possible to base a synchrotron radiation light source of the ultimate brightness on them. Short pulses and high peak currents will also allow the generation of coherent radiation. Although terahertz radiation can be emitted from short bunches, low-gain free-electron laser (FEL) operation (i.e., an FEL amplifier) is also possible. Even high-gain FEL operation is feasible, as long as the beam degradation remains within the machine acceptance (transverse and momentum). Another application option is to use an ERL as a Compton source, generating hard X-rays from low-energy electrons.

In storage rings, the beam dimensions result from an equilibrium state between radiation excitation and damping, and hence are totally independent of the quality of the beam from the source, and totally independent of all pre-accelerators. In ERLs, as they are single- or few-turn machines, the passage time is much too short to reach this equilibrium and the beam quality is defined by the electron source. Whereas for a given storage ring the emittance ε scales as $\varepsilon \sim E^2$, in ERLs adiabatic damping causes an $\varepsilon \sim E^{-1}$ scaling. Thus, with increasing energy, the bunch quality in an ERL improves. Using present-day high-brightness electron sources, based on laser-induced photoemission from a gun cathode, ERLs have the potential to significantly exceed the bunch quality of modern storage-ring-based, third-generation light sources. The physics of these sources is a large topic of its own and will not be covered here [1, 2].

The central goal parameter for almost all kinds of present and future accelerators, and especially for synchrotron radiation light sources, is the brilliance $B \sim N/\varepsilon_x\varepsilon_y$, which scales with the number of electrons per second N and the transverse emittances ε_x and ε_y. Small emittances of bunches with a high charge and a high repetition rate maximize the average brilliance. Short pulses from ERLs enable insights into the dynamics of subpicosecond processes and can produce an extreme peak brilliance $\hat{B} \sim B/\sigma_s$. The spectral brilliance obtained from long insertion devices scales inversely with the energy spread, i.e., $B(\omega) \sim 1/\sigma_E$. Thus achieving the ultimate spectral brilliance, average as well as peak, requires beams of the highest electron densities, not only in 3D but also in the six-dimensional phase space.

As ERLs can reach and exceed storage ring beam parameters in any phase space dimension, many of the beam dynamics challenges known for storage rings are relevant to ERLs as well and can affect

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.439

their performance, possibly even to a higher degree. The beam dynamics challenges in an ERL arise from its general layout and target parameters, and vary with beam energy and with the function of the various machine sections.

– *Injector.* This provides high-brightness beam generation and low-energy beam transport under the influence of strong space charge forces.
– *Merger.* This guides both the low-energy fresh beam and the high-energy used beam into the same linac section.
– *Linac section(s).* This provides acceleration and deceleration of the beam. Depending on the target energy, available linac length, and average accelerating gradients, a layout based on single linac, a split linac, or a multipass linac can be chosen.
– *Spreader.* Using a multiturn layout, the various beams must be merged into the linac section, and after acceleration/deceleration be sent to beam lines according to their energy.
– *Recirculation section.* This provides lossless beam transport with conserved beam quality, with the option of beam manipulation; variants of arc lattices (Bates, double-bend achromat (DBA), triple-bend achromat (TBA), multi-bend achromat (MBA), and fixed-field alternating-gradient (FFAG)) generate the conditions for the most efficient energy recovery.
– *Splitter and dump line.* This section is analogous to the merger: downstream of the linac section, the fresh and used bunches need to be separated, for further acceleration or light generation with the former and to guide the latter into the dump line.

For this overview report, we separate the beam dynamics issues of ERL-based synchrotron radiation facilities into two main categories:

– *beam optics,* dealing mostly with charge- and current-independent problems of linear and non-linear beam transport, manipulation, and acceleration; and
– *collective effects,* caused by the high electron density and average current, which can degrade the beam quality, drive instabilities, and ultimately even lead to partial or total beam loss.

In the first, ‘beam optics’ part we will introduce general magnet optics designs applicable to ERLs, and discuss the design philosophies of the subcomponents. The requirements on the beam optics are collected together and the magnet lattice configurations that best satisfy them are compared. Non-linear effects and their compensation by adjusting the linear optics and by the use of higher-order magnetic multipole elements are considered. In the second part, the physics of potentially harmful collective effects is introduced. Options to counteract these effects by the use of special optics settings are discussed.

Since derivations of the fundamental formulas presented here are far beyond the scope of this report, a selection of references to specialized papers is given for each issue considered. In general, the online journal *Physical Review Accelerators and Beams* [3] and the proceedings of the International Particle Accelerator Conference (IPAC) and ERL workshops hosted by JACoW, the Joint Accelerator Conferences Website [4], provide an excellent source of information on all fields of ERL beam dynamics issues. As the authors were involved in the design of two ERL projects, we would like to refer the reader also to the conceptual design reports for these projects, which give a good insight into the beam dynamics aspects of low- and high-energy ERLs: bERLinPro [5, 6], currently under construction at the Helmholtz-Zentrum Berlin (HZB), and FSF, the Femto-Science-Factory [7], an HZB design study for a 6 GeV ERL-based synchrotron light source.

2 Beam optics

The magnetic lattice is defined by the type, number, and arrangement of multipole magnets and radio frequency (RF) structures. These devices are tuned to form a beam optics system, capable of transporting the beam (including acceleration and deceleration) throughout the machine while:

– maintaining the beam quality delivered from the source;
– ensuring minimum electron losses;
– merging or splitting beams of various energies, for example injected and recirculated beam(s);
– performing bunch manipulations, for example compression, emittance exchange, and plane rotation;
– establishing conditions for efficient energy recovery.

There are many challenges related to specific parts of an ERL. In contrast, particle losses and beam size are issues in all machine sections and thus will be covered here first.

2.1 Beam size and losses

In optics simulations, the beam is described by its 6D phase space size $\sigma_x, \sigma_{x'}, \sigma_y, \sigma_{y'}, \sigma_s, \sigma_E$, its emittances $\varepsilon_x, \varepsilon_y, \varepsilon_s$, and its Twiss parameters $\beta_{x,y}, \alpha_{x,y}, \gamma_{x,y}$, assuming Gaussian particle distributions. The behaviour of energy-deviating electrons is described by the dispersion function η. Various partially contradictory demands are made on the beam size.

– For a high-brilliance light source, suitable electron bunches at the point(s) of radiation generation are required. Small beam sizes in all dimensions enable generation of diffraction-limited light pulses with high transverse and longitudinal coherence fractions. For the minimum radiation wavelength λ_γ to be generated, diffraction puts a lower limit on the transverse emittances $\varepsilon_{x,y} \approx \lambda_\gamma/4n$, such that smaller electron beam emittances do not further reduce the photon beam size [8].
– Particle losses are at least as important as they are in storage rings. Although beam decay (as in storage rings) is not an issue, radiation and activation issues, as well as RF power limits, are of great importance. Especially, losses in high-energy turns need to be minimized as far as possible. The beam size is directly involved in two mechanisms:
  – *losses at the machine aperture:* the transverse beam size must be small compared with the dimensions A of the vacuum chamber: $A_{x,y}(s) > N\sigma_{x,y}(s)$. In large storage rings, N is quite high, of order 10^2–10^3, whereas in the lower-energy parts of ERLs this number can be much smaller. In dispersive sections passed through by a chirped beam, N can be of the order of 10 or even below. To reach storage-ring-like relative loss rates of 10^{-10} per turn, one needs $N \geq 7$ for a Gaussian-distributed beam. Starting from the electron source, any emission of electrons into the extreme tails of the distribution must be prevented. Nevertheless, halo electrons independent of any assumed distribution function can contribute to particle losses.
  – *Touschek losses (see Section 2.4):* electron collisions within a bunch (intrabeam scattering) lead to momentum transfer between the transverse and longitudinal motions and can be a source of beam halo formation and losses in ERLs. The loss rate from these Touschek events scales with the electron density and thus with the bunch charge and volume. A low density, i.e., a large bunch volume, reduces Touschek losses.
– Collective effects, for example space charge, coherent synchrotron radiation (CSR), and other kinds of wake fields (see Sections 3.1 and 3.2), act on the beam and imprint an energy modulation along the bunch, which ultimately deteriorates the beam quality. As the strength of all these effects scales with the peak current and thus inversely with the bunch length, bunches should be kept long during transport if possible, and only tuned short when generating radiation.
– RF curvature is important: while passing through the RF structures for acceleration or deceleration, the bunches scan the temporal and spatial field variation in the cavities, generating a correlation in the longitudinal phase space. The non-linear part of this correlation can limit bunch manipulation techniques, for example bunch compression, and increases the energy spread. As short bunches scan a smaller RF phase range, non-linearities are reduced compared with longer bunches.

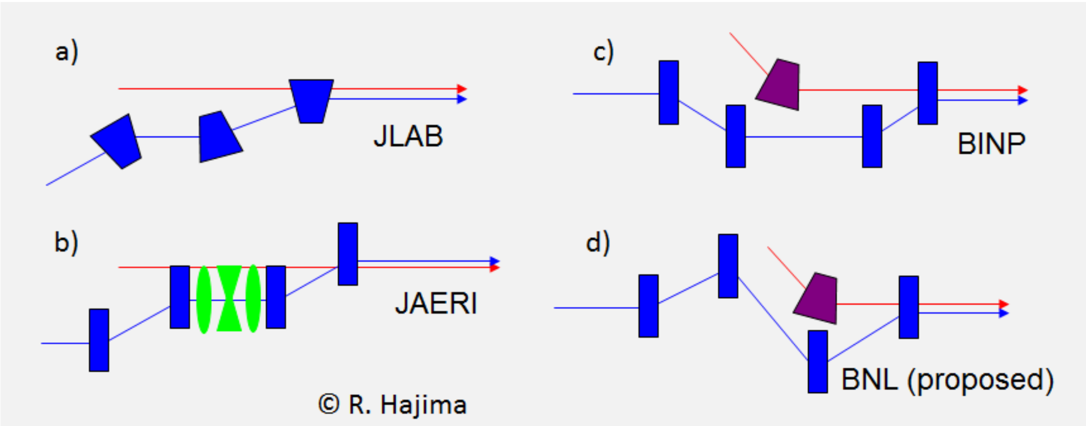


Fig. 1: Mergers for existing and proposed ERLs: (a) deflecting three-bend dog-leg, (b) four-bend dog-leg, (c) four-bend chicane, (d) 'zigzag' merger.

The optimal beam size is a compromise between these demands and has to be found for the various machine sections. Besides the beam size, many more aspects needs to be considered—the most important ones for the various machine sections of an ERL will be covered in the following.

2.2 Injector line and merger

The first machine section, which guides the beam from the source to the first multibeam linac, is referred to as the injection line here. On exit, the low-energy beam must be merged with the high-energy beam to pass through the linac on the same centered trajectory.

Beam transport in the injection line at energies of a few MeV is space charge dominated. Spatially varying forces due to self-generated fields in the bunch can cause significant emittance growth. By following an emittance compensation scheme [9, 10], a sophisticated beam optics system can reverse these space charge effects and cancel the emittance degradation to a major degree. The basic concept is described in Section 3.1. As space charge effects scale strongly with the beam energy, pre-acceleration in the injection line and before the first major acceleration will reduce the initial emittance growth. On the other hand, the pre-acceleration energy is not recovered in an ERL, and RF and dump power considerations will limit its value.

At the end of the injector line, the new and the recirculated beam have to be merged into the linac section. This is achieved by a series of bending magnets, where the last one is passed through by both beams, which are bent at different angles according to their respective energies. Whereas at the beginning of the injector line the optics can be kept axially symmetric and solenoids provide sufficient focusing, the symmetry is broken in the merger. Quadrupole magnets are used here to control dispersion (to form an achromatic bump) and to shape the beam size throughout the merger. Mergers with four different layouts, shown in Fig. 1, have been considered for ERL test facilities [11]: dog-leg-type (Fig. 1(a) and (b)), chicane-type (Fig. 1(c)), and zigzag-type (Fig. 1(d)) mergers.

In contrast to the start of the injection line, in the merger the longitudinal space-charge-induced energy modulation takes place in a dispersive section. Thus, with any energy change, an oscillation around the shifted, new reference path is excited. Since the energy modulation varies along a bunch from its tail to its head, the centroids of longitudinal slices through the bunch oscillate as well. On leaving the merger, the projected emittance in the merger plane can be significantly increased. The emittance growth of parts of the bunch with a linear energy modulation $\Delta E(s) \sim s$ (where s is the longitudinal position in the bunch) can be removed by adjusting the dispersion at the merger exit. When this is done, however, the achromaticity of the merger is broken, so that variations in the initial energy now cause an emittance growth at the merger exit. Finally, the merger is set up to minimize the overall emittance growth due to space charge dispersion and unclosed merger dispersion ($\eta_x = 0$ m out of the merger).

The same physics applies to the splitter, which divides the accelerated, high-energy beam from the decelerated, low-energy beam, which is sent into the dump line.

Stray fields also need to be considered. Although they are unwanted in general, interfering fields such as the Earth's magnetic field, remanent fields from the optics magnets, and magnetic fields from vacuum pumps and gauges are most distorting in the injection line owing to the low beam energy and low rigidity. Shielding of fields, magnet-cycling procedures, and careful placing of vacuum devices reduce these stray fields. For the remaining fields, trajectory offsets have to be corrected with a sufficient number of steerer magnets.

2.3 Linac sections

One or more ERL sections are equipped with linacs to accelerate the beam in one or several turns up to its final energy. Several aspects of the beam dynamics have to be considered.

- *RF focusing.* The cavity fields focus the beam [12], both horizontally and vertically, when it enters the cavity, and defocus it when it leaves. During acceleration, owing to the energy increase in the cavity, the focusing on the low-energy side prevails over the defocusing on the high-energy side. The opposite effect happens during deceleration. Especially at low energies, the focal strength is high and needs to be carefully considered when the linac section beam optics are being set up.
- *RF phase slip.* At low injection energies, the beam is not sufficiently relativistic and time-of-flight effects can cause a phase slip relative to the recirculated, high-energy beam. A power mismatch in the RF cavities is the consequence, and beam-loading problems arise. The effect can be reduced by increasing the injection energy, but clearly at the cost of the RF and dump power.
- *Multienergy beam lines.* The linac sections are passed through by beams of different energies, sharing the same focusing elements, namely magnets and RF structures. The difficulty of finding suitable optics for all beams scales with the range of energy in the beam line. The optics are mainly tuned with respect to the lowest-energy beam, because it has the lowest magnetic rigidity. Any other strategy would lead to strong overfocusing and an unsuitable beam size. The lack of focusing for the high-energy beam has to be compensated in a separate beam transport section or sections.
- *Spreader.* The separation of the multiple beams into energy-adjusted beam lines is done by a spreader, using the energy dependence of the bending angle in the first, shared dipole magnet(s). The challenge here is to create a compact layout, using a small number of magnets even for several beams of different energies. The dispersion in the spreader plane should be closed at its exit, and the beam size must be matched to the recirculation arcs to avoid emittance degradation.
- *Beam break-up (BBU).* The BBU instability (see Section 3.5) is driven by a positive feedback of the beam into higher-order-mode (HOM) fields of the superconducting RF cavities. Although the most important countermeasure is the use of cavities with a minimized HOM spectrum, the beam optics also influence BBU: a betatron phase advance of $\Delta\psi = n\pi$ between consecutive cavity passages sets the transport matrix element R_{12} to zero, so that the beam passes through the cavity on axis after recirculation and no power is fed into the HOMs (see Eq. (2)). In addition, optimized Twiss parameters for the linac can be calculated [13]. Both of these measures can significantly increase the instability threshold. Effective measures against BBU become even more important for multiturn ERLs, where various beams (multiplying the total current) traverse the linac sections simultaneously.

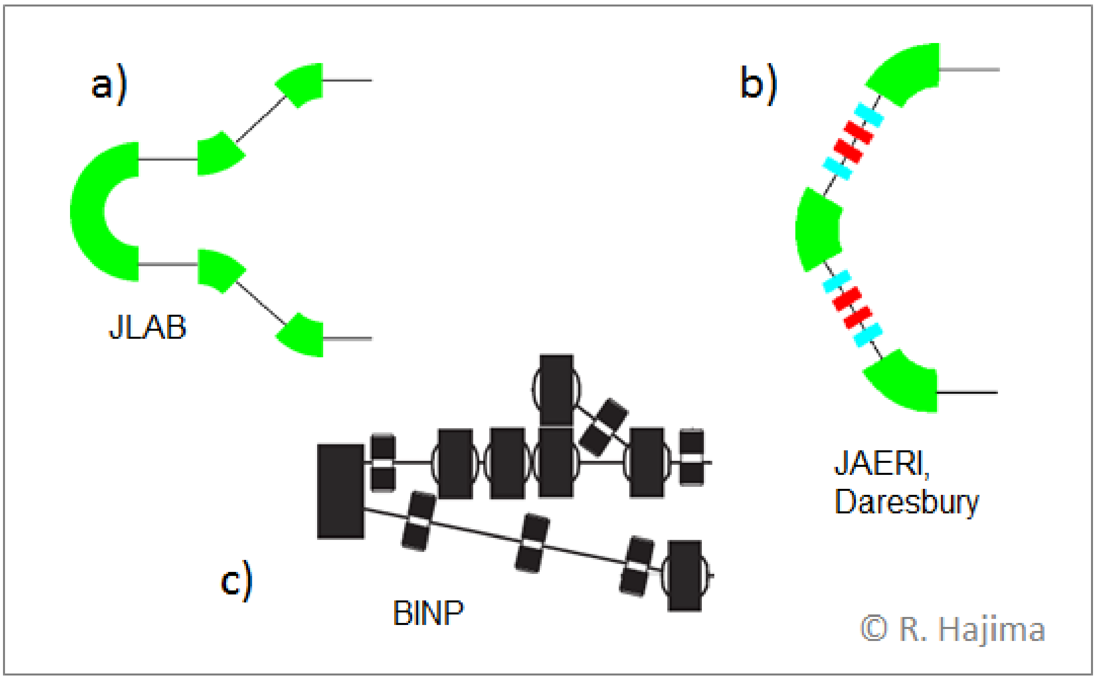


Fig. 2: Recirculator arc lattice types: (a) Bates arc, (b) TBA, (c) BINP arc

2.4 Recirculators

The transfer lines connecting the ERL linac sections and the full-energy section dedicated to radiation generation are referred to here as recirculators. Together with the linac sections, they form the majority of the machine sections in an ERL. A careful beam optics set-up, fulfilling a variety of demands, is mandatory. Several basic lattice concepts are suitable for ERL recirculator arcs, depending on the energy and on the available space and number of magnets [14]. In low- to medium-energy ERLs of moderate size, DBAs [15], TBAs [16], Bates arcs [17], and also individual, non-standard schemes have been applied, as shown in Fig. 2.

For large-scale ERL-based light sources with energies in the GeV range, multibend achromat lattices and FFAG lattices have been considered [18–22]. The various lattice types differ in their tunability, space, and magnet number requirements, and in performance with respect to emittance conservation, lossless beam transport, and beam manipulation capabilities. Flexible control of the linear and non-linear beam optics is the key to covering all of the aspects mentioned above.

2.4.1 Lossless beam transport

As mentioned earlier, Touschek scattering is one of the two dominant loss processes. Besides a large bunch volume, which is contrary to radiation generation requirements, the momentum acceptance A_p of the optics is of crucial significance. Although energy transfer due to intrabeam scattering into the transverse motion is of minor importance, the longitudinal momentum change is Lorentz transformed into the laboratory frame and is thus strongly enhanced. With a momentum change $\Delta p/p$ from a scattering event, the downstream reference trajectory shifts to a dispersive path $x_{\text{ref}}(s) = \eta(s) \cdot \Delta p/p$. Depending on the dispersion function at the scattering position, a betatron oscillation of initial amplitude

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \frac{\Delta p}{p} \begin{pmatrix} \eta \\ \eta' \end{pmatrix}$$

may be excited in addition. This is equivalent to a single-particle emittance of

$$\varepsilon_0 = \gamma x^2 + 2\alpha x x' + \beta x'^2 \quad = \quad (\Delta p/p)^2 (\gamma \eta^2 + 2\alpha \eta \eta' + \beta \eta'^2) = (\Delta p/p)^2 \mathcal{H} \,, \tag{1}$$

$$\mathcal{H} = \gamma\eta^2 + 2\alpha\eta\eta' + \beta\eta'^2 \tag{2}$$

(the Twiss parameter, dispersion, and $\mathcal{H}$-function are evaluated at the scattering position $s = s_0$). The general expression for the downstream trajectory of the scattered electrons is

$$\begin{aligned} x(s) &= \sqrt{\varepsilon_0\beta(s)}\cos(\psi(s) - \phi_0) + \eta(s)\,\Delta p/p & (3)\\ &= \Delta p/p\left[\sqrt{\mathcal{H}_0\beta(s)}\cos(\psi(s) - \phi_0) + \eta(s)\right], & (4)\end{aligned}$$

which is directly proportional to $\Delta p/p$. Scattering events that cause a downstream offset larger than the available horizontal aperture $A_x(s)$ lead to particle losses. The maximum deviation $\Delta p/p$, where $x(s) = f(\Delta p/p) \le A_x(s)$, defines the momentum acceptance $A_p(s)$ of the optics. The Touschek loss rate [23] scales as $\dot{N}/N \sim 1/A_p^3$, and therefore a large momentum acceptance is essential for low losses. A small overall dispersion and lower maxima of the beta and dispersion functions optimize the $\mathcal{H}$-function (reducing its maximum value) and thus increase the momentum acceptance (see also Section 3.2). Lattices with lower bending angles of the dipole magnets are advantageous, but require more magnets at increased cost.

The loss rate due to elastic scattering from the atomic nuclei of the residual gas is the second main loss mechanism. The loss rate scales as $\dot{N}/N \sim 1/(\theta_x^2 + \theta_y^2)$, with the angular acceptance $\theta_{x,y}^2 \approx A_{x,y}^2/(\langle\beta_{x,y}\rangle\beta_{x,y}^{\max})$. Smaller transverse beta functions with lower maximum values increase the angular acceptance, thus reducing the loss rate. Besides the intentionally generated 'wanted beam', there are a few sources of unwanted beam, for example stray light from the gun laser, extreme tails of the distribution, ghost pulses, and dark current from field emission from the (superconducting) RF structures. This unwanted beam, often referred to as beam halo, can be simulated if the generating process is known. Unfortunately, the dominating contributor only becomes apparent in the real machine, and may even change its origin. The best measure to control beam halo is a large acceptance of the magnet optics to transport both the core beam and the halo.

2.4.2 Bunch manipulation

To generate the most brilliant light pulses, several manipulation techniques are applied that exchange parts of phase space between two planes by means of quasi-phase-space rotations. Conservation of the uninvolved phase space dimensions and the overall beam quality is mandatory.

In many linear accelerators and ERLs, bunch compression is used, where in the first step a chirp (mostly a linear z–p_z correlation, $\Delta p/p = C \cdot \Delta z$) is imprinted by passing the beam through the RF structures off crest. In the second step, the beam passes through a dispersive section with $\eta \neq 0$, where the path length depends on the particle momentum in accordance with $\Delta L = R_{56}\,\Delta p/p + T_{566}(\Delta p/p)^2 + \ldots$, with $R_{56} = \int \eta/\rho\, ds$ and T_{566} as the first- and second-order beam transport matrix elements. Nonlinearities (RF curvature, T_{566}, etc.) can be corrected using higher-order multipole magnets, starting with sextupole magnets at the lowest order. Whereas extra bunch compressor sections are often provided in linacs, in ERLs the recirculation arcs can be used as an alternative. The various lattice types offer different amounts of variability for tuning the optics: for an achromatic arc ($\eta_{\text{in}} = \eta_{\text{out}} = 0$), the DBA lattice offers no R_{56} tunability at all, whereas, for example, in a TBA lattice R_{56} can be tuned via the dispersion function of the middle bend. With more quadrupole magnets in the more complex lattice types, one generates 'free knobs' to adjust the dispersion and beta functions for non-linear corrections, minimizing the required multipole strengths. Also, the phase advance in certain sections can be tuned with respect to emittance-degrading effects, for example CSR.

Another manipulation that can be done in ERLs is so-called 'beam rotation', where the two transverse phase spaces are completely switched. This can increase the BBU threshold for polarized cavities, since no further excitation of the kicking HOM occurs on the return pass. A section with a set of skew quadrupole magnets is required to swap the transverse planes, ideally transforming the beam in accordance with

$$\vec{X}_1 = M\vec{X}_0, \qquad \text{with} \qquad M = \begin{bmatrix} 0 & E \\ E & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

There are more manipulation techniques, for example emittance exchange, but they have either not been used or not been required in ERLs so far.

2.4.3 Energy recovery

A further important task of recirculator arcs is to provide path length adjustment options, to enable one to set up accurately the required RF phase advances of 0 or 180° between linac sections. Depending on the ERL layout, these tuning options may be needed not only in the deceleration but also possibly in the acceleration pass. Common options are movable arcs for small ERLs with 180° DBA, TBA, or Bates arcs (the latter only in the large centre magnet), two longitudinal movable bends within an arc (e.g., in bERLinPro), or high-amplitude steering bumps in large recirculators with sufficient mechanical aperture. Chicanes outside the recirculators can also be used, but they can significantly contribute to the R_{56} budget and can only lengthen the pass (compared with the straight option with all bends off). Moreover, a lengthening of the order of the RF wavelength with a chicane requires large offsets and is hard to achieve in a single wide vacuum chamber.

Beam matching, especially in the last deceleration to the lowest energy, is of vital importance for the efficiency of the recovery process. Minimization of the energy spread at the low-energy side is a precondition for a high recovery rate and for safe transport of the high-power beam into the dump. One option to cancel out RF curvature effects is to adjust the bunch length so that it is equal to that during acceleration. In this case, any bunch compression needs to be reversed. Owing to beam-loading effects, this can be only done by inverting the sign of R_{56} in the corresponding recirculator sections, which again favours highly tunable lattice types with a wide range of values of R_{56}. If the bunch length differs between acceleration and deceleration, sextupole magnets can be used to remove RF non-linearities. Any remaining non-linearities arising from the magnet optics or from collective effects (CSR, wakes, ...) have to be minimized using higher-order multipole magnets.

For efficient recovery, the transverse beam size in the linac needs to be adjusted to take RF focusing into account and also to provide suitable BBU conditions.

2.4.4 Radiation excitation

Despite the transfer line character of ERLs and the short passage times, the emission of incoherent synchrotron radiation (ISR) can cause considerable emittance growth. Since the energy loss due to emission of synchrotron radiation scales as $\Delta E \sim E^4/\rho \sim E^3 B$, high-energy ERLs are most affected. Moreover, the critical photon energy ε_c scales non-linearly with energy, as $\varepsilon_\mathrm{c} \sim E^3/\rho \sim E^2 B$, extending the photon spectrum equivalently to higher values, and thus increasing the resulting energy change and energy spread of the emitting electrons. Similarly to a Touschek event, an energy change in a dispersive section excites a betatron oscillation around the new reference orbit. The corresponding emittance growth is described by the function $\mathcal{H}$ (see Eq. (1)), which relates the momentum change to the amplitude of the downstream transverse betatron oscillations. A low $\mathcal{H}$-function represents an optics system where momentum changes cause a smaller transverse-phase-space blow-up and thus reduced emittance growth. Assuming an achromatic arc tuning, minimal form factors $F \sim \mathcal{H}$ can be calculated for the various lattice types for comparison [24]. Compared with the DBA lattice, the theoretical minimum for an MBA lattice is reduced by a factor of 3 when the Twiss parameters and the dispersion function are optimized to reduce $\mathcal{H}$ and extra bends at the beginning and end of the cell close the dispersion. Thus, the emittance growth will be smaller with an MBA lattice, but zero-dispersion sections, for example for insertion devices, are not available without lattice modifications.

2.5 Dump line and beam dump

The ERL's last section is the dump line, which guides the low-energy but high-power beam into the beam dump, where sufficient cooling power is provided to safely absorb the beam in the walls. Losses in the dump line are no longer relevant for the RF budget, but the high-power beam has substantial damage potential. When mis-steered, the beam is able to melt holes in vacuum chamber components on a very short timescale. Thus lossless beam transport is the main task of the beam optics. Large apertures are essential, therefore, to allow safe beam transport with increased emittance (compared with that in the injector) and a moderate further emittance degradation due to space charge effects in the dump line.

In the dump, the full beam power of hundreds of kilowatts or even megawatts is mostly transferred into heat (and radiation). Clearly, this must not happen in an area of a few square millimetres or less, not even an area of a few square centimetres. Instead, the beam power needs to be carefully distributed over the inner surface of the dump, usually over an area of the order of 1–2 m^2. Two options exist:

– beam widening by massively increasing (by orders of magnitude) the beta functions in the last part of the dump line and in the dump;
– beam sweeping using two rapid-cycling (at tens of hertz) transverse steerers, to distribute equally the beam impact points in the dump.

Ideally, a combination of both is used to relax the hardware requirements and to improve the reaction time of the 'machine protection system' in the case of device failures.

3 Collective effects

The intensity and quality of the beam in an accelerator are usually limited by collective effects. In the following, the characteristic effects and their peculiarities in the case of ERLs are discussed.

3.1 Space charge

Space charge effects typically limit the performance of the low-energy beam transport in high-brightness photoinjectors. One direct effect is transverse defocusing of the beam by space charge forces within bunches. The linear part of the forces can be compensated by external focusing (by solenoids or quadrupoles), but the non-linear part still affects the beam quality. Emittance degradation due to collective space charge forces is one of the important issues in the design of injector optics. Flat-top cathode laser profiles, both transversely and longitudinally, which linearize the space charge forces in the central part of the beam, are routinely used to achieve the highest beam brightness [1, 25].

If the aim is to achieve a high-brightness electron beam in an ERL, the injector should be designed with the emittance compensation technique [9] in mind. The critical difference between an ERL and a conventional linac injector is the merger section, where axial symmetry of the beam can no longer be assumed. This means that emittance compensation with a solenoid is not enough any more to achieve the minimum beam emittance in both planes. A theory of so-called '2D emittance compensation' was developed in Ref. [10]. The application of this method to the superconducting RF photoinjector in the bERLinPro project [5] is described in [26].

Space charge effects determine the choice of the merger geometry. An overview of practical merger designs can be found, for example, in Ref. [11]. One problem with a space-charge-dominated merger is the longitudinal space charge force, which affects the transverse motion of individual bunch slices in a dispersive section. Transverse defocusing and changes in the energy of a slice, caused by space charge forces, can modify the achromatic condition significantly. This effect favours merger designs that are short and have a low dispersion [26]. The linear part of the effect can be corrected if the bunch has a sufficiently large correlated energy spread.

Particle-tracking codes (e.g., Parmela [27], ASTRA [28], and GPT [29]) can be used to model space-charge-dominated beams. Usually, these programs require extensive resources for tracking, which makes optimization of beam lines time- and resource-consuming. There are space charge codes (e.g., Trace3D [30], SCO [31], and HOMDYN [32]), which allow fast tracking of a model charge distribution (a Kapchinsky–Vladimirsky distribution, applied to a whole bunch or slicewise). These codes allow an initial optimization to be achieved quickly; afterwards, tracking with 'full' space charge codes can be done.

See specialized contributions in this CAS, especially "Space Charge Mitigation" (Massimo Ferrario, INFN-LNF).

3.2 Coherent synchrotron radiation

Although synchrotron radiation is usually emitted incoherently (i.e., ISR), very short electron bunches generate CSR with wavelengths comparable to the bunch length. The resulting energy loss can become very significant. For typical bunch lengths in storage rings (20–100 ps), CSR is shielded by the vacuum chamber and plays only a minor role in the beam dynamics. In linacs, however, where bunches can easily be compressed, CSR can strongly influence the beam parameters. Moreover, with the high average beam currents typical of ERLs, CSR can cause damage to vacuum system components as a result of its high average power. As an example [5], the CSR losses of bERLinPro in normal operation with 2 ps bunches and 100 mA average current were estimated to be 2.5 kW. For short-pulse operation mode at full current (100 mA) and with bunch lengths down to 150 fs, the losses would increase to about 25 kW.

The main problems and solutions related to the effects of CSR on the beam emittance have been investigated, for example, for short-wavelength FELs (FLASH, LCLS, and XFEL). If the key effect of the CSR wake on a bunch is a longitudinal-position-dependent transverse kick to the bunch slices, a 1D model can be a good approximation. A comprehensive derivation of 1D CSR wake functions for different cases is presented in [33]. This model has also been implemented in a number of packages (e.g, Elegant and Opal) [34, 35].

For very short bunches (if the bunch is, in its reference frame, equal in length to or shorter than its transverse sizes), the full 3D radiation field should be taken into account. Appropriate simulation codes (e.g., CSRTrack [36]) should be used in this case. However, the 1D model usually gives an overestimation of the effect and can still be used for quick checks.

CSR-induced emittance growth can be reduced by several methods. The increase in the transverse emittance is proportional to the $\mathcal{H}$-function (Eq. (1)), so keeping this function low reduces emittance degradation. In an ERL, this measure is essential in the magnets, where the bunches are at their shortest.

If the effect on the bunches is small, magnetic optics with a repetitive symmetry and an appropriate betatron phase advance between cells can cancel out the CSR kicks (see, e.g., [37] and references therein for the implementation of this method in FERMI@Elettra). The idea is easy to understand when the bunch length does not change along the beam line, so that the CSR emission conditions don't vary, i.e. with an isochronous arc or a bunch without correlated energy spread. In this case the energy change imprinted on every bunch slice is the same in each cell of the periodic focusing system. The final displacement and angle of the slice are the sums of the displacements and angles arising from each cell (i.e., a superposition). If the betatron phase advance from cell to cell is $2\pi k/N$, where k is any integer and N is the number of cells, then

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \sum_{k=1}^{N} \exp\left[2\pi i k/N\right] = 0 \, ,$$

i.e., all slices are aligned again.

A similar approach is possible even for a periodic arc with bunch compression. In this case the assumption of a self-similar CSR wake is necessary, which is not always satisfied. For example, the CSR wakes in the drifts are not self-similar. The implementation of such emittance correction is described, for example, in [7].

3.3 Microbunching instability

The average power coherently emitted from a short bunch is $P_{\text{total}} \sim Q_{\text{B}}^2/R^{2/3}\sigma_z^{4/3}$, and is capable of causing significant beam quality degradation. The effect can be greatly intensified if the bunch is structured on scales much shorter than the bunch length.

The mechanisms of such 'microbunching' can involve different wakes, the most important being those due to the longitudinal space charge (LSC) and the CSR itself [38–40]. The CSR wake shows only a weak dependence on the beam energy, whereas the LSC wake scales with $1/\gamma^2$ and therefore plays an important role in the low-energy, injector part of an accelerator. The wake imprints an energy modulation on the bunch, which can be transformed further into a longitudinal density modulation (microbunching). In a storage ring, the momentum compaction factor α_{c} and, in a linear accelerator, the element R_{56} of the transport matrix is responsible for this. This is the same matrix element that is necessary for bunch compression, so the two processes are intrinsically dependent on each other. The amplification factor of the density modulation (gain) in a beam line can be found, for example, in Ref. [40]. The initial density modulation can be imprinted in the RF photogun, for example by the longitudinal profile of the laser beam, which may be generated with a pulse shaper. Some details of the analysis that was done for LCLS can be found in Ref. [41]. Shot noise in the beam is another possibility, which usually gives a much lower initial modulation amplitude. The gain of this instability scales with the peak current of the bunch. An uncorrelated energy spread in the bunch smears out the bunching and can be used to suppress the instability [42]. A laser heater [43] is one option to increase the energy spread in a slice controllably; using a strong wiggler to induce an energy spread through emission of ISR is another.

3.4 Wakes and impedances

Resistive walls, surface roughness, and geometric wakes are other sources of distortions in ERL beam dynamics. Usually these distortions are smaller than those due to the CSR and LSC wakes. However, if countermeasures are taken to reduce or (ideally) completely compensate the effects of the CSR wake, they can become the main concern.

The resistive-wall impedance is usually higher for ERLs than for storage rings owing to the short bunch length achievable. The scaling is $k_{\text{L}} \sim \sigma_z^{-3/2}$ (for $\sigma_z > a/\gamma$), where k_{L} is the longitudinal loss factor, σ_z is the bunch length, a is the radius (half gap) of the vacuum chamber, and γ is the relativistic factor [44]. Surface roughness can also be an issue. For example, smooth NEG coating of the vacuum chambers may be necessary. Resonances in geometric wakes (when spectral lines of the beam coincide with resonances of the structure) should be avoided at the design stage, as was done, for example, for bERLinPro [45].

3.5 Linac configuration and beam break-up

Dipole-mode-driven transverse BBU can be a serious limitation in high-current operation of an ERL. This is primarily an ERL-specific problem, since accelerators that have high-quality-factor cavities (superconducting) and operate with a high average current are vulnerable to this instability.

Transverse BBU was observed and understood well at the JLab ERL [46]. A simple analytical scaling can be derived for the 'one cavity, one mode, one turn' case:

$$I_{\text{th}} = -\frac{2pc^2}{e\omega(R/Q)QR_{12}\sin(\omega T)} , \tag{5}$$

where I_{th} is the threshold current for the instability, p is the beam momentum, ω is the dipole mode frequency, $(R/Q)Q$ is the mode impedance, R_{12} is the element of the transport matrix of the recirculation, and T is the recirculation time. In the case of coupled optics and an arbitrary polarization angle α of the mode,

$$R_{12}^* = R_{12}\cos^2\alpha + (R_{14} + R_{32})\sin\alpha\cos\alpha + R_{34}\sin^2\alpha$$

has to be used instead of R_{12} [47]. The threshold current is proportional to the beam energy, so the most problematic cavities are those where the beam has its lowest energy. The threshold current for transverse BBU in the case of a single cavity and a single TM_{110} mode for a multipass ERL can be estimated as [48]

$$I_\mathrm{b} \approx I_0 \frac{\lambda^2/4\pi^2}{QL_\mathrm{eff}\sqrt{\sum\limits_{m=1}^{2N-1}\sum\limits_{n=m+1}^{2N}(\beta_m\beta_n/\gamma_m\gamma_n)}}\,,$$

where I_0 is the Alfvén current, Q is the quality factor of the HOM, λ is the wavelength corresponding to the resonant frequency of the mode, γ_m is the relativistic factor at the mth pass through the cavity, β_m is the Twiss parameter, L_eff is the effective length of the cavity, and N is the number of acceleration passes. This expression indicates the limitation on the number of passes and suggests an optical design with beta functions as low as possible for cavities with low beam energy.

As was shown in Ref. [49], the BBU threshold current for an N-turn ERL may be estimated as roughly $N(2N-1)$ times smaller than that for a single-turn machine. The worst-case scenario, where the betatron phase advances between all pairs of passes through the cavity were $\sin(\mu_{nm}) = 1$, was assumed in that estimate. The expression in Ref. [48] gives another estimate, assuming random phases, which is closer to reality for a 'large' number of cavities and passes. Numerical modelling of the transverse BBU instability is necessary to take many linac cavities into account, with all relevant modes. A number of codes for this purpose exist [50–52]. Also, a high arc chromaticity has been proposed as a measure to stabilize the beam against transverse BBU [53].

Longitudinal BBU driven by monopole modes is another issue for ERLs. In this case the longitudinal dispersion R_{56} replaces R_{12} in the estimate of the threshold current in Eq. (5) (see, e.g., [54]). If a single-turn ERL operates with $R_{56} = 0$, it is not vulnerable to this instability (at least in theory). However, in a multiturn ERL with bunch compression in the arcs, $R_{56} \neq 0$ and an analysis of this instability becomes necessary.

3.6 Ion trapping

ERLs are vulnerable to the effects of ions accumulated in the potential well of the electron beam. These effects include:

- optical errors due to strong focusing of the electron beam by the space charge of the ion cloud;
- higher electron-beam scattering rates, leading to the formation of a beam halo and increased beam losses;
- ion-induced beam instabilities.

The ions are produced by electron ionization of the residual gas (ionization by synchrotron radiation is also possible). Confined inside the 'time-averaged electrostatic potential' of the electron beam, ions can be 'trapped' in the beam for a relatively long time, oscillating near the minima of the potential. These minima coincide with the minima of the beam size for axially symmetrical beams.

Simulation of the formation and dynamics of the ion cloud is complicated by the complex trajectories of ions in the potential of a non-axisymmetric electron beam and the fact that the dynamic equilibrium that determines the neutralization factor of the electron beam is defined by a competing process of ion heating (by scattering from the electrons). Modelling of these processes is therefore a complex task; some results can be found, for example, in Ref. [55]. The methods for clearing ions in an ERL are basically the same as those used in storage rings. Clearing electrodes, gaps in the bunch train, and resonant excitation of the ion cloud are discussed, for example, in [56, 57]. For small-scale machines, a gap is not a good option owing to the short recirculation time. The variable beam loading due to the fluctuating beam current is a general concern.

References

[1] S.J. Russell, *Nucl. Instrum. Methods Phys. Res. A* **507**, Issues 1-2, p. 304-309, (2003).

[2] A. Arnold and J. Teichert, *Phys. Rev. ST Accel. Beams* **14** (2011) 024801. https://doi.org/10.1103/PhysRevSTAB.14.024801

[3] http://journals.aps.org/prstab/.

[4] http://www.jacow.org/.

[5] Conceptual design report, bERLinPro, edited by B. Kuske, N. Paulick, A. Jankowiak, and J. Knobloch (2012).

[6] M. Abo-Bakr *et al.*, Progress report of the Berlin energy recovery project bERLinPro, Proc. IPAC15, Richmond, VA, 2015.

[7] T. Atkinson, A. Bondarenko, A. Matveenko, and Y. Petenev, 'Conceptual design report for a multi-turn energy recovery linac-based synchrotron light facility (femto-science factory) (Helmholtz-Zentrum Berlin, 2015), Section 5.2.3. https://doi.org/10.5442/R0002

[8] H. Wiedemann, *Particle Accelerator Physics* (Springer, 1993). https://doi.org/10.1007/978-3-662-02903-9

[9] L. Serafini and J. Rosenzweig, *Phys. Rev. E* **55** (1997) 7565. https://doi.org/10.1103/PhysRevE.55.7565

[10] S.V. Miginsky, *Nucl. Instrum. Methods Phys. Res. A* **603** (2009) 32. https://doi.org/10.1016/j.nima.2008.12.124

[11] V.N. Litvinenko, R. Hajima, and D. Kayran, *Nucl. Instrum. Methods Phys. Res. A* **557** (2006) 165. https://doi.org/10.1016/j.nima.2005.10.065

[12] J. Rosenzweig and L. Serafini, *Phys. Rev. E* **49** (1994) 1599. https://doi.org/10.1103/PhysRevE.49.1599

[13] Y. Petenev, Dissertation, HU Berlin, 2014.

[14] R. Hajima, *Nucl. Instrum. Methods Phys. Res. A* **557** (2006) 45. https://doi.org/10.1016/j.nima.2005.10.050

[15] R. Chasman *et al.*, Preliminary design of a dedicated synchrotron radiation facility, Proc. PAC 1975, Washington, DC, 1975.

[16] H. Owen and B. Muratori, Choice of arc design for the ERL prototype at Daresbury Laboratory, Proc. EPAC 2004, Lucerne, Switzerland, 2004.

[17] J.B. Flanz *et al.*, An isochronous beam recirculation magnet system, Proc. PAC 1981, Washington, DC, 1981.

[18] D. Trbojevic *et al.*, ERL with non-scaling fixed field alternating gradient lattice for eRHIC, Proc. IPAC'15, Richmond, VA, 2015.

[19] D. Trbojevic *et al.*, Lattice design for the future ERL-based electron hadron colliders eRHIC and LHeC, Proc. PAC11, New York, 2011.

[20] E. Jensen *et al.*, Design study of an ERL test facility at CERN, Proc. IPAC'14, Dresden, Germany, 2014.

[21] E. Cruz-Alaniz *et al.*, Tracking studies in the LHeC lattice, Proc. IPAC15, Richmond, VA, 2015.

[22] E.C. Aschenauer *et al.*, eRHIC design study: An electron–ion collider at BNL. https://arxiv.org/abs/1409.1633.

[23] A. Piwinski, The Touschek effect in strong focusing storage rings. https://arxiv.org/abs/physics/9903034.

[24] S.Y. Lee, *Phys. Rev. E* **54** (1996) 1940. https://doi.org/10.1103/PhysRevE.54.1940

[25] F. Stephan *et al.*, *Phys. Rev. ST Accel. Beams* **13** (2010) 020704. https://doi.org/10.1103/PhysRevSTAB.13.020704

[26] A.V. Bondarenko and A.N. Matveenko, Emittance compensation scheme for the bERLinPro injector, Proc. IPAC'11, San Sebastian, Spain, 2011.

[27] L.M. Young and J.H. Billen, PARMELA documentation, LA-UR-96-1835, Los Alamos National Laboratory, Revised April 18, 2005.

[28] K. Floettmann, ASTRA: A space charge tracking algorithm, http://www.desy.de/ mpyflo/.

[29] S.B. van der Geer *et al.*, GPT: The general particle tracer, http://www.pulsar.nl/gpt.

[30] D. Rusthoi, W. Lysenko, and K. Crandall, Further improvements on trace 3-D, Proc. PAC (1997), Vancouver, B.C., Canada.

[31] A.V. Bondarenko and A.N. Matveenko, Implementation of 2D-emittance compensation scheme in the bERLinPro injector, Proc. FEL 2011, pp. 564–567, Shanghai, China.

[32] M. Ferrario, HOMDYN user guide, SLAC 3/1/02 (1999).

[33] E.L. Saldin, E.A. Schneidmiller, and M.V. Yurkov, *Nucl. Instrum. Methods Phys. Res. A* **398** (1997) 373. https://doi.org/10.1016/S0168-9002(97)00822-X

[34] M. Borland, *Phys. Rev. ST Accel. Beams* **4** (2001) 070701. https://doi.org/10.1103/PhysRevSTAB.4.070701

[35] https://amas.psi.ch/OPAL.

[36] M. Dohlus and T. Limberg, CSRTrack user manual, http://www.desy.de/xfel-beam/csrtrack/.

[37] S. Di Mitri, M. Cornacchia, and S. Spampinati, *Phys. Rev. Lett.* **110** (2013) 014801. https://doi.org/10.1103/PhysRevLett.110.014801

[38] G. Stupakov, Theory and observations of microbunching instability in electron machines, SLAC-PUB-9880 (2003).

[39] G. Stupakov and S. Heifets, *Phys. Rev. ST Accel. Beams* **5** (2002) 054402. https://doi.org/10.1103/PhysRevSTAB.5.054402

[40] Ya.S. Derbenev, J. Rossbach, E.L. Saldin, and V.D. Shiltsev, Microbunch radiative tail–head interaction, Deutsches Elektronen-Synchrotron Report No. TESLA-FEL 95-05 (1995).

[41] J. Wu *et al.*, Temporal profile of the LCLS photocathode ultraviolet drive laser tolerated by the microbunching instability, SLAC-PUB-10430 (2004).

[42] S. Heifets *et al.*, *Phys. Rev. ST Accel. Beams* **5** (2002) 064401. https://doi.org/10.1103/PhysRevSTAB.5.064401

[43] Z. Huang *et al.*, *Phys. Rev. ST Accel. Beams* **7** (2004) 074401. https://doi.org/10.1103/PhysRevSTAB.7.074401

[44] A. Chao and M. Tigner (Eds.), *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1999).

[45] H.-W. Glock *et al.*, Loss factor and impedance analysis of warm components of bERLinPro, Proc. IPAC15, Richmond, VA, 2015.

[46] C. Tennant *et al.*, *Phys. Rev. ST Accel. Beams* **8** (2005) 074403. https://doi.org/10.1103/PhysRevSTAB.8.074403

[47] E. Pozdeyev, *Phys. Rev. ST Accel. Beams* **8** (2005) 054401. https://doi.org/10.1103/PhysRevSTAB.8.054401

[48] N.A. Vinokurov *et al.*, *Proc. SPIE* **2988** (1997) 221. https://doi.org/10.1117/12.274385

[49] G.H. Hoffstaetter and I.V. Bazarov, *Phys. Rev. ST Accel. Beams* **7** (2004) 054401. https://doi.org/10.1103/PhysRevSTAB.7.054401

[50] E. Pozdeyev *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **557** (2006) 176. https://doi.org/10.1016/j.nima.2005.10.066

[51] I. Bazarov, bi—beam instability BBU code. http://www.lepp.cornell.edu/ ib38/bbu/

[52] L. Merminga and I.E. Campisi, Higher-order-models and beam breakup simulations in the Jefferson Lab FEL recirculating linac, 19th International Linear Accelerator Conference, 1998.

[53] V.N. Litvinenko, *Phys. Rev. ST Accel. Beams* **15** (2012) 074401. https://doi.org/10.1103/PhysRevSTAB.15.074401

[54] J. Bisognano and M. Fripp, Requirements for longitudinal HOM damping in superconducting linacs, CEBAF-PR-89-018, 1989.

[55] G. Pöplau, U. van Rienen, and A. Meseck, *Phys. Rev. ST Accel. Beams* **18** (2015) 044401. https://doi.org/10.1103/PhysRevSTAB.18.044401

[56] G.H. Hoffstaetter and M. Liepe, *Nucl. Instrum. Methods Phys. Res. A* **557** (2006) 205. https://doi.org/10.1016/j.nima.2005.10.069

[57] S. Full *et al.*, Detection and clearing of trapped ions in the high current Cornell photoinjector. http://arxiv.org/pdf/1508.00923.pdf

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

High-Gain Regime: 1D[1]

K.-J. Kim[1], *L. Ryan*[2] *and Z. Huang*[3]
[1]The University of Chicago, Illinois, USA
[2]Argonne National Lab, Illinois, USA
[3]Stanford University, California, USA

Abstract
We discuss the free-electron laser physics in high-gain regime in 1D regime, which contains the most important aspects of the free-electron laser dynamics. The high-gain regime is particularly important when mirrors are not available to build oscillators, and has been used as the most straightforward way to produce intense X-rays from FELs.

Keywords
High-gain regime; FELs.

1. Introduction

A free-electron laser (FEL) can act as a high-gain amplifier, in which case the energy exchange during a single pass through the undulator is large and the field amplitude cannot be regarded as a constant. Therefore, it is necessary to consider the field evolution, so that we must study the pendulum equations for the electron motion and Maxwell equation for the radiation. We will derive these equations, limit them to 1D case. An approximate solution of the coupled Maxwell-pendulum equations are obtained to exhibit the basic characteristics.

1 Maxwell equation

An FEL is a natural extension of spontaneous undulator radiation once we include the self-consistent electron motion in the radiation field. Thus we may begin our FEL derivation starting from the paraxial approximation of Maxwell equation [1] for N_e electrons arbitrarily distributed:

$$\left[\frac{\partial}{\partial z} + \frac{\mathrm{i}k}{2}\boldsymbol{\phi}^2\right]\tilde{\mathcal{E}}_\omega(\boldsymbol{\phi}; z) = \sum_{j=1}^{N_\mathrm{e}} \frac{e\left[\boldsymbol{\beta}_j(z) - \boldsymbol{\phi}\right]}{4\pi\epsilon_0 c\lambda^2} \mathrm{e}^{\mathrm{i}k[ct_j(z)-z]} \times \int \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}k\boldsymbol{\phi}\cdot\boldsymbol{x}} \delta(\boldsymbol{x} - \boldsymbol{x}_j) \,. \tag{1}$$

Here, we have rewritten the angular dependence of the current so that we can replace the point-like electron source with a constant charge density in the transverse plane by making the replacement $\delta(\boldsymbol{x} - \boldsymbol{x}_j) \to \mathcal{A}_\mathrm{tr}^{-1}$, where $\mathcal{A}_\mathrm{tr}$ is the transverse area. Then, in the one-dimensional (1D) limit we have

$$\int \mathrm{d}\boldsymbol{x}\, \mathrm{e}^{-\mathrm{i}k\boldsymbol{\phi}\cdot\boldsymbol{x}} \delta(\boldsymbol{x} - \boldsymbol{x}_j) \to \frac{1}{\mathcal{A}_\mathrm{tr}} \int \mathrm{d}\boldsymbol{x} \mathrm{e}^{-\mathrm{i}k\boldsymbol{\phi}\cdot\boldsymbol{x}} = \frac{\lambda^2}{\mathcal{A}_\mathrm{tr}} \delta(\boldsymbol{\phi}) \,, \tag{2}$$

and the source is directed entirely in the forward direction. We complete the 1D limit by defining the 1D electric field $\tilde{E}_\omega(z)$ via

[1] from Kwang-Je Kim, Zhirong Huang, Ryan Lindberg, Synchrotron Radiation and Free-Electron Lasers (2017), © Cambridge University Press 2017, reproduced with permission. This material cannot be reproduced, shared, altered, or exploited commercially in any way without the permission of Cambridge University Press, as it is copyrighted material and therefore not subject to the allowances permitted by a CC licence.

https://doi.org/10.23730/CYRSP-2018-001.455

$$\tilde{\mathcal{E}}_\omega(\boldsymbol{\phi};z) = \tilde{E}_\omega(z)\delta(\boldsymbol{\phi}). \tag{3}$$

The $\delta(\boldsymbol{\phi})$ enforces the field to be in the forward direction only, which also implies that the spatial representation of the electric field is independent of $\boldsymbol{x}$. We insert the field Eq. (3) and the transverse electron velocity $\beta_{x,j} = (K/\gamma_j)\cos(k_u z)$ into Eq. (1), and then integrate over angles to find the 1D field equation

$$\frac{\partial}{\partial z}\tilde{E}_\nu(z) = \frac{1}{2\pi}\sum_{j=1}^{N_e}\frac{eK\cos(k_u z)}{4\epsilon_0 c\mathcal{A}_{\rm tr}\gamma_j}\mathrm{e}^{\mathrm{i}k[ct_j(z)-z]}. \tag{4}$$

Here $\nu = \frac{\omega}{\omega_1} = \frac{k}{k_1}$ is the dimensionless frequency. We have assumed that the field $\tilde{E}_\nu(z)$ describes a slowly varying envelope, so that consistency requires that we also identify the slowly varying current in Eq. (4). As discussed previously, we can do this by introducing the average particle time $\bar{t}_j = t_j - (K^2/8k_u\gamma^2)\sin(2k_u z)$ that subtracts off the oscillatory figure-eight component from t. In terms of the slowly varying ponderomotive phase, we have

$$\begin{aligned} k[ct_j(z) - z] &= \nu[\omega_1\bar{t}_j(z) - (k_1 + k_u)z] + \nu k_u z + \frac{\nu K^2}{4+2K^2}\sin(2k_u z) \\ &= -\nu\theta_j(z) + \Delta\nu k_u z + hk_u z + \nu\xi\sin(2k_u z), \end{aligned} \tag{5}$$

where we recall that h is an odd integer identifying the harmonic number, the normalized frequency difference $\Delta\nu \equiv \nu - h \equiv k/k_1 - h$, and we have introduced the shorthand notation $\xi \equiv K^2/(4+2K^2)$. Then, the wave equation Eq. (4) becomes

$$\begin{aligned} \frac{\partial}{\partial z}\tilde{E}_\nu(z) = -\frac{eK}{4\epsilon_0 c\mathcal{A}_{\rm tr}}\frac{1}{2\pi}\sum_{j=1}^{N_e}\frac{1}{\gamma_j}\mathrm{e}^{-\mathrm{i}\nu\theta_j(z)+\mathrm{i}\Delta\nu k_u z} \\ \times\left[\mathrm{e}^{\mathrm{i}(h-1)k_u z} + \mathrm{e}^{\mathrm{i}(h+1)k_u z}\right]\mathrm{e}^{\mathrm{i}\nu\xi\sin(2k_u z)}. \end{aligned} \tag{6}$$

The envelope $\tilde{E}_\nu(z)$, energy γ_j, and phase $\theta_j(z)$ all vary slowly over one undulator period, as does $\mathrm{e}^{\mathrm{i}\Delta\nu k_u z}$ if we restrict our attention to small frequency detunings, $|\Delta\nu| \ll 1$. We can extract the slowly varying terms from the second line of Eq. (6) by averaging over an undulator period λ_u as follows:

$$\begin{aligned} \frac{1}{\lambda_u}\int_0^{\lambda_u} dz\left[\mathrm{e}^{\mathrm{i}(h-1)k_u z} + \mathrm{e}^{\mathrm{i}(h+1)k_u z}\right]\mathrm{e}^{\mathrm{i}\nu\xi\sin(2k_u z)} \\ = J_{-(h-1)/2}(\nu\xi) + J_{-(h+1)/2}(\nu\xi) \equiv [\mathrm{JJ}]_h, \end{aligned} \tag{7}$$

where we have used the Jacobi–Anger identity to evaluate the integral, from which we find the harmonic Bessel function factor $[\mathrm{JJ}]_h$.

We are now in a position to write the frequency domain wave equation for the 1D FEL. However, there are a few notational issues that we would like to simplify. First, we will find it convenient to have the temporal and frequency representations of the field be related by a Fourier transform with respect to the scaled frequency ν. To do this, we write

$$\begin{aligned} E_x(\boldsymbol{x},t;z) &= \int d\omega d\boldsymbol{\phi}\, \mathrm{e}^{-\mathrm{i}(\omega - k\boldsymbol{\phi}\cdot\boldsymbol{x})}\mathrm{e}^{\mathrm{i}kz}\tilde{\mathcal{E}}_\omega(\boldsymbol{\phi};z) = \int d\omega\, \mathrm{e}^{-\mathrm{i}(\omega - kz)}\tilde{E}_\omega(z) \\ &= \mathrm{e}^{-\mathrm{i}h(\omega_1 - k_1 z)}\int d\nu\, \mathrm{e}^{\mathrm{i}\Delta\nu\theta}ck_1\mathrm{e}^{-\mathrm{i}\Delta\nu k_u z}\tilde{E}_\omega(z). \end{aligned} \tag{8}$$

The integrand contains the slowly varying field, and we can both simplify this and eliminate the phase $\mathrm{e}^{\mathrm{i}\Delta\nu k_u z}$ from the source current in Eq. (6) by defining the phase-shifted electric field amplitude

$$E_\nu(z) = ck_1 \mathrm{e}^{\mathrm{i}\Delta\nu k_u z}\tilde{E}_\omega(z) . \tag{9}$$

Note that this phase shift must be retained even though $\Delta\nu \ll 1$, since we may also have $k_u z \gg 1$. Finally, the field equation for $E_\nu(z)$ is

$$\begin{aligned}\left[\frac{\partial}{\partial z} + \mathrm{i}\Delta k_u\right] E_\nu(z) &= \frac{ek_1 K[\mathrm{JJ}]_h}{4\epsilon_0\gamma_r \mathcal{A}_{\mathrm{tr}}} \frac{1}{2\pi} \sum_{j=1}^{N_\mathrm{e}} \mathrm{e}^{-\mathrm{i}\nu\theta_j(z)} \\ &= -\kappa_h n_e \frac{1}{N_\lambda} \sum_{j=1}^{N_\mathrm{e}} \mathrm{e}^{-\mathrm{i}\nu\theta_j(z)} .\end{aligned} \tag{10}$$

Here, N_λ is the number of electrons in one wavelength, and the harmonic coupling and electron volume density are, respectively,

$$\kappa_h \equiv \frac{eK[\mathrm{JJ}]_h}{4\epsilon_0\gamma_r}, \qquad n_e \equiv \frac{I/ec}{\mathcal{A}_{\mathrm{tr}}} \equiv \frac{N_\lambda}{\lambda_1 \mathcal{A}_{\mathrm{tr}}} . \tag{11}$$

Note that while approximating γ_j by γ_r in k_h is a very good approximation, such a replacement in the particle phase would eliminate the FEL interaction entirely.

Equation (10) is in frequency domain, and is the most convenient to analytically study the FEL dynamics [1]. There are situations in which the time domain approach is more useful. The time domain equations are also well-suited for efficient numerical simulation codes. In the rest of this paper we will use the time domain formulation to obtain some basic understanding of the high-gain behaviour and its scalings.

The time domain wave equation basically follows from the inverse Fourier transform of Eq. (10). To make this connection explicit, we use the definitions Eq. (8) and Eq. (9) to find that 1D slowly varying envelopes are related by the Fourier transforms

$$E(\theta;z) = \int \mathrm{d}\nu\, \mathrm{e}^{\mathrm{i}\Delta\nu\theta} E_\nu(z) , \qquad E_\nu(z) = \frac{1}{2\pi}\int \mathrm{d}\theta\, \mathrm{e}^{-\mathrm{i}\Delta\nu\theta} E(\theta;z) . \tag{12}$$

Therefore, multiplying Eq. (10) by $\mathrm{e}^{\mathrm{i}\Delta\nu\theta}$ and integrating over ν yields

$$\left[\frac{\partial}{\partial z} + k_u \frac{\partial}{\partial\theta}\right] E(\theta;z) = -\kappa_1 n_e \frac{2\pi}{N_\lambda} \sum_{j=1}^{N_e} \mathrm{e}^{-\mathrm{i}\theta_j(z)} \delta\left[\theta - \theta_j(z)\right] , \tag{13}$$

at the fundamental frequency ω_1.

It may appear that our work is done, but the transverse current in Eq. (13) is composed of a sum of delta functions that is unfortunately both difficult to treat and apparently in violation of our assumption that E varies slowly. To establish a well-defined, slowly varying current, we average Eq. (13) over some number of periods in θ. This 'slice-averaging' has the same physical significance as our previous assumption that $|\Delta\nu| \ll 1$, and is valid provided the averaging time is much shorter than the characteristic time over which the field amplitude changes. For a high-gain FEL we require the averaging time Δt to be much less than the coherence time, $\Delta t = \Delta\theta/\omega_h \ll t_{\mathrm{coh}}$, which at the fundamental frequency reduces to $\Delta t \ll \lambda_1/(4\pi c\rho)$ or $\Delta\theta \ll 1/2\rho$. The time window over which the beam average is taken is sometimes referred to as an FEL slice.

We average Eq. (13) over an FEL slice by applying

$$\frac{1}{\Delta\theta}\int_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} \mathrm{d}\theta' \Bigg|_{\text{at fixed } z} \tag{14}$$

to both sides. Averaging the left-hand side of Eq. (13) leaves it unchanged since it is slowly varying, while applying Eq. (14) to the right-hand side picks out those electrons whose ponderomotive phase θ_j is within the interval $\theta - \Delta\theta/2$ and $\theta + \Delta\theta/2$. In other words, the source for $E(\theta)$ includes the $N_\Delta = N_\lambda(\Delta\theta/2\pi)$ electrons that satisfy $|\theta_j - \theta| \leq \Delta\theta/2$ when they arrive at location z. Then, we find that the wave equation in the time domain is

$$\left[\frac{\partial}{\partial z} + k_u \frac{\partial}{\partial\theta}\right] E(\theta; z) = -\kappa_1 n_e \frac{1}{N_\Delta}\sum_{j\epsilon\Delta} \mathrm{e}^{-\mathrm{i}\theta_j(z)} \tag{15}$$

$$= -\kappa_1 n_e \langle \mathrm{e}^{-\mathrm{i}\theta_j(z)} \rangle_\Delta \, . \tag{16}$$

The notation in Eq. (15) denotes that we are to sum over the N_Δ particles within the FEL slice at position z and phase θ. Hence, the average $\langle \mathrm{e}^{-\mathrm{i}\theta_j} \rangle_\Delta$, which is often referred to as the local bunching factor (or just bunching factor), is a function of both z and θ. For any given z, the bunching factor quantifies the spectral content of the current near the fundamental frequency by a complex number whose magnitude is between 0 and 1[2]. Finally, we note while that the Maxwell equation in the time and frequency domain look quite similar, they differ as follows: the driving current in Fourier version Eq. (10) is a sum over all electrons with the phase $\mathrm{e}^{-\mathrm{i}\nu\theta_j}$, while the time domain Eq. (15) sums only over those electrons within the FEL time slice using the phase $\mathrm{e}^{-\mathrm{i}\theta}$.

2 FEL equations and energy conservation

All equations governing the 1D FEL in the time domain are as follows:

$$\left[\frac{\partial}{\partial z} + k_u \frac{\partial}{\partial\theta}\right] E(\theta; z) = -\kappa_1 n_e \langle \mathrm{e}^{-\mathrm{i}\theta_j} \rangle_\Delta \, , \tag{17}$$

$$\frac{\mathrm{d}\theta_j}{\mathrm{d}z} = 2k_u\eta_j \, , \tag{18}$$

$$\frac{\mathrm{d}\eta_j}{\mathrm{d}z} = \chi_1\left(E\mathrm{e}^{\mathrm{i}\theta_j} + E^*\mathrm{e}^{-\mathrm{i}\theta_j}\right) , \tag{19}$$

with

$$\kappa_1 \equiv \frac{\mathrm{e}K[\mathrm{JJ}]}{4\epsilon_0\gamma_r} \, , \qquad \chi_1 \equiv \frac{\mathrm{e}K[\mathrm{JJ}]}{2\gamma_r^2 mc^2} \, . \tag{20}$$

Equations (17) is Maxwell equation (the same as Eq. (16)) and Eqs. (18) and (19) are the pendulum equations describing the electron motion [1]. These equations conserve total (particle + field) energy. To show this, we first integrate the electromagnetic energy density u_{EM} over length and multiply the result by the transverse area $\mathcal{A}_{\mathrm{tr}}$ to obtain the field energy

$$U_{\mathrm{EM}} = \frac{\mathcal{A}_{\mathrm{tr}}\lambda_1}{2\pi}\int \mathrm{d}\theta \; u_{\mathrm{EM}} = \frac{\mathcal{A}_{\mathrm{tr}}\lambda_1}{2\pi}\int \mathrm{d}\theta \frac{\epsilon_0}{2}(\boldsymbol{E}^2 + c^2\boldsymbol{B}^2)$$

[2] Harmonic generalizations of the bunching factor can also be defined as $b_h \equiv \langle \mathrm{e}^{-\mathrm{i}h\theta_j} \rangle_\Delta$.

$$= \frac{\mathcal{A}_{\text{tr}}\lambda_1}{2\pi}\int \mathrm{d}\theta\ 2\epsilon_0|E|^2 \ . \tag{21}$$

Hence, an equation for the electromagnetic field energy can be obtained by multiplying (Eq. (17)) by $(\mathcal{A}_{\text{tr}}\lambda_1/\pi)\epsilon_0 E^*$, adding the complex conjugate, and integrating over θ; we find that

$$\frac{\mathrm{d}}{\mathrm{d}z}U_{\text{EM}} = -\frac{\mathrm{e}K[\text{JJ}]}{2\gamma_r}\frac{N_\lambda}{2\pi N_\Delta}\int \mathrm{d}\theta \sum_{j\epsilon\Delta} E^*\mathrm{e}^{-\mathrm{i}\theta j} + c.c.$$

$$= -\frac{\mathrm{e}K[\text{JJ}]}{2\gamma_r}\sum_j \frac{\mathrm{e}^{-\mathrm{i}\theta_j}}{\Delta\theta}\int_{\theta_j-\Delta\theta/2}^{\theta_j+\Delta\theta/2} \mathrm{d}\theta\ E^*(\theta) + c.c.$$

$$= -\frac{\mathrm{e}K[\text{JJ}]}{2\gamma_r}\sum_j E^*(\theta_j)\mathrm{e}^{-\mathrm{i}\theta_j} + c.c.\,, \tag{22}$$

where in the last line we assumed that $E(\theta)$ is constant over the length $\Delta\theta$; this assumption is required because of our slice averaging, but is not necessary if one uses the frequency representation (Eq. (10)) or the unaveraged Eq. (13). The change in the total kinetic energy is obtained by multiplying (Eq. (19)) by $\gamma_r mc^2$ and summing over all electrons,

$$\frac{\mathrm{d}}{\mathrm{d}z}U_{\text{KE}} = \frac{\mathrm{d}}{\mathrm{d}z}\sum_j \gamma_r(1+\eta_j)mc^2 = \frac{\mathrm{e}K[\text{JJ}]}{2\gamma_r}\sum_j E(\theta_j)\mathrm{e}^{\mathrm{i}\theta_j} + c.c. \tag{23}$$

Adding Eqs. (22) and (23) shows that energy is conserved:

$$\frac{\mathrm{d}}{\mathrm{d}z}[U_{\text{EM}} + U_{\text{EM}}] = \frac{\mathrm{d}}{\mathrm{d}z}\left[\sum_j \gamma_r n_j mc^2 + \frac{\mathcal{A}_{\text{tr}}\lambda_1}{2\pi}\int \mathrm{d}\theta\ 2\epsilon_0|E(\theta;z)|^2\right] = 0\,. \tag{24}$$

3 Dimensionless FEL scaling parameter ρ

By expressing the governing equations of physical systems in terms of dimensionless quantities, one can identify important time and length scales and characterize the relevant magnitudes of the physical variables. In this section we cast the FEL equations into dimensionless form and find the fundamental scaling parameter ρ. We will subsequently see that ρ, which is also called the Pierce parameter, characterizes most properties of a high-gain FEL, while the dimensionless beam and radiation variables will give us some sense of the dynamics without any additional computation.

We introduce the as-yet-unspecified parameter ρ by defining the scaled longitudinal coordinate $\hat{z} \equiv 2k_u\rho z$ that leads to the phase equation

$$\frac{\mathrm{d}\theta_j}{\mathrm{d}\hat{z}} = \hat{\eta}_j \text{ for } \hat{\eta}_j \equiv \frac{\eta_j}{\rho} \quad \text{(the new 'momentum' variable).} \tag{25}$$

To simplify the energy equation for $\hat{\eta}_j$, we define the dimensionless complex field amplitude

$$a = \frac{\chi_1}{2k_u\rho^2}E\,, \tag{26}$$

in terms of which the energy equation reduces to

$$\frac{\mathrm{d}\hat{\eta}_j}{\mathrm{d}\hat{z}} = a(\theta_j,\hat{z})\mathrm{e}^{\mathrm{i}\theta_j} + a(\theta_j,\hat{z})^*\mathrm{e}^{-\mathrm{i}\theta_j}\,. \tag{27}$$

Writing the field Eq. (17) in terms of $\hat{z}$ and a, we have

$$\left[\frac{\partial}{\partial \hat{z}} + \frac{1}{2\rho}\frac{\partial}{\partial \theta}\right] a(\theta, \hat{z}) = -\frac{\chi_1}{2k_u\rho^2}\frac{n_e\kappa_1}{2k_u\rho}\langle \mathrm{e}^{-\mathrm{i}\theta_j}\rangle_\Delta \,. \tag{28}$$

To simplify the field equation, we choose to set the coefficient on the right-hand side of Eq. (28) to unity. Thus, the dimensionless Pierce parameter ρ must be [1]

$$\rho = \left[\frac{n_e\kappa_1\chi_1}{(2k_u)^2}\right]^{1/3} = \left(\frac{e^2K^2[\mathrm{JJ}]^2 n_e}{32\epsilon_0\gamma_r^3 mc^2 k_u^2}\right)^{1/3}$$
$$= \left[\frac{1}{8\pi}\frac{I}{I_\mathrm{A}}\left(\frac{K[\mathrm{JJ}]}{1+K^2/2}\right)^2 \frac{\gamma\lambda_1^2}{2\pi\sigma_x^2}\right]^{\frac{1}{3}}, \tag{29}$$

where $I_\mathrm{A} = ec/r_e = 4\pi\epsilon_0 mc^3/e \approx 17045$ A is the Alfvén current and we have set the cross-sectional area of the electron beam $\mathcal{A}_\mathrm{tr} \to 2\pi\sigma_x^2$ assuming a Gaussian transverse profile.

The scaled FEL equations have all coefficients unity, so that the dimensionless form allows one to make a number of order-of-magnitude estimates regarding the dynamics. First, one may *a priori* expect that the scaled variation $\mathrm{d}/\mathrm{d}\hat{z} \lesssim 1$. Thus, in the exponential growth regime we may anticipate the 1D gain length $L_{G0}{\sim}(2k_u\rho)^{-1}$. Additionally, since resonant energy exchange proceeds if the ponderomotive phase is nearly constant, this implies that saturation of the FEL interaction occurs when the scaled energy deviation $\hat{\eta}_j \sim 1$ (or $\eta_j \sim \rho$). At this point we expect that the bunching will approach its maximum value $\left|\langle \mathrm{e}^{-\mathrm{i}\theta_j}\rangle_\Delta\right| \to 1$, which in turn implies that the maximum scaled amplitude of the radiation $|a| \sim 1$. Furthermore, if we had included the transverse derivatives in the wave equation we would expect

$$\frac{1}{4k_u k_1\rho}\boldsymbol{\nabla}_\perp^2 \to 1\,. \tag{30}$$

Identifying the transverse Laplacian with the radiation size via $\boldsymbol{\nabla}_\perp^2 \sim 1/\sigma_r^2$ we find that the RMS mode size of the laser is roughly given by

$$\sigma_r \sim \sqrt{\frac{\lambda_1}{4\pi}\frac{\lambda_u}{4\pi\rho}}\,. \tag{31}$$

While these arguments are heuristic, they give useful predictions of FEL performance. Besides the observation that the gain length is approximately $\lambda_u/4\pi\rho$, we use the definition Eq. (26) to translate the scaled radiation amplitude $|a| \to 1$ at saturation to $|E| \to 2k_u\rho^2/\chi_1$, so that the maximum field energy density

$$2\epsilon_0|E|^2 \sim 2\epsilon_0\rho\frac{4k_u\rho^3}{\chi_1^2} = 2\epsilon_0\rho\frac{\kappa_1}{\chi_1} = \rho n_e\gamma_r mc^2\,. \tag{32}$$

Because $n_e mc^2\gamma_r$ is the electron energy density, we see that ρ also gives the FEL efficiency at saturation:

$$\rho = \frac{\text{field energy generated}}{\text{e-beam kinetic energy}}\,. \tag{33}$$

To determine the distance at which the FEL gain saturates and $P \sim \rho P_\mathrm{beam}$, we consider the motion of the electron in the pendulum potential. The period of motion is characterized by the synchrotron wavenumber

$$\Omega_s \equiv \sqrt{\frac{eE_0k_uK[\mathrm{JJ}]}{\gamma^2mc^2}} = 2\rho k_u|2a_0|^{1/2}\,, \tag{34}$$

and that the radiation field gains or loses energy depending on the oscillation phase of the particles. Since the energy exchange to the radiation ends when most of the particles make one-half oscillation in the ponderomotive bucket, we have $\langle\Omega_s\rangle z_{\text{sat}} \approx \pi$, where $\langle\Omega_s\rangle$ is the average value of the synchrotron wavenumber over the FEL length z_{sat}. Taking $\langle\Omega_s\rangle$ to be one-quarter of its maximum value at saturation where $|a_0| \sim 1$, we have $\rho k_u z_{\text{sat}}/\sqrt{2} \sim \pi$, or $z_{\text{sat}} \sim \lambda_u/\rho$. It is interesting to note that the power saturates when the synchrotron wavenumber is roughly equal to the exponential growth rate,

$$P \sim \rho P_{\text{beam}} \Leftrightarrow \Omega_s \sim 2\rho k_u \,. \tag{35}$$

This is to be expected, since when $\Omega_s \sim 2\rho k_u$ the particles can rotate to the accelerating phase of the potential during one growth length, in which case they then extract energy from the field.

Therefore, the FEL (or Pierce) parameter ρ determines the main characteristics of high-gain FEL systems, including the following.

1. Gain length $\sim \lambda_u/4\pi\rho$.
2. Saturation power $\sim \rho\times$ (e-beam power).
3. Saturation length $_{\text{sat}} \sim \lambda_u/\rho$.
4. Transverse mode size $\sigma_r \sim \sqrt{\lambda_1\lambda_u/16\pi^2\rho}$.

In the following sections we will analyse the FEL equations and demonstrate that the dynamics indeed exhibit these simple scalings.

4 1D solution using collective variables

In this section, we illustrate the essentials of FEL gain by neglecting the θ dependence of the electromagnetic field. This ignores the propagation (slippage) of the radiation, and is equivalent to assuming that a has only one frequency component. This model will be useful to illustrate the basic physics of the electron beam and radiation field in a high-gain device, but will be insufficient to fully understand the spectral properties of self-amplified spontaneous emission (SASE). A more rigorous discussion of SASE can be found in literature [1]. The 1D FEL equations ignoring radiation slippage are as follows

$$\frac{d\theta_j}{d\hat{z}} = \hat{\eta}_j, \tag{36}$$

$$\frac{d\hat{\eta}_j}{d\hat{z}} = a e^{i\theta_j} + a^* e^{-i\theta_j}, \tag{37}$$

$$\frac{da}{d\hat{z}} = -\langle e^{-i\theta_j}\rangle_\Delta \,. \tag{38}$$

These are $2N_\Delta + 2$ coupled first-order ordinary differential equations, $2N_\Delta$ for the particles, and 2 for the complex amplitude a. In general, these can only be solved via computer simulation. However, the system can be linearized in terms of three collective variables as in Ref. [2]:

$$\begin{aligned} &a && \text{(field amplitude);}\\ &b = \langle e^{-i\theta_j}\rangle_\Delta && \text{(bunching factor);}\\ &P = \langle \hat{\eta}_j e^{-i\theta_j}\rangle_\Delta && \text{(collective momentum) .} \end{aligned}$$

The equations of motion for the bunching b and the field amplitude a follow directly from Eqs. (36) and (38). Differentiating the collective momentum yields

$$\frac{dP}{d\hat{z}} = \left\langle \frac{d\hat{\eta}_j}{d\hat{z}} e^{-i\theta_j} \right\rangle - i\langle \hat{\eta}_j^2 e^{-i\theta_j} \rangle = a + a^* \langle e^{-2i\theta_j} \rangle - i\langle \hat{\eta}_j^2 e^{-i\theta_j} \rangle . \tag{39}$$

Note that Eq. (39) contains additional field variables, and the resulting system of equations is not closed. Nevertheless, these other terms are nonlinear, which we therefore expect to result in negligible higher-order corrections when a, b, and P are much smaller than unity before saturation. Thus, linearizing Eq. (39) and including the equations for b and a from Eqs. (36) and (38) yields the following closed system in the small-signal regime:

$$\frac{da}{d\hat{z}} = -b \qquad \text{bunching produces coherent radiation}\,, \tag{40a}$$

$$\frac{db}{d\hat{z}} = -iP \qquad \text{energy modulation becomes density bunching}, \tag{40b}$$

$$\frac{dP}{d\hat{z}} = a \qquad \text{coherent radiation drives energy modulation}\,. \tag{40c}$$

These are three coupled first-order equations, which can be reduced to a single third- order equation for a as

$$\frac{d^3 a}{d\hat{z}^3} = ia\,. \tag{41}$$

We solve the linear equation by assuming that the field dependence is $\sim e^{-i\mu\hat{z}}$, which results in the following dispersion relation for μ:

$$\mu^3 = 1\,. \tag{42}$$

This is the well-known cubic equation, whose three roots are given by

$$\mu_1 = 1\,, \qquad \mu_2 = \frac{-1-\sqrt{3}i}{2}\,, \qquad \mu_3 = \frac{-1+\sqrt{3}i}{2}\,. \tag{43}$$

The root μ_1 is real and gives rise to an oscillatory solution, while μ_2 and μ_3 are complex conjugates that lead to exponentially decaying and growing modes, respectively. Furthermore, the roots obey

$$\sum_{\ell=1}^{3} \mu_\ell = 0, \qquad \sum_{\ell=1}^{3} \frac{1}{\mu_\ell} = \sum_{\ell=1}^{3} \mu_\ell^* = \sum_{\ell=1}^{3} \mu_\ell^2 = 0\,, \tag{44}$$

and the general solution to Eq. (41) is composed of a linear combination of the exponential solutions:

$$a(\hat{z}) = \sum_{\ell=1}^{3} C_\ell e^{-i\mu_\ell \hat{z}}\,. \tag{45}$$

The three constants C_ℓ are determined from the initial conditions $a(0)$, $b(0)$, and $P(0)$. By differentiating the expression for a and using Eq. (40), we find

$$a(0) = C_1 + C_2 + C_3\,, \tag{46}$$

$$\left.\frac{da}{d\hat{z}}\right|_0 = -b(0) = -i[\mu_1 C_1 + \mu_2 C_2 + \mu_3 C_3]\,, \tag{47}$$

$$\left.\frac{d^2 a}{d\hat{z}^2}\right|_0 = iP(0) = -[\mu_1^2 C_1 + \mu_2^2 C_2 + \mu_3^2 C_3]\,. \tag{48}$$

Using Eq. (44), this yields the electromagnetic field evolution as

$$a(\hat{z}) = \frac{1}{3}\sum_{\ell=1}^{3}\left[a(0) - \mathrm{i}\frac{b(0)}{\mu_\ell} - \mathrm{i}\mu_\ell P(0)\right]\mathrm{e}^{-\mathrm{i}\mu_\ell \hat{z}} \,. \tag{49}$$

The general solution for the radiation requires all three roots of μ. For long propagation distances, however, the relative importance of the oscillating root μ_1 and decaying root μ_2 becomes insignificant in comparison with the growing solution associated with μ_3. Thus, the radiation field is completely characterized by μ_3 in the exponential growth regime where $\hat{z} \gg 1$, so that

$$a(\hat{z}) \approx \frac{1}{3}\left[a(0) - \mathrm{i}\frac{b(0)}{\mu_3} - \mathrm{i}\mu_3 P(0)\right]\mathrm{e}^{-\mathrm{i}\mu_3 \hat{z}} \,. \tag{50}$$

The first term in the bracket describes the coherent amplification of an external radiation signal, while the second and the third term show how modulations in the electron beam density and energy may lead to FEL output. When the source of these modulations is the electron beam shot noise then the exponential growth is considered to be SASE.

5 Qualitative description of SASE

SASE results from the FEL amplification of the initially incoherent spontaneous undulator radiation, Refs. [2, 3, 4]. It is of primary importance for FEL applications in wavelength regions where mirrors (and, hence, oscillator configurations) are unavailable.

For our first look at SASE, we use the formula for the radiation in the high-gain regime Eq. (50) assuming that there is no external field $a(0) = 0$ and that the beam has vanishing energy spread with $P(0) = 0$. In this case, the radiation intensity in the exponential growth regime is

$$\langle |a(\hat{z})|^2\rangle \approx \frac{1}{9}\langle |b(0)|^2\rangle \mathrm{e}^{\sqrt{3}\hat{z}} \,. \tag{51}$$

Here, the scaled propagation distance $\sqrt{3}\hat{z} = \sqrt{3}(2k_u z\rho) = z/L_{G0}$, and the ideal 1D power gain length is

$$L_{G0} \equiv \frac{\lambda_u}{4\pi\sqrt{3}\rho} \,. \tag{52}$$

The bunching factor at the undulator entrance $\langle |b(0)|^2\rangle$ derives from the initial shot noise of the beam, which is subsequently amplified by the FEL process. This level of shot noise is determined by the number of particles in the radiation coherence length, and it can be shown that

$$\langle |b(0)|^2\rangle = \left\langle \frac{1}{N_{l_\mathrm{coh}}^2}\left|\sum_{j\epsilon l_\mathrm{coh}} \mathrm{e}^{-\mathrm{i}\theta_j}\right|^2\right\rangle \approx \frac{1}{N_{l_\mathrm{coh}}} \,, \tag{53}$$

where N_{l_coh} is the number of electrons in a coherence length l_coh. It turns out that the normalized bandwidth of SASE is $\Delta\omega/\omega \sim \rho$, so that the coherence time $t_\mathrm{coh} \sim \lambda_1/c\rho$ and the coherence length $l_\mathrm{coh} \sim \lambda_1/\rho$ [2]. Alternatively, one can recognize the coherence length as the amount the radiation slips ahead of the electron beam in a few gain lengths. Hence, the start-up noise of a SASE FEL is characterized by

$$N_{l_\mathrm{coh}} \sim \frac{I}{ec}\frac{\lambda_1}{\rho} \,. \tag{54}$$

Figure 1 is a schematic plot that illustrates the initial start-up, exponential growth, and saturation of a SASE FEL. As is clear from the figure and from the previous discussion, ρ plays a fundamental role in the high-gain FEL physics for SASE. While we have not yet derived all the radiation properties, some of the important ones include:

1. saturation length $L_{\mathrm{sat}} \sim \lambda_u/\rho$;
2. output power $\sim \rho \times P_{\mathrm{beam}}$;
3. frequency bandwidth $\Delta\omega/\omega \sim \rho$;
4. 1D power gain length $L_{G0} = \lambda_u/(4\pi\sqrt{3}\rho)$;
5. transverse coherence: radiation emittance $\varepsilon_r = \lambda/4\pi$;
6. transverse mode size: $\sigma_{\mathrm{r}} \sim \sqrt{\varepsilon_r L_{G0}}$;
7. for the SASE power $P = P_{\mathrm{in}} \exp(z/L_G)$, the effective noise $P_{\mathrm{in}} \sim \rho\gamma mc^2/N_{l_{\mathrm{coh}}}$.

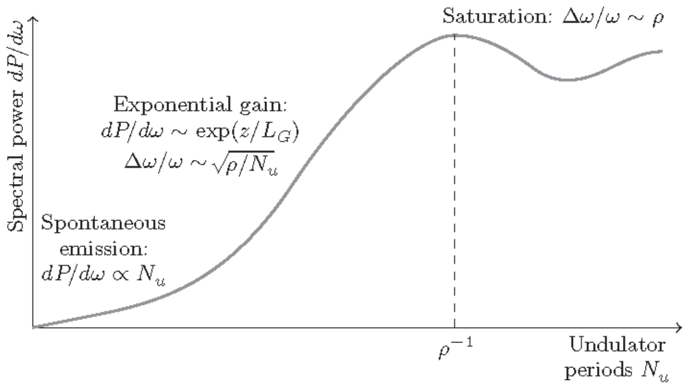


Fig. 1: Illustration of basic SASE processes. Adapted from Ref. [5]

While these basic scalings and the plot of Fig. 1 describes the ensemble averaged SASE properties, we should keep in mind that any individual SASE pulse is essentially amplified undulator radiation, and therefore has the same basic power and spectral fluctuations as the chaotic light discussed in "Temporal Coherence of Radiation Beam from a collection of Electrons" (previous lecture from these proceedings). We can understand the connection of SASE to amplified undulator radiation by considering the undulator energy as computed from the 1D power spectral density,

$$U_{\mathrm{und}} = T\int \mathrm{d}\omega\, \mathrm{d}\boldsymbol{\phi}\, \frac{\mathrm{d}P}{\mathrm{d}\omega\, \mathrm{d}\boldsymbol{\phi}} \xrightarrow{\mathrm{1D}} T\int\, d\omega \frac{\lambda^2}{\mathcal{A}_{\mathrm{tr}}} \frac{\mathrm{d}P}{\mathrm{d}\omega\, \mathrm{d}\boldsymbol{\phi}}\bigg|_{\boldsymbol{\phi}=0}\,, \tag{55}$$

where the quantity $\lambda^2/\mathcal{A}_{\mathrm{tr}}$ can be understood as the characteristic angular spread from a source of area $\mathcal{A}_{\mathrm{tr}}$: $\Delta\phi_x\Delta\phi_y \sim \lambda^2/\mathcal{A}_{\mathrm{tr}}$. In the 1D limit this tends to zero and we identify $\delta(\boldsymbol{\phi}) = \mathcal{A}_{\mathrm{tr}}/\lambda^2$, so that

$$\frac{\mathrm{d}P}{\mathrm{d}\omega}\bigg|_{\mathrm{1D}} = \frac{\lambda^2}{\mathcal{A}_{\mathrm{tr}}}\delta(\boldsymbol{\phi})\frac{\mathrm{d}P}{\mathrm{d}\omega} = \frac{\lambda^2}{\mathcal{A}_{\mathrm{tr}}}\frac{\mathrm{d}P}{\mathrm{d}\omega\boldsymbol{\phi}}\bigg|_{\boldsymbol{\phi}=0}\,. \tag{56}$$

The same factor $\lambda^2/\mathcal{A}_{\mathrm{tr}}$ appeared for the 1D limit in Eq. (2). Inserting the power density in the forward direction [1], we find that

$$\begin{aligned} U_{\mathrm{und}} &= T\left[\frac{\lambda_1^2}{\mathcal{A}_{\mathrm{tr}}}\frac{I}{I_A}\left(\frac{K[\mathrm{JJ}]}{1+K^2/2}\right)^2 \gamma_r^2 mc^2 N_u^2\right]\frac{\omega_1}{\pi N_u}\int\, \mathrm{d}x\left(\frac{\sin x}{x}\right)^2 \\ &= 8\pi\omega_1 T\gamma_r mc^2 N_u \rho^2 \;\to 8\pi\omega_1 T\gamma_r mc^2\rho^2 \end{aligned} \tag{57}$$

at the FEL saturation distance $N_u \approx 1/\rho$. Now, we use Eq. (57) to rewrite the FEL energy at saturation as

$$U_{\mathrm{FEL}} = N_{\mathrm{e}}\rho\gamma_r mc^2 = \frac{N_{\mathrm{e}}}{\rho\omega_1 T}\frac{U_{\mathrm{und}}}{8\pi} \sim \frac{t_{\mathrm{coh}}N_{\mathrm{e}}}{T}U_{\mathrm{und}} = N_{l_{\mathrm{coh}}}U_{\mathrm{und}} \tag{58}$$

$$= \frac{T}{t_{\mathrm{coh}}}N_{l_{\mathrm{coh}}}^2\frac{U_{\mathrm{und}}}{N_{\mathrm{e}}}\,. \tag{59}$$

The first line Eq. (58) shows that in the forward direction the FEL output at saturation is larger than that of the undulator radiation by the number of particles in a coherence time $N_{l_{\mathrm{coh}}} \gtrsim 10^5$. The second result Eq. (59) interprets the FEL energy as being proportional to the undulator field energy due to a single electron times the square of the number of electrons in a coherence length times the number of coherent regions T/t_{coh}.

Finally, we would like to emphasize that X-ray FELs based on SASE would not have been realized without incredible improvements in the production, transport, and manipulation of electron beams, since very high brightness electron beams are essential for X-ray FELs. In particular, SASE FELs have been made possible through recent advances in photocathode gun design (see Ref. [6] and a review in Ref. [7]), and tremendous improvements of radiofrequency linac and undulator technology. These advances have made it possible to produce sufficient gain in the undulator for transversely coherent radiation, meaning that the electron beam meets the following criteria:

1. energy spread $\Delta\gamma/\gamma < \rho$;
2. emittance $\varepsilon_x \lesssim \lambda/(4\pi)$;
3. beam size $\sigma_x \gtrsim \sigma_r \sim \sqrt{\frac{\lambda}{4\pi}\frac{\lambda_u}{4\pi\rho}}$ to have 1D scalings approximately apply;
4. high peak current to achieve $\rho \sim 10^{-3}$ and, hence, a reasonable saturation length and power efficiency.

References

[1] See, for example, K.-J. Kim, Z. Huang, and R. Lindberg, “*Synchrotron Radiation and Free-Electron Lasers*,” Cambridge University Press (2017).

[2] R. Bonifacio, C. Pellegrini and L.M. Narducci, *Opt. Commun.* **50** (1984) 373. https://doi.org/10.1016/0030-4018(84)90105-6

[3] A. Kondratenko and E. Saldin, *Part. Accel.* **10** (1980) 207.

[4] Y.S. Derbenev, A.M. Kondratenko and E.L. Saldin, *Nucl. Instrum. Methods Phys. Res.* **193** (1982) 415. https://doi.org/10.1016/0029-554x(82)90233-6

[5] K.-J. Kim, *Phys. Rev. Lett.* **57** (1986) 1871. https://doi.org/10.1103/PhysRevLett.57.1871

[6] J.S. Fraser, R.L. Sheffield and E.R. Gray, *Nucl. Instrum. Methods Phys. Res., Sect. A* **250** (1986) 71. https://doi.org/10.1016/0168-9002(86)90862-4

[7] I. Ben-Zvi, in *Accelerator Physics, Technology, and Applications*, Eds. A.W. Chao, H.O. Moser and Z. Zhao (World Scientific, 2004), p. 158. https://doi.org/10.1142/9789812702807_0007

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

The Quantum Free-Electron Laser

Gordon R.M. Robb
University of Strathclyde & Scottish Universities Physics Alliance (SUPA), John Anderson Building, 107 Rottenrow, Glasgow, G4 0NG, UK

Abstract
The quantum regime of the free-electron laser (FEL) interaction, where the recoil associated with photon emission plays a significant role, is discussed. The role of quantum effects and their relation to electron beam coherence are considered. An outline derivation of a 1D quantum high-gain FEL model and some of its predictions for the behaviour of a quantum FEL in the linear and non-linear regimes are presented. The effect of slippage and, consequently, the quantum regime of self-amplified spontaneous emission are discussed. Suggestions on how to realise a quantum FEL are presented and some recent related work is summarized.

Keywords
Free-electron laser; quantum; recoil.

1 Introduction

Free-electron lasers (FELs) have demonstrated the ability to produce coherent radiation ranging from the microwave to the X-ray region of the electromagnetic spectrum. Although the first theoretical studies of the FEL involved a quantum mechanical analysis (e.g., [1]), it is generally accepted that FEL experiments *to date* are well described by classical models in which an electron beam consisting of classical particles interacts with a classical electromagnetic radiation field. Until relatively recently, most theoretical studies using a quantum analysis have been restricted to the regime of low gain (e.g., [2–6]). More recently, however, there has been a revival of interest in the role of quantum effects in the FEL interaction (e.g., [7–9]), stimulated in part by the development of short-wavelength, high-gain FEL amplifiers producing progressively higher-energy photons (see, e.g., Ref. [10] for a review of X-ray FELs). In this article, methods for describing the quantum regime of high-gain FEL operation are outlined and some of the effects which are beyond description by the usual classical models are discussed.

2 The role of quantum effects in the free-electron laser interaction

2.1 Recap of classical high-gain free-electron laser theory

In this section, a brief summary of 1D classical high-gain FEL theory is presented, which will be used both as a benchmark with which to compare the results from the quantum model, and to illustrate where the limitations of classical FEL models occur.

Classical high-gain FEL models (e.g., [11–13]) describe the self-consistent evolution of a collection of relativistic electrons interacting with an electromagnetic radiation field as they propagate through an undulator/wiggler magnet. The Newton–Lorentz equations of motion describing the dynamics of each electron and Maxwell's wave equation can be written in a dimensionless, universally scaled form where the number of free parameters is minimized [11]:

$$\frac{\mathrm{d}\theta_j}{\mathrm{d}\bar{z}} = \bar{p}_j \, , \tag{1}$$

$$\frac{\mathrm{d}\bar{p}_j}{\mathrm{d}\bar{z}} = -\left(\bar{A}\mathrm{e}^{\mathrm{i}\theta_j} + \mathrm{c.c.}\right) \, , \tag{2}$$

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.467

$$\frac{d\bar{A}}{d\bar{z}} = \left\langle e^{-i\theta} \right\rangle + i\delta\bar{A}\,, \tag{3}$$

where $\theta_j = (k + k_w)z - \omega t_j$ is the electron ponderomotive phase, $\bar{p}_j = (\gamma_j - \gamma_r)/\rho\gamma_r$ is the scaled energy/momentum change of each electron, $\bar{z} = z/L_g$ is the scaled position in the wiggler, $L_g = \lambda_w/4\pi\rho$ is the gain length, $\delta = (\gamma_0 - \gamma_r)/\rho\gamma_r$ is the scaled detuning of the beam energy from its resonance value, where γ_0 is the initial Lorentz factor of the electron beam, γ_r is the Lorentz factor of the electron beam at resonance, which satisfies the FEL resonance condition $\gamma_r = \sqrt{\lambda_w(1 + a_w^2)/2\lambda}$, where λ is the radiation wavelength and $a_w = eB_w/mck_w$ is the wiggler deflection parameter for a wiggler with magnetic field, B_w, and wiggler wavenumber, $k_w = 2\pi/\lambda_w$, where λ_w is the wiggler period. The dimensionless radiation field amplitude $\bar{A}$ is defined so that $|\bar{A}|^2 = \epsilon_0|E|^2/\rho\hbar\omega n_e$, where $|E|$ is the electric field amplitude, n_e is the electron beam density, $k = 2\pi/\lambda$ is the wavenumber of the radiation, $\omega = ck$ is the angular frequency of the radiation and c is the speed of light. The average, $\langle(...)\rangle \equiv \frac{1}{N}\sum_{j=1}^{N}(...)_j$, where $j = 1, \ldots, N$ is the electron index, and N is the number of electrons. The FEL parameter ρ is defined as

$$\rho = \frac{1}{\gamma_r}\left(\frac{a_w\omega_p}{4ck_w}\right)^{2/3}\,, \tag{4}$$

where $\omega_p = \sqrt{e^2 n_e/\epsilon_0 m}$ is the plasma frequency associated with the electron beam, m is the electron mass, and e is the magnitude of the electron charge. In deriving Eqs. (1)–(3), the relative slippage of the radiation with respect to the electron beam has been neglected.

Assuming an initial condition corresponding to equally distributed electron phases, a monoenergetic electron beam, and no electromagnetic field, i.e.,

$$\left\langle e^{-i\theta} \right\rangle = 0,\ \ \bar{p}_j = 0\ \forall j,\ \ \bar{A} = 0\,,$$

then a linear stability analysis [11] shows that this initial condition is unstable to fluctuations in the electromagnetic field amplitude and electron phase, i.e., shot noise. A numerical solution of Eqs. (1)–(3) with initial conditions

$$\left\langle e^{-i\theta} \right\rangle = 0\,,\ \ \bar{p}_j = 0\ \forall j\,,\ \ |\bar{A}| = 10^{-4}\,,\ \ \delta = 0$$

is shown in Fig. 1, showing exponential amplification of the electromagnetic field intensity and the bunching factor $|b| = |\langle e^{-i\theta}\rangle|$ before saturation, when $|\bar{A}|^2 \approx 1.4$ and $|b| \approx 0.8$. This high-gain field amplification therefore occurs simultaneously with the development of strong electron bunching on the scale of the radiation wavelength. Figure 2 shows the corresponding electron phase space distribution at different stages of the interaction (starting from $\bar{z} = 0$) and at saturation ($\bar{z} \approx 12$ here). It can be seen that during the interaction, the electron distribution becomes strongly modulated on the scale of the radiation wavelength λ.

2.2 Energy and momentum considerations

The FEL process fundamentally involves electrons emitting photons, with each photon carrying a finite momentum $\hbar k$. Consequently, in the relativistic limit in which the Lorentz factor of the electrons $\gamma \gg 1$ so that $k \gg k_w$, then each photon emission event will result in the electron recoiling, reducing its momentum by an amount $\hbar k$. The classical model of Section 2.1 neglects the discrete nature of this recoil process, and assumes that the interaction involves a continuous exchange of energy/momentum between the electrons and the radiation field, with the electrons moving along continuous trajectories in phase space as shown in Fig. 2. This assumption is valid as long as the momentum associated with each photon recoil is negligible compared with the momentum exchange between the electrons and the radiation field. A condition for this classical approximation can be derived by inspection of the electron beam phase space in Fig. 2.

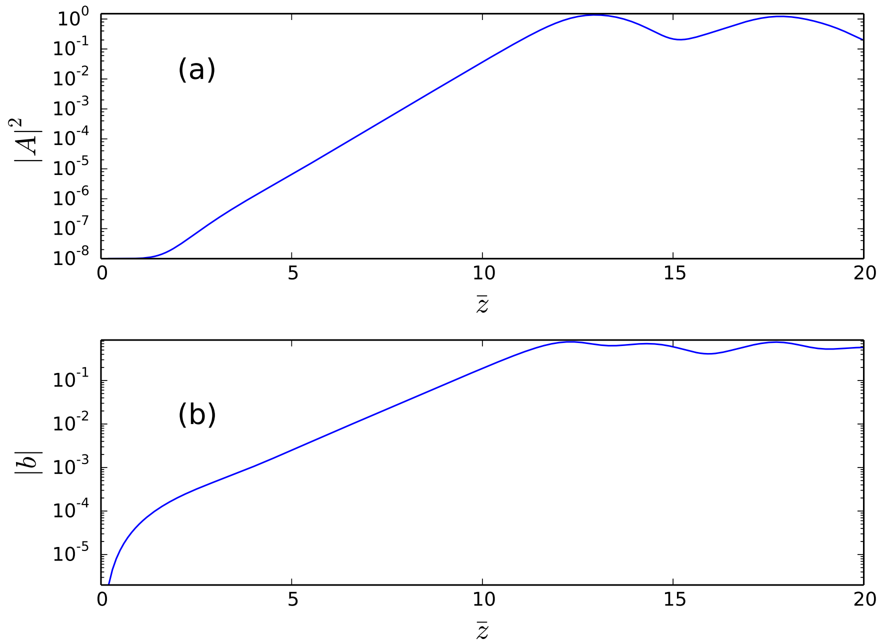


Fig. 1: Evolution of (a) radiation intensity, $|\bar{A}|^2$, and (b) bunching factor amplitude, $|b|$, in the classical FEL model of Eqs. (1)–(3) for an ideal, resonant ($\delta = 0$) monoenergetic electron beam.

It can be seen from Fig. 2 that, as a consequence of the FEL interaction, the initially monoenergetic electron beam acquires a spread in the scaled energy/momentum variable $\bar{p}$ of $\Delta\bar{p} \sim 1$. Consequently, from the definition of $\bar{p}$ above, this implies a relative energy spread $\Delta\gamma/\gamma_{\rm r} \sim \rho$, or a momentum spread $\Delta p = \Delta\gamma\, mc \sim \rho\gamma_{\rm r} mc$. The validity of the classical model will therefore depend on the relative size of this 'classical' momentum spread and the single-photon recoil $\hbar k$, i.e., the quantity

$$\frac{\Delta p}{\hbar k} = \frac{\gamma_{\rm r} mc}{\hbar k}\rho \equiv \bar{\rho}\,, \tag{5}$$

where $\bar{\rho}$ is the quantum FEL parameter [7]. The classical regime will therefore be valid when $\bar{\rho} = \Delta p/\hbar k \gg 1$, so that the discrete nature of each photon recoil is insignificant and the electron energy/momentum and position evolve along continuous trajectories in phase space as in Fig. 2. In the opposite limit, where $\bar{\rho} \geq 1$, this continuous classical picture breaks down owing to the finite momentum of each photon recoil being significant. Consequently, to describe this quantum FEL regime it is necessary to replace the particle–field model described by Eqs. (1)–(3) with a different model whose behaviour reduces to that of Eqs. (1)–(3) in the classical limit where $\bar{\rho} \gg 1$.

2.3 Electron beam coherence considerations

Although electron beams are usually described as moving collections of point-like classical charged particles, there are situations where electron beams display coherent phenomena which require a wave-like description; for example, it has been shown [14] that an electron beam which is split into two parts which follow different paths before being recombined produces an electron density interference pattern if the path difference involved is less than the electron beam (temporal) coherence length, defined as

$$L_{\rm e_c} = \frac{\lambda_{\rm e}^2}{\Delta\lambda_{\rm e}}\,, \tag{6}$$

where $\lambda_{\rm e} = h/p$ is the electron de Broglie wavelength and h is Planck's constant.

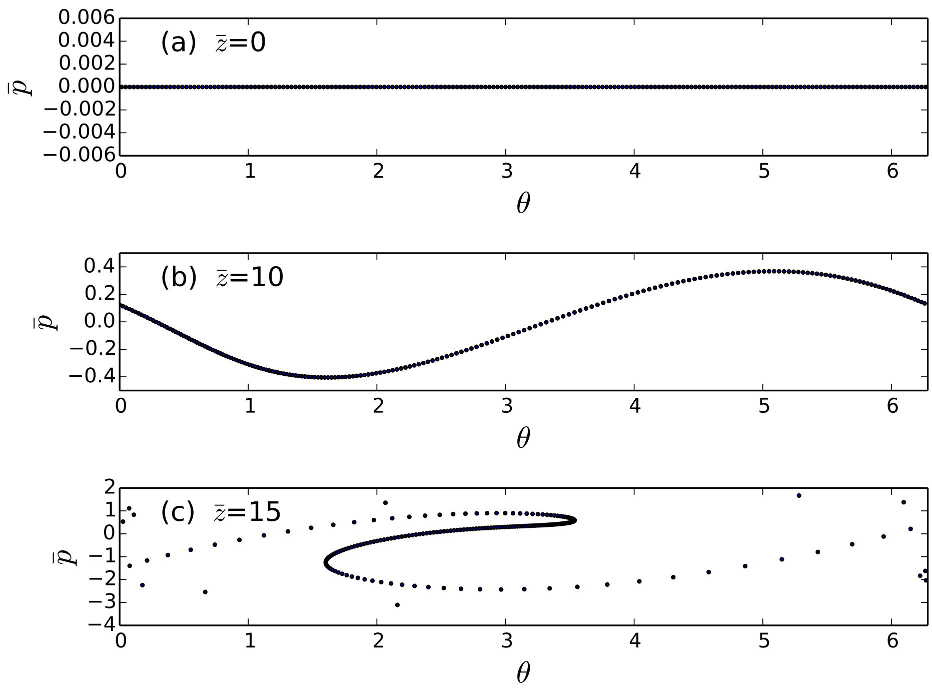


Fig. 2: Evolution of electron trajectories in phase space as described by the classical model of Eq. (1) for an ideal, resonant ($\delta = 0$) monoenergetic electron beam.

The interference patterns observed in Ref. [14] cannot be described in terms of electrons as particles, but can be described in terms of two wave functions, one for each part of the split electron beam, i.e., $\Psi_{1,2} = |\Psi_{1,2}|e^{i\phi_{1,2}}$, where $\phi_{1,2}$ are the phases of parts 1 and 2, respectively. The total electron density after recombination can therefore be written as

$$|\Psi|^2 = |\Psi_1 + \Psi_2|^2 = |\Psi_1|^2 + |\Psi_2|^2 + 2|\Psi_1||\Psi_2|\cos(\phi_1 - \phi_2) , \tag{7}$$

which displays interference.

For the FEL, it is expected that the wave-like nature of the electron beam should be significant during the FEL interaction if the electron beam coherence length becomes comparable with or exceeds the radiation wavelength λ, i.e., $L_{\mathrm{e_c}} \geq \lambda$. Rewriting $L_{\mathrm{e_c}}$ in terms of the electron momentum p from Eq. (6) produces

$$L_{\mathrm{e_c}} = \frac{h^2}{p^2}\frac{p^2}{h\,\Delta p} = \frac{h}{\Delta p} , \tag{8}$$

so $L_{\mathrm{c}} \geq \lambda$ implies $h/\Delta p \geq \lambda$, i.e., $\hbar k \geq \Delta p$, which is just the condition for the breakdown of the classical FEL model described in Section 2.2. Consequently, when $\hbar k \geq \Delta p$, i.e., $\bar{\rho} \geq 1$, the classical, particle model of the FEL interaction must be replaced with a wave function model, or its equivalent. The arguments above raise the interesting question of what would happen in the case of an FEL using a fully temporally coherent electron bunch [15] such that $L_{\mathrm{e_c}} > L_{\mathrm{e}}$, where L_{e} is the electron bunch length.

3 A one-dimensional quantum free-electron laser model

In this section, an outline derivation of a model capable of describing both the classical and the quantum regime of high-gain FEL operation is presented. More rigorous derivations can be found in [7, 9, 16].

3.1 Electron dynamics

It is possible to rewrite the pendulum-like Newton–Lorentz equations of motion shown in Eqs. (1) and (2) in the following slightly modified form, where the parameter $\bar{\rho}$ appears explicitly:

$$\frac{\mathrm{d}\theta_j}{\mathrm{d}\bar{z}} = \frac{\tilde{p}_j}{\bar{\rho}} , \tag{9}$$

$$\frac{\mathrm{d}\tilde{p}_j}{\mathrm{d}\bar{z}} = -\bar{\rho}\left(\bar{A}\mathrm{e}^{\mathrm{i}\theta_j} + \mathrm{c.c.}\right) , \tag{10}$$

where the dimensionless variable $\tilde{p} = \bar{\rho}\bar{p} \approx mc(\gamma - \gamma_0)/\gamma_{\mathrm{r}}$. Equations (9) and (10) can be written in the form of Hamilton's equations, i.e.,

$$\frac{\mathrm{d}\theta_j}{\mathrm{d}\bar{z}} = \frac{\partial H_j}{\partial \tilde{p}_j}, \quad \frac{\mathrm{d}\tilde{p}_j}{\mathrm{d}\bar{z}} = -\frac{\partial H_j}{\partial \theta_j} ,$$

in terms of the single-electron Hamiltonian function

$$H_j(\theta_j, \tilde{p}_j) = \frac{\tilde{p}_j^2}{2\bar{\rho}} - \mathrm{i}\bar{\rho}\left(\bar{A}\mathrm{e}^{\mathrm{i}\theta_j} - \mathrm{c.c.}\right) . \tag{11}$$

The Hamiltonian in Eq. (11) can in turn be used to write a Schrödinger equation of the form

$$\mathrm{i}\frac{\partial\Psi(\theta,\bar{z})}{d\bar{z}} = H_j\Psi(\theta,\bar{z})$$

for the single-electron wave function $\Psi(\theta, \bar{z})$, where $\tilde{p}$ is now treated as an operator, i.e., $\tilde{p} = -\mathrm{i}\partial/\partial\theta$, so that

$$\mathrm{i}\frac{\partial\Psi(\theta,\bar{z})}{\partial\bar{z}} = -\frac{1}{2\bar{\rho}}\frac{\partial^2\Psi}{\partial\theta^2} - \mathrm{i}\bar{\rho}\left(\bar{A}\mathrm{e}^{\mathrm{i}\theta} - \mathrm{c.c.}\right)\Psi, \tag{12}$$

where the index j has been dropped.

3.2 Radiation field dynamics

To describe the evolution of the radiation field in terms of the wave function $\Psi(\theta, \bar{z})$, the ensemble average in Eq. (3) is replaced by a corresponding average involving the probability distribution of the electron positions in the beam, so that Eq. (3) becomes

$$\frac{\mathrm{d}\bar{A}}{\mathrm{d}\bar{z}} = \int_0^{2\pi} |\Psi(\theta,\bar{z})|^2\mathrm{e}^{-\mathrm{i}\theta}\,\mathrm{d}\theta + \mathrm{i}\delta\bar{A} , \tag{13}$$

where the wave function $\Psi(\theta, \bar{z})$ is normalized such that

$$\int_0^{2\pi} |\Psi(\theta,\bar{z})^2|\,\mathrm{d}\theta = 1 .$$

The Maxwell–Schrödinger equations in Eqs. (12) and (13) together constitute a quantum model of the high-gain FEL which is valid for any value of $\bar{\rho}$.

3.3 Momentum state representation

The quantum FEL equations, Eqs. (12) and (13), can be solved directly numerically, but it is convenient to write them in terms of a set of discrete momentum state amplitudes rather than a spatially dependent wave function. To do this, the fact that the states $|n\rangle = \frac{1}{\sqrt{2\pi}}\exp(\mathrm{i}n\theta)$ are momentum eigenstates is used, where n is an integer. This means that $|n\rangle$ satisfies the eigenvalue equation

$$\hat{p}|n\rangle = n|n\rangle$$

where $\hat{p} = -\mathrm{i}\partial/\partial\theta$ is the momentum operator.

It is possible to expand the electron wave function $\Psi(\theta, \bar{z})$ in terms of these momentum eigenstates, i.e.,

$$\Psi(\theta, \bar{z}) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} c_n(\bar{z}) \exp(\mathrm{i}n\theta) , \tag{14}$$

where $|c_n|^2$ is the probability of an electron having momentum $(\gamma - \gamma_0)mc = n\hbar k$. Substituting for $\Psi(\theta, \bar{z})$ using Eq. (14) in Eqs. (12) and (13), the quantum FEL equations in the momentum state representation are

$$\frac{\mathrm{d}c_n(\bar{z})}{\mathrm{d}\bar{z}} = -\mathrm{i}\frac{n^2}{2\bar{\rho}} c_n - \bar{\rho}\left(\bar{A}c_{n-1} - \bar{A}^* c_{n+1}\right) , \tag{15}$$

$$\frac{\mathrm{d}\bar{A}(\bar{z})}{\mathrm{d}\bar{z}} = \sum_{n=-\infty}^{\infty} c_n c_{n-1}^* + \mathrm{i}\delta\bar{A} . \tag{16}$$

In this momentum state representation, the electron-beam–light interaction is described as an exchange of population between different electron momentum states via the electromagnetic field in discrete amounts $\hbar k$. The evolution of the electromagnetic field is driven by spatial bunching of electrons (as in Eq. (3) or (13)), which from Eq. (16) can be seen to be equivalent to coherence between adjacent momentum states.

3.4 Linear stability analysis

One of the advantages of using the momentum state representation given in the previous section is that it easily allows an analysis of the linear stability of stationary solutions, and consequently identification of the conditions under which instability can occur. A stationary solution of the quantum FEL equations in the momentum state representation, Eqs. (15) and (16), is $\bar{A} = 0$, i.e., no radiation field, and $c_0 = 1, c_m = 0\ \forall m \neq 0$, i.e., a spatially uniform electron distribution. Introducing small fluctuations in c_n and $\bar{A}$ about these stationary values, denoted by $c_n^{(1)}$ and $\bar{A}^{(1)}$ respectively, then

$$\begin{aligned} \bar{A} &= 0 + \bar{A}^{(1)}, \\ c_0 &= 1 + c_0^{(1)}, \\ c_k &= 0 + c_k^{(1)} \text{ for all } k \neq 0 . \end{aligned}$$

Retaining only terms linear in the fluctuation variables produces

$$\frac{\mathrm{d}c_1}{\mathrm{d}\bar{z}} = -\frac{\mathrm{i}}{2\bar{\rho}} c_1 - \bar{\rho}\bar{A} , \tag{17}$$

$$\frac{\mathrm{d}c_{-1}}{\mathrm{d}\bar{z}} = -\frac{\mathrm{i}}{2\bar{\rho}} c_{-1} + \bar{\rho}\bar{A}^* , \tag{18}$$

$$\frac{\mathrm{d}\bar{A}}{\mathrm{d}\bar{z}} = c_{-1}^* + c_1 + \mathrm{i}\delta\bar{A} , \tag{19}$$

where the $^{(1)}$ superscript has been dropped throughout as all the dependent variables represent fluctuating quantities. Looking for solutions to Eqs. (17)–(19) of the form $(c_1, c_{-1}, \bar{A}) \propto \exp(\mathrm{i}\Lambda\bar{z})$ results in the dispersion relation

$$(\Lambda - \delta)\left(\Lambda^2 - \frac{1}{4\bar{\rho}^2}\right) + 1 = 0 . \tag{20}$$

The solutions Λ of Eq. (20) with $\Im(\Lambda) < 0$ correspond to instability, resulting in exponential growth of the radiation field amplitude $\bar{A}$ and of the momentum state amplitudes $c_{\pm 1}$, and consequently a spatially periodic electron density modulation with wavelength λ, i.e., bunching.

Figure 3 shows a graph of the instability growth rate $|\Im(\Lambda)|$ versus the detuning δ for different values of $\bar{\rho}$. For $\bar{\rho} \to \infty$, it can be seen that Eq. (20) reduces to

$$\Lambda^2 (\Lambda - \delta) + 1 = 0 , \tag{21}$$

which is the dispersion relation which would be obtained from a linear stability analysis of the classical FEL model, Eqs. (1)–(3) [11]. Similarly, the curve for $\bar{\rho} = 10$ in Fig. 3 is almost identical to the classical FEL gain curve which would be obtained from Eq. (21), with the maximum growth rate at resonance ($\delta = 0$). As $\bar{\rho}$ is decreased, the region of gain decreases in size and the value of detuning at which maximum gain occurs shifts to increasing values of δ such that $\delta_{\text{max gain}} \approx 1/2\bar{\rho}$, i.e., $(\gamma_0 - \gamma_\text{r}) \approx \hbar k/2mc$.

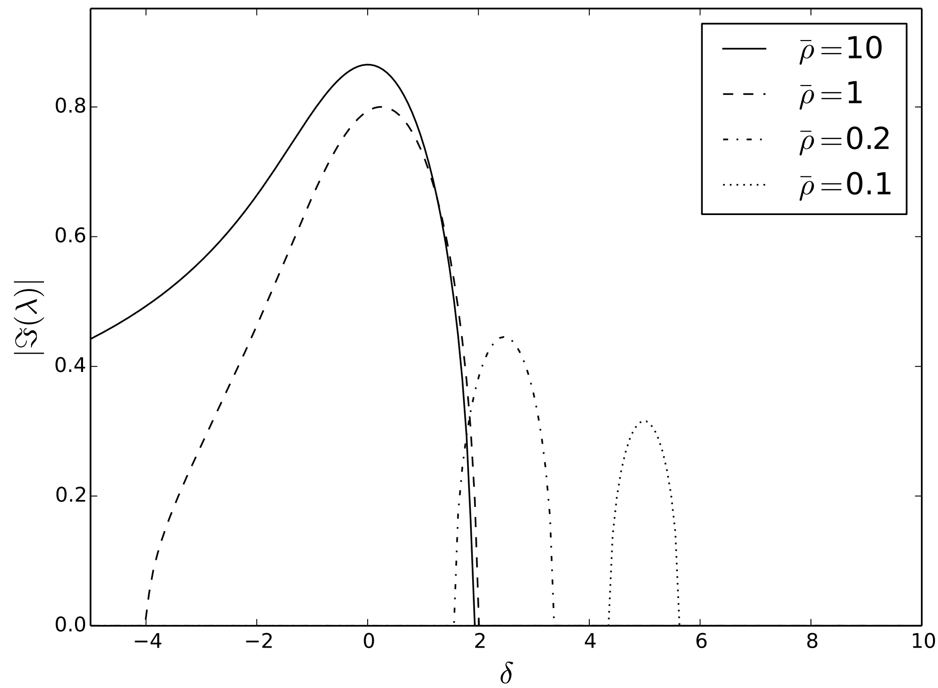


Fig. 3: Graph of growth rate against detuning, δ, for different values of $\bar{\rho}$ as calculated from Eq. (20)

4 One-dimensional quantum free-electron laser simulations

Using numerical simulations of the quantum FEL equations in either their position representation/Schrödinger form (Eqs. (12) and (13)) or their momentum representation (Eqs. (15) and (16)) allows investigation of the non-linear regime of the quantum FEL interaction. Here, results from simulations of the momentum representation equations are presented. In all cases shown, the initial condition corresponds to a monoenergetic/spatially uniform electron distribution such that $c_0 = 1$ and $c_k = 0 \; \forall \; k \neq 0$, and that the radiation field amplitude is extremely small ($\bar{A} = 10^{-4}$).

Figure 4 shows the evolution of the radiation field intensity as calculated from Eqs. (15) and (16) for a case where $\bar{\rho} = 10$ and $\delta = 0$. A comparison with Fig. 1 calculated from the classical equations, Eqs. (1)–(3), shows that the evolution of the radiation field is almost identical when calculated using either the quantum, wave-function-based FEL model or the classical, particle FEL model when $\bar{\rho} \gg 1$, which is consistent with this being the classical limit as discussed in Section 2. Figure 5 shows snapshots of the corresponding electron energy/momentum distribution for the case where $\bar{\rho} = 10$ and $\delta = 0$. It can be seen that during the interaction, a large number of momentum states ($\sim \bar{\rho}$) become populated, again consistent with the idea that this represents a classical regime of interaction.

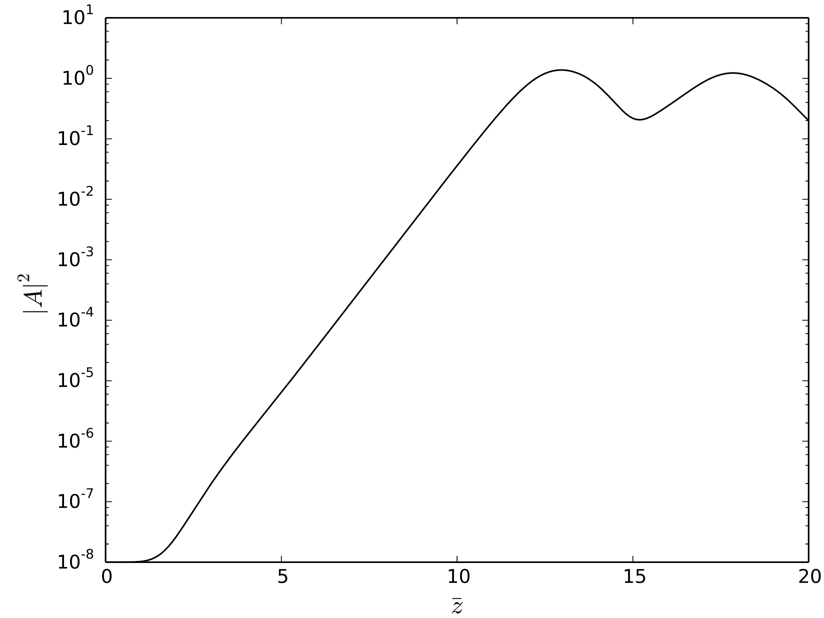


Fig. 4: Classical regime: evolution of radiation intensity, $|\bar{A}|^2$, when $\bar{\rho} = 10$, $\delta = 0$ as calculated from Eqs. (15) and (16).

In contrast, Fig. 6 shows the evolution of the radiation field intensity as calculated from Eqs. (15) and (16) for a case where $\bar{\rho} = 0.1$ and $\delta = 1/2\bar{\rho} = 5$, which should correspond to a quantum regime of interaction as discussed in Section 2. It can be seen that the field intensity evolves in a very different manner from that of the classical behaviour shown in Fig. 4 as a sequence of hyperbolic secant pulses. Figure 7 shows snapshots of the corresponding electron energy/momentum states. It can be seen that during the interaction, now only a maximum of two momentum states are populated at any stage of the interaction, the population cycling periodically between momentum states $n = 0$ and $n = -1$, with one radiation pulse being emitted as the population cycles from state $0 \to -1 \to 0$. It can be shown that this evolution is well described by a model consisting of only two momentum states.

4.1 Including slippage

So far, it has been assumed that the relative slippage between the radiation field and the electrons due to their different velocities is negligible, so that the electron beam can be described in terms of a single ponderomotive potential with periodic boundary conditions such that every ponderomotive potential in the electron beam evolves identically. Here, the quantum FEL model is extended to include the effects of slippage, which allows the description of radiation pulse propagation and evolution during the quantum FEL interaction.

The inclusion of slippage is essential to model the process of self-amplified spontaneous emission (SASE) in an FEL, as SASE involves the FEL interaction being initiated by electron beam shot noise rather than a coherent seed radiation field. Consequently, as electron shot noise is stochastic in nature,

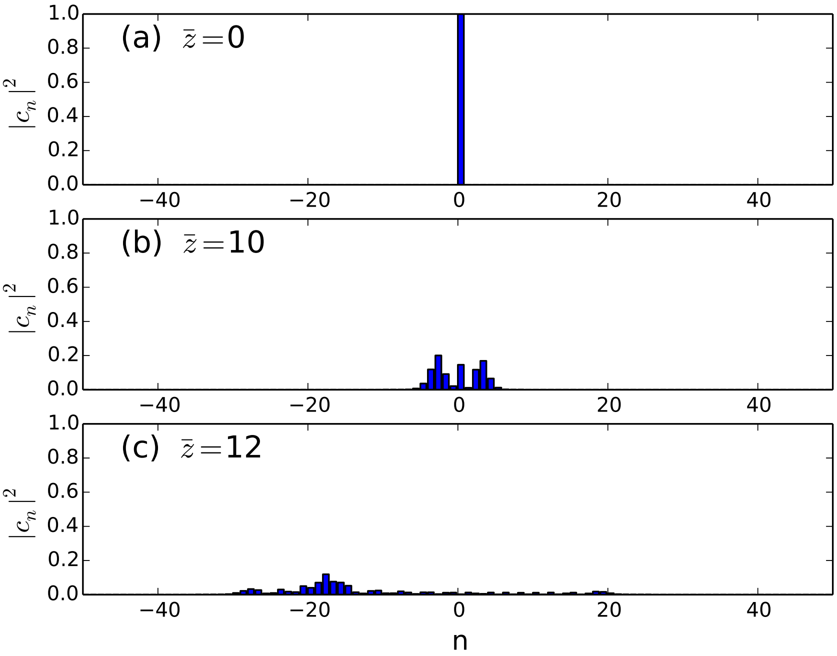


Fig. 5: Classical regime: snapshots of the electron momentum distribution, $|c_n|^2$, when $\bar{\rho} = 10$, $\delta = 0$ as calculated from Eqs. (15) and (16).

the (weak) initial bunching of electrons due to shot noise will vary randomly between different parts of the electron beam. It is important, therefore, to describe the effect of one part of the electron beam on the rest of the beam as mediated by the radiation field as it propagates forward through the electron beam.

To incorporate slippage into the models described so far, it is necessary to introduce an additional length scale which represents position in the electron bunch, i.e.,

$$z_1 = \frac{z - v_z t}{L_c} , \tag{22}$$

where $L_c = \lambda/4\pi\rho$ is the 'cooperation length' [17] and v_z is the mean axial velocity of the electron beam. .

Using this new independent variable, we can now define the radiation field amplitude $\bar{A}$ and the momentum state amplitudes c_n at each position along the electron bunch, i.e.,

$$\bar{A}(\bar{z}) \to \bar{A}(\bar{z}, z_1) ,$$
$$c_n(\bar{z}) \to c_n(\bar{z}, z_1) ,$$

so that the quantum FEL equations in the momentum state representation, Eqs. (15) and (16), become the set of coupled *partial* differential equations [18]

$$\frac{\partial c_n(\bar{z}, z_1)}{\partial \bar{z}} = -\mathrm{i}\frac{n^2}{2\bar{\rho}} c_n - \bar{\rho}\left(\bar{A} c_{n-1} - \bar{A}^* c_{n+1}\right) , \tag{23}$$

$$\left(\frac{\partial}{\partial \bar{z}} + \frac{\partial}{\partial z_1}\right) \bar{A}(\bar{z}, z_1) = \sum_{n=-\infty}^{\infty} c_n c_{n-1}^* + \mathrm{i}\delta \bar{A} . \tag{24}$$

As this model now describes the spatio-temporal evolution of the radiation field, it is possible to investigate the spectral properties of SASE radiation in a quantum FEL. To simulate SASE, the initial

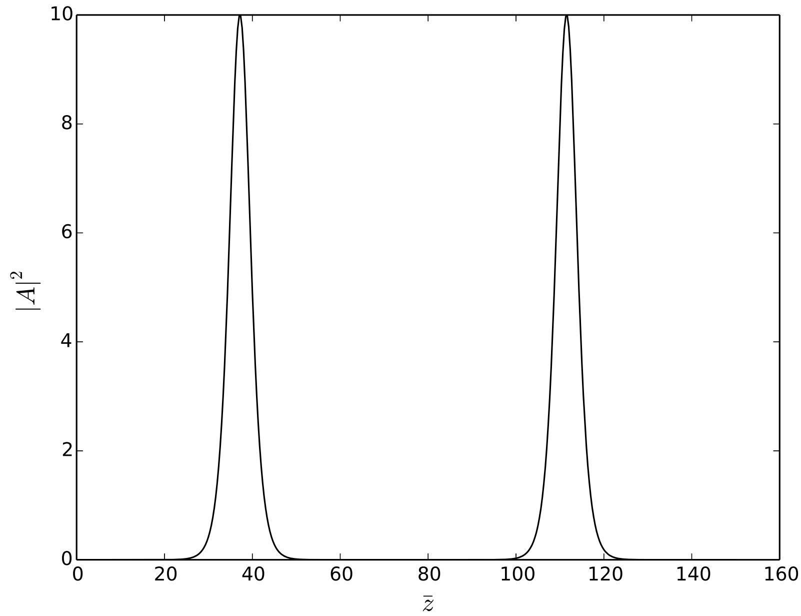


Fig. 6: Quantum regime: evolution of radiation intensity, $|\bar{A}|^2$, when $\bar{\rho} = 0.1$, $\delta = 5$ as calculated from Eqs. (15) and (16).

momentum amplitudes c_n are distributed with random phases to simulate a stochastically fluctuating bunching parameter along the electron bunch.

Snapshots of the spatio-temporal evolution of classical SASE radiation and its corresponding frequency spectrum are shown in Figs. 8 and 9, respectively. It can be seen that the evolution of the interaction results in a radiation field consisting of a sequence of random spikes of radiation, similar to what is produced in the classical model [11, 13]. Similarly, the corresponding frequency spectrum of classical SASE, shown in Fig. 9, is broad and chaotic. Note that the scaled frequency variable is defined as $\bar{\omega} = (\omega' - \omega_{\mathrm{r}})/\rho\omega_{\mathrm{r}}$, where $\omega_{\mathrm{r}} \approx 2c\gamma_{\mathrm{r}}^2 k_{\mathrm{w}}$ is the resonant (angular) frequency of the radiation.

Corresponding graphs showing the spatio-temporal evolution of SASE radiation and its frequency spectrum in the quantum regime ($\bar{\rho} < 1$) are shown in Figs. 10 and 11, respectively. It can be seen that the evolution of the interaction in the quantum regime results in a radiation field which is amplified at a lower rate, but consists of a smoother, less spiky profile, with a frequency spectrum which evolves as series of discrete, narrow lines. Each discrete line is produced by transitions between successive momentum states, i.e., $n = 0 \to -1 \to -2, \ldots$. It can be shown [19] that the frequency separation between these lines is $\Delta\omega = \hbar k^2/\gamma m$, the recoil frequency associated with the emission of a photon with momentum $\hbar k$ by a relativistic electron. The presence of these different discrete frequencies results in a beating between these frequencies, which can be seen in the spatio-temporal evolution of the field for large values of $\bar{z}$ (see, e.g., Fig. 10(b) and (c)). The existence of this beat note illustrates the phase coherence between the frequencies being generated.

The spectral features of the quantum regime of SASE, for example higher temporal coherence than that produced by classical SASE, are attractive for potential applications of the radiation generated.

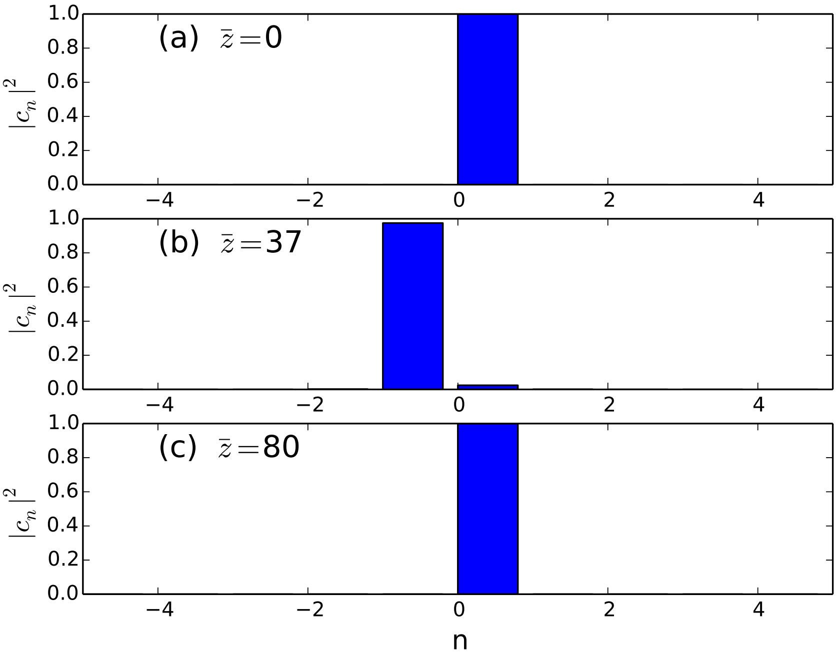


Fig. 7: Quantum regime: snapshots of the electron momentum distribution $|c_n|^2$ when $\bar{\rho} = 0.1$, $\delta = 5$ as calculated from Eqs. (15) and (16).

5 Realising the quantum free-electron laser regime

It was shown in Section 2 that the quantum FEL regime requires

$$\bar{\rho} = \frac{mc\gamma}{\hbar k}\rho < 1 ,$$

which can be rewritten as

$$\frac{\gamma\lambda}{\lambda_{\mathrm{c}}}\rho < 1 , \tag{25}$$

where $\lambda_{\mathrm{c}} = h/mc \approx 2.4 \times 10^{-12}$ m is the Compton wavelength. All current short-wavelength FELs operate in the classical regime; for example, the Linac Coherent Light Source (LCLS) has approximate parameters $\gamma \approx 3 \times 10^4$, $\rho \approx 5 \times 10^{-4}$, and $\lambda \approx 1$ Å $\approx 40\lambda_{\mathrm{c}}$, which from Eq. (25) gives $\bar{\rho} \approx 600$.

Equation (25) shows that to attain the quantum regime where $\bar{\rho} < 1$ without drastically reducing the FEL gain ($\propto \rho$), then it is best to reduce λ as far as possible (i.e., high photon momentum), but keep γ as small as possible. This would seem to be unfeasible for an FEL using a magnetostatic wiggler, such as the LCLS, but may be possible if an electromagnetic wiggler is used [19], where the much shorter effective wiggler period ($\lambda_{\mathrm{w}} < 1$ μm compared with $\sim$ cm for magnetostatic wigglers) allows generation of short radiation wavelengths with much lower-energy electrons than with a magnetostatic wiggler. A challenging feature of laser wigglers is that the effective interaction length is much shorter than for a magnetostatic wiggler, but recent studies [19, 20] suggest that current high-power technology is capable of producing sufficiently high intensities and pulse durations to potentially satisfy the conditions required to achieve the high-gain quantum FEL regime.

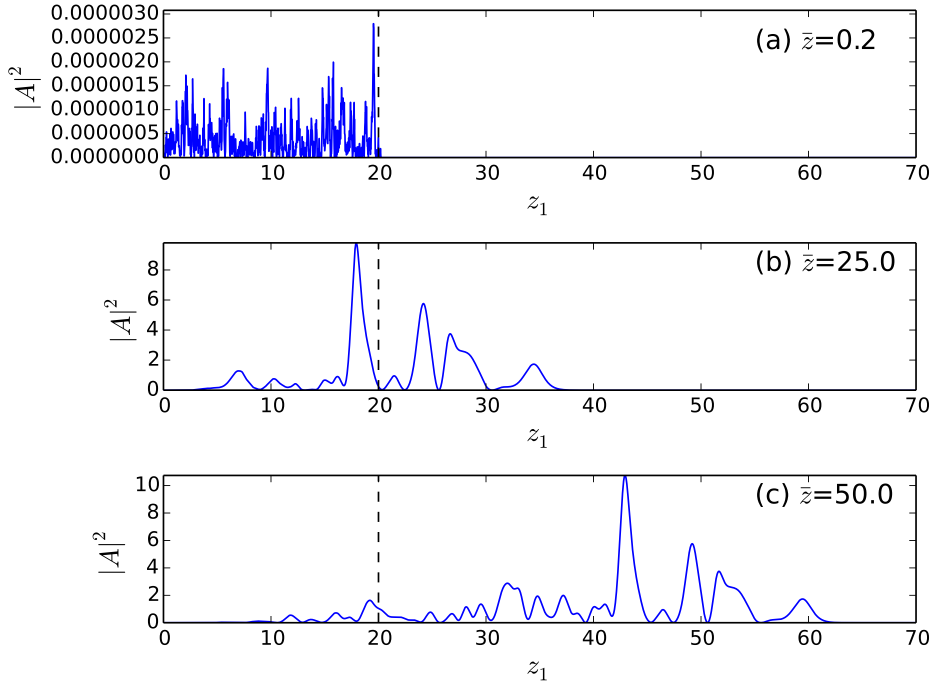


Fig. 8: Snapshots of the spatio-temporal evolution of the classical regime of SASE ($\bar{\rho} = 5$) when the electron bunch length $L_\mathrm{e} = 20L_\mathrm{c}$.

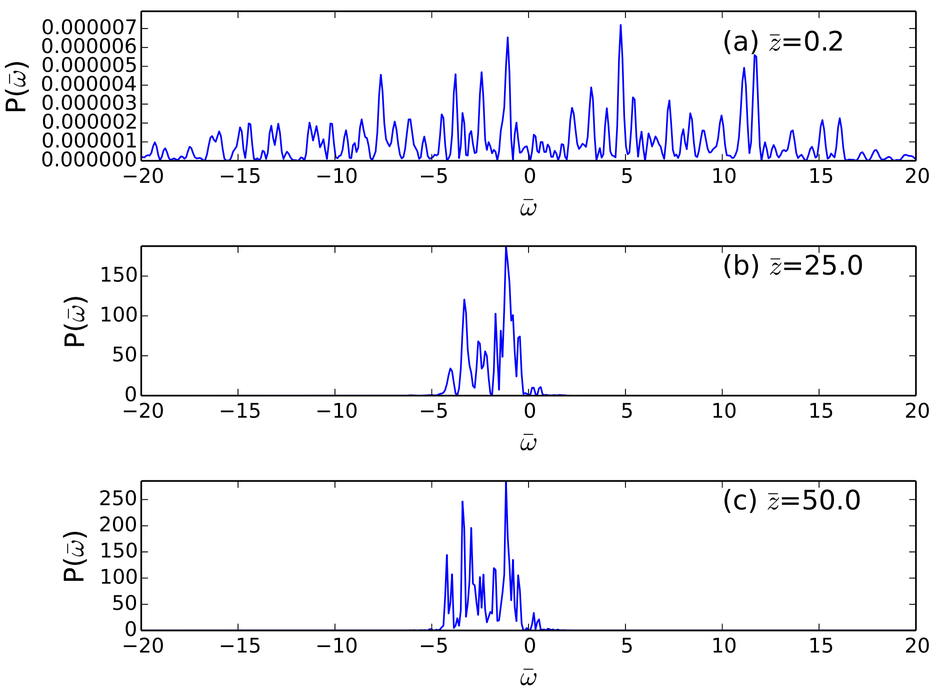


Fig. 9: Snapshots of the evolution of the frequency spectrum for classical SASE when the electron bunch length $L_\mathrm{e} = 20L_\mathrm{c}$.

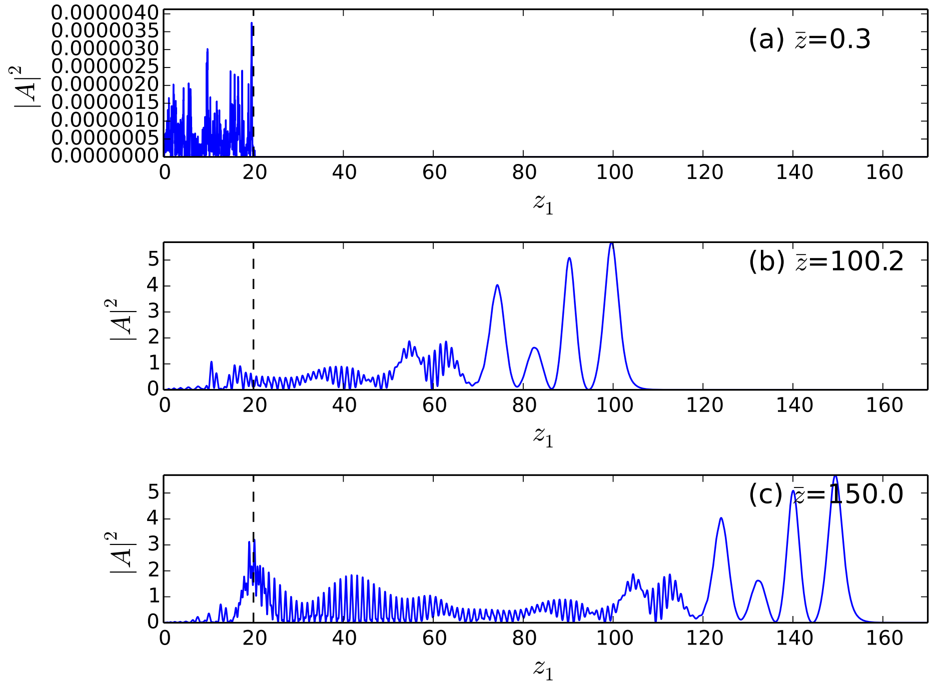


Fig. 10: Snapshots of the spatio-temporal evolution of the quantum regime of SASE ($\bar{\rho} = 0.2$) when the electron bunch length $L_e = 20L_c$.

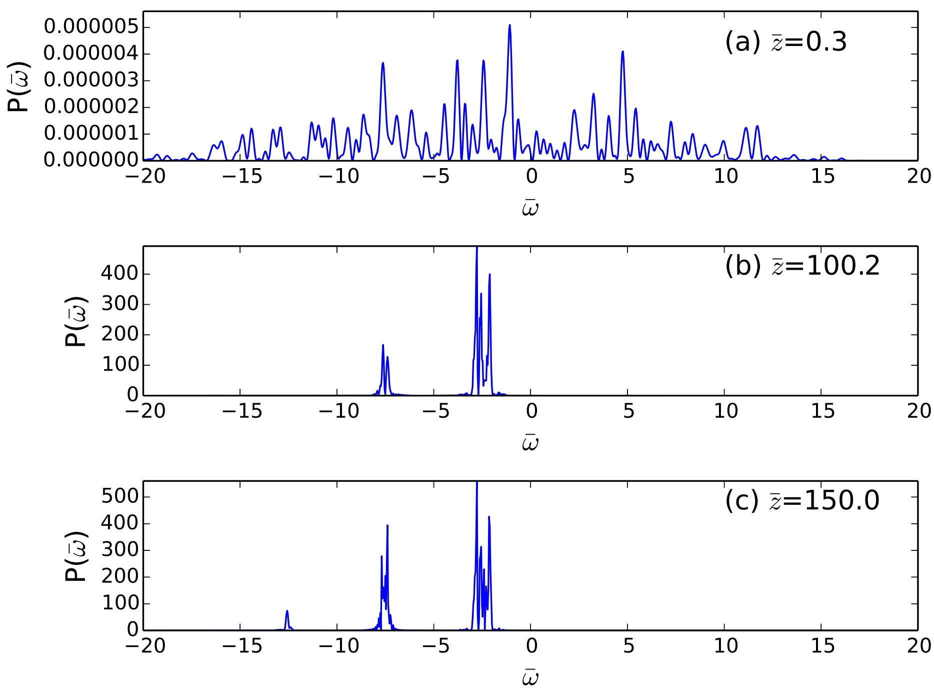


Fig. 11: Snapshots of the evolution of the frequency spectrum for quantum SASE when the electron bunch length $L_e = 20L_c$.

6 Conclusions

In this article, the possibility of a quantum regime of FEL operation, the conditions under which quantum effects due to large photon recoil may be significant in the FEL interaction, and some of the features of the quantum FEL interaction have been discussed using a 1D model. Although the quantum regime of high-gain FEL operation has not yet been achieved, it appears that with current technology, realisation of the quantum FEL regime may be a possibility. Given that some spectral properties of the quantum FEL regime appear attractive relative to those of classical SASE FELs, for example improved temporal coherence, then the quantum FEL seems to have potential as a compact source of coherent, short-wavelength (X-ray/γ-ray) radiation.

Some recent and current work on quantum FELs which has not been covered in detail here is summarised below.

6.1 2D and 3D models

Although this article has concentrated on 1D models to illustrate concepts, both 2D and 3D quantum FEL simulations and models have been investigated [21,22]. Such studies are important to determine the conditions under which transverse inhomogeneities in and diffraction of the intense laser pulses used as electromagnetic wigglers will affect the quantum FEL gain process.

6.2 Spontaneous emission

It is well known that the stochastic kicks produced by spontaneously emitted photons can produce momentum diffusion, which eventually quenches the classical FEL gain process at short wavelengths and provides an effective lower limit to the wavelength which can be produced by classical FEL operation of $\lambda \approx 1$ Å. There has recently been some debate about the significance of the role which spontaneous emission will play in attempts to achieve the quantum regime of FEL operation [23,24].

Finally, although the terminology of 'quantum' FEL has been used throughout this article, based on the conventional terminology used in the quantum optics literature, the models described here are more correctly described as 'semiclassical', as they use a quantum description of the medium (the electron beam) but a classical description of the electromagnetic field. Fully quantum FEL models involving quantized radiation fields were developed long ago for the low-gain regime and used to investigate, for example, the photon statistics of the low-gain FEL interaction [3], and recent work has extended these models into the high-gain regime [16,25]. Such models offer the exciting prospect of opening up a whole range of new coherent, quantum, X-ray, and γ-ray phenomena, as has been done in the optical region of the spectrum.

Acknowledgements

This article is dedicated to the memory of Rodolfo Bonifacio, who introduced me to the concept of the quantum free-electron laser. Many of the results described here arose from collaboration with Rodolfo and with Nicola Piovella.

References

[1] J. Madey, *J. Appl. Phys.* **42** (1971) 1906. https://doi.org/10.1063/1.1660466

[2] W. Becker and S. Zubairy, *Phys. Rev. A* **25** (1982) 2200. https://doi.org/10.1103/PhysRevA.25.2200

[3] W. Becker and J.K. McIver, *Phys. Rev. A* **27** (1983) 1030. https://doi.org/10.1103/PhysRevA.27.1030

[4] J. Gea-Banacloche, *Phys. Rev. A* **31** (1985) 1607. https://doi.org/10.1103/PhysRevA.31.1607

[5] A. Friedman, A. Gover, G. Kurizki, S. Ruschin, and A. Yariv, *Rev. Mod. Phys.* **60** (1988) 471. https://doi.org/10.1103/RevModPhys.60.471

[6] P. Kling, E. Giese, R. Endrich, P. Preiss, R. Sauerbrey, and W.P. Schleich, *New J. Phys.* **17** (2015) 123019. https://doi.org/10.1088/1367-2630/17/12/123019

[7] R. Bonifacio, N. Piovella, and G.R.M. Robb, *Nucl. Instrum. Methods Phys. Res. A* **543** (2005) 645. https://doi.org/10.1016/j.nima.2005.01.324

[8] C.B. Schroeder, C. Pellegrini, and P. Chen, *Phys. Rev. E* **64** (2001) 056502. https://doi.org/10.1103/PhysRevE.64.056502

[9] G. Preparata, *Phys. Rev. A* **38** (1988) 233. https://doi.org/10.1103/PhysRevA.38.233

[10] B.W.J. McNeil and N.R. Thompson, *Nat. Photonics* **4** (2010) 814. https://doi.org/10.1038/nphoton.2010.239

[11] R. Bonifacio *et al.*, *Riv. Nuovo Cim.* **13** (1990) 1. https://doi.org/10.1007/BF02770850

[12] J.B. Murphy and C. Pellegrini, in *Laser Handbook*, Vol. 6, Eds. W.B. Colson *et al.* (North-Holland, Amsterdam, 1990), p. 9.

[13] E. Saldin, E.V. Schneidmiller, and M.V. Yurkov, *The Physics of Free Electron Lasers* (Springer, 2000).

[14] A. Tonomura *et al.*, *Am. J. Phys.* **57** (1989) 117. https://doi.org/10.1119/1.16104

[15] M. Brown, J.R. Henderson, L.T. Campbell, and B.W.J. McNeil, in preparation.

[16] R. Bonifacio, N. Piovella, G.R.M. Robb, and A. Schiavi, *Phys. Rev. Accel. Beams* **9** (2006) 090701. https://doi.org/10.1103/PhysRevSTAB.9.090701

[17] R. Bonifacio, P. Pierini, and B.W.J. McNeil, *Phys. Rev. A* **40** (1989) 4467. https://doi.org/10.1103/PhysRevA.40.4467

[18] R. Bonifacio, N. Piovella, G.R.M. Robb, and M.M. Cola, *Opt. Commun.* **252** (2005) 381. https://doi.org/10.1016/j.optcom.2005.04.037

[19] R. Bonifacio, N. Piovella, M.M. Cola, and L. Volpe, *Nucl. Instrum. Methods Phys. Res. A* **577** (2007) 745. https://doi.org/10.1016/j.nima.2007.03.024

[20] R. Bonifacio, H. Fares, M. Ferrario, B.W.J. McNeil, and G.R.M. Robb, *Opt. Commun.* **382** (2017) 58. https://doi.org/10.1016/j.optcom.2016.07.007

[21] A. Schiavi, N. Piovella, G.R.M. Robb, and R. Bonifacio, *Int. J. Mod. Phys. A* **22** (2007) 4245. https://doi.org/10.1142/S0217751X07037792

[22] N. Piovella, M.M. Cola, L. Volpe, A. Schiavi, and R. Bonifacio, *Phys. Rev. Lett.* **100** (2008) 044801. https://doi.org/10.1103/PhysRevLett.100.044801

[23] G.R.M. Robb and R. Bonifacio, *Phys. Plasmas* **19** (2012) 073101. https://doi.org/10.1063/1.4729337

[24] G.R.M. Robb and R. Bonifacio, *Phys. Plasmas* **20** (2013) 124702. https://doi.org/10.1063/1.4838175

[25] R. Bonifacio and H. Fares, *Europhys. Lett.* **115** (2016) 34004. https://doi.org/10.1209/0295-5075/115/34004

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

High-Gain Regime: 3D[1]

K.-J. Kim[1], L. Ryan[2] and Z. Huang[3]
[1]The University of Chicago, Illinois, USA
[2]Argonne National Lab, Illinois, USA
[3]Stanford University, California, USA

Abstract
Although the FEL interaction is predominantly longitudinal in nature, transverse physics cannot be neglected if one wants to have a complete picture of the FEL. Specifically, we must understand the roles of radiation diffraction and how the electron's betatron motion in the undulator affects performance. We discuss these effects emphasizing the underlying physical picture. A high-gain FEL has a set of transverse modes, of which the fundamental mode has the largest growth rate and thus become dominant as the radiation-electron beam system travels along the undulator. To maximize the growth rate, the electron beam phase space distribution should be matched to the guided optical beam, leading to criteria on electron beam parameters. The FEL gain length is presented near the end of this chapter.

Keywords
High-Gain Regime; FELs.

1 Diffraction and guiding

A remarkable feature of a self-amplified spontaneous emission [SASE] FEL is its transverse coherence. The spontaneous undulator radiation has a transverse phase space area that is determined by the electron beam emittance $(2\pi\varepsilon_x)^2$. This area is typically much larger than the diffraction-limited phase space area $(\lambda/2)^2$, especially at X-ray wavelengths, so that undulator radiation is composed of many transverse modes. Thus, in a SASE FEL, the initial transverse phase space of the spontaneous emission also consists of an incoherent sum of many spatial modes. However, since the FEL interaction is localized within the electron beam near the peak electron density, there is one 'dominant' mode whose transverse size σ_r is dictated by the beam area, and whose natural divergence satisfies $\sigma_r\sigma_{r'} = \lambda/4\pi$. Higher-order spatial modes either diffract more, which results in greater effective losses, or are of larger spatial extent and couple less efficiently to the particles. Thus, the fundamental mode has the highest effective gain, so that it eventually becomes the preferred spatial distribution for the SASE radiation. This surviving fundamental mode appears to be guided after a sufficient undulator distance, a phenomenon commonly referred to as 'optical guiding' or 'gain guiding' [1, 2].

We illustrate the general idea of gain guiding schematically in Fig. 1. Since gain is only effective within the central area, one 'matched' transverse mode shape is selected over all others, and this mode then appears to be guided over many vacuum Rayleigh lengths due to the gain. The transverse mode selection is also clearly evident in Fig. 2, which was obtained from a three-dimensional (3D) GENESIS simulation of SASE. Initially, the radiation power is randomly distributed in the transverse plane, but after a sufficient propagation length only one localized coherent mode survives. For one Gaussian-like

[1]from Kwang-Je Kim, Zhirong Huang, Ryan Lindberg, Synchrotron Radiation and Free-Electron Lasers (2017), ©Cambridge University Press 2017, reproduced with permission. This material cannot be reproduced, shared, altered, or exploited commercially in any way without the permission of Cambridge University Press, as it is copyrighted material and therefore not subject to the allowances permitted by a CC licence.

https://doi.org/10.23730/CYRSP-2018-001.483

transverse mode to completely dominate in this way, there must be enough propagation distance for the competing modes to communicate transversely via diffraction.

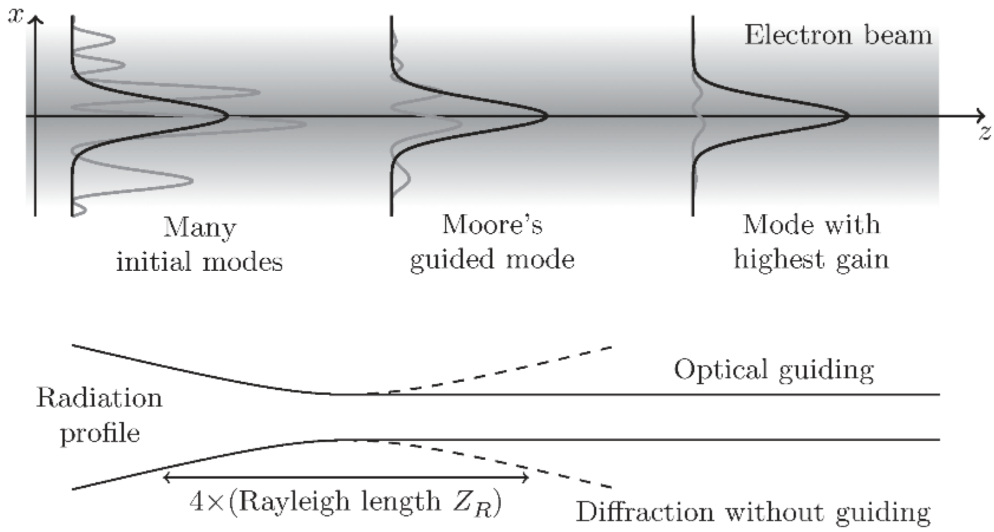


Fig. 1: Illustration of Moore's guided mode. In the top panel the preferentially guided mode is plotted in black, while the higher-order modes are in grey. The intensity at each z location is scaled to keep the height of the guided (black) mode invariant, so that what appears to be a decrease in the power in the higher-order (grey) modes is actually the larger gain of the Gaussian guided profile outstripping the smaller gain associated with all other modes. The bottom panel compares the natural diffraction of the radiation with that of the guided mode generated by FEL gain.

In the one-dimensional (1D) analysis, we introduced the important FEL scaling or Pierce parameter ρ, defined through the relation $n_e\kappa_1\chi_1 = 4k_u^2\rho^3$ which is equivalent to

$$\rho = \left(\frac{e^2K^2[\mathrm{JJ}]^2 n_e}{32\epsilon_0\gamma_r^3 mc^2k_u^2}\right)^{1/3} = \left[\frac{1}{8\pi}\frac{I}{I_{\mathrm{A}}}\left(\frac{K[\mathrm{JJ}]}{1+K^2/2}\right)^2\frac{\gamma_r\lambda_1^2}{2\pi\sigma_x^2}\right]^{1/3}, \tag{1}$$

where $I_{\mathrm{A}} = ec/r_e \approx 17045$ A is the Alfvén current. Many important characteristics of the FEL scale with ρ: the gain length and saturation length scale inversely with ρ, while the bandwidth is proportional to ρ. As shown in the paper titled "High-Gain Regime: 1D" in these proceedings, for vanishing e-beam energy spread the ideal gain length is given by

$$L_{G0} = \frac{\lambda_u}{4\sqrt{3}\pi\rho}\,. \tag{2}$$

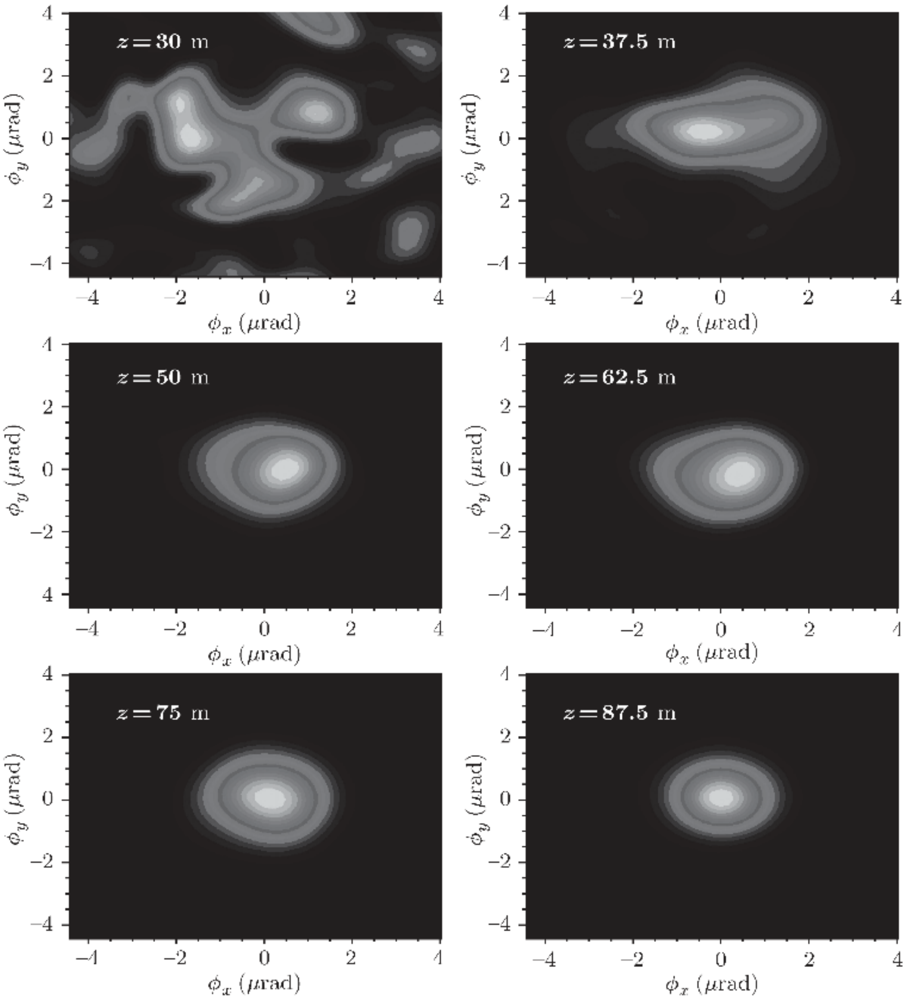


Fig. 2: Evolution of the LCLS radiation angular distribution at different z location. Courtesy of S. Reiche

When 3D effects are included, a different dimensionless combination of parameters may govern the gain characteristics of the FEL. To see this, consider the extreme case where the effect of diffraction is 'large', meaning that the radiation mode size is significantly larger than the electron beam size. To better describe the interaction between the electrons and the radiation in this 3D limit, the beam area $\mathcal{A}_{\mathrm{tr}} = 2\pi\sigma_x^2$ in Eq. (1) should be replaced by the diffraction-limited cross-section which is as follows:

$$2\pi\sigma_x^2 \to 2\pi\frac{\lambda_1}{4\pi}Z_R\,. \tag{3}$$

Here Z_R is the Rayleigh length of the radiation, which from our discussion on gain guiding ought to be of order a few gain lengths. Thus, by inserting $2\pi\sigma_x^2 \to \lambda_1 L_G$ into Eq. (1) and then the resulting expression for ρ into Eq. (2), one can solve the resulting algebraic equation for the gain length L_G to find

$$L_G^{-1} = \frac{4\pi}{\lambda_u}\frac{3^{3/4}}{2}\sqrt{\frac{I}{\gamma I_{\mathrm{A}}}\frac{K^2[\mathrm{JJ}]^2}{(1+K^2/2)}}\,. \tag{4}$$

This equation gives an approximate formula for the growth rate when the 3D effect of diffraction dominates, specifically, when the optical mode is larger than the electron beam cross-sectional area. Thus, it may be convenient to introduce the diffraction D-scaling for certain FEL applications as was done in Ref. [3]. Notice that L_G^{-1} scales as $I^{1/2}$ in the 3D diffractive limit, which is in contrast to the $I^{1/3}$ behaviour that characterizes the 1D limit when the electron beam size is larger than that of the optical mode. Additionally, the D-scaling shows that shrinking the electron beam cross-section much below that of the radiation mode does not further reduce the gain length. In fact, reducing the beam size beyond a certain point actually tends to increase the gain length, since decreasing the physical beam size necessarily increases the angular spread of an electron beam with non-zero emittance. It then follows from this discussion that the optimal electron beam size should roughly match the size of the radiation beam:

$$\sigma_x \sim \sigma_r = \sqrt{\varepsilon_r Z_R} \sim \sqrt{\varepsilon_r L_G}\,, \tag{5}$$

where $\varepsilon_r = \lambda_1/4\pi$ is the radiation emittance.

The above qualitative arguments are useful for understanding the effect of diffraction and for estimating the gain length of certain high-gain FEL projects operating in the infrared and visible wavelengths, where the optical mode size is larger than the e-beam size. Nevertheless, we will continue to scale quantities by the dimensionless parameter ρ for two reasons. First, ρ-scaling is more relevant for X-ray FELs because the typical optical mode size is smaller than the RMS beam size. Second, ρ does not require introducing the (formally undetermined) Rayleigh range, and instead relies on the electron beam cross-sectional area as shown in Eq. (1).

2 Beam emittance and focusing

An electron beam with finite emittance ε_x has a RMS angular spread $\sigma_{x'} = \varepsilon_x/\sigma_x$, so that its size will expand in free space. Hence, to keep a nearly constant e-beam size and maximize the FEL interaction in a long undulator channel requires proper electron focusing. The undulator magnetic field does provide a 'natural' focusing effect. The natural focusing strength, however, is typically too weak, so that external focusing by quadrupole magnets is often required. This focusing is used to decrease the beam size, thereby increasing the ρ parameter and decreasing the gain length. As mentioned in the previous section, decreasing the beam size below that of the optical mode may actually degrade the FEL performance, because the increasing angular spread introduces a spread in the resonant wavelength. This effect is similar to that of energy spread, and can be understood by considering the FEL resonance condition

$$\lambda_1(\psi) = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\psi^2\right), \tag{6}$$

where ψ is the angle the particle trajectory[1] makes with respect to the z axis. From Eq. (6), we see that the spread in particle angles given by $\psi = \sigma_{x'}$ causes a spread in the resonant wavelength

$$\frac{\Delta\lambda}{\lambda_1} = \sigma_{x'}^2\frac{\lambda_u}{\lambda_1} = \frac{\varepsilon_x}{\beta_x}\frac{\lambda_u}{\lambda_1}\,. \tag{7}$$

To not adversely affect the FEL gain, we demand that the induced wavelength variation due to the angular spread be less than the FEL bandwidth $\sim \rho$, namely that

$$\frac{\Delta\lambda}{\lambda_1} = \sigma_{x'}^2\frac{\lambda_u}{\lambda_1} \lesssim \rho \approx \frac{\lambda_u}{4\pi L_G}\,. \tag{8}$$

Due to optical guiding, the radiation Rayleigh range is of order the gain length, $Z_R \sim L_G$, so that Eq. (8) implies that the electron beam angular divergence should be no more than that of the radiation:

$$\sigma_{x'} = \sqrt{\frac{\varepsilon_x}{\beta_x}} \leq \sqrt{\frac{\varepsilon_r}{L_G}} \sim \sigma_{r'}\,. \tag{9}$$

The inequalities regarding the beam size (see Eq. (5)), and angular divergence (see Eq. (9)), together require

$$\varepsilon_{x,y} \lesssim \varepsilon_r = \frac{\lambda_1}{4\pi}\,, \tag{10}$$

while the optimal focusing beta function for a given emittance saturates the inequality seen in Eq. (9):

$$\beta_x \sim L_G \frac{\varepsilon_x}{\varepsilon_r}\,. \tag{11}$$

A smaller beam emittance allows for a tighter focused beam size and hence a smaller gain length.

References

[1] G.T. Moore, *Nucl. Instrum. Methods Phys. Res., Sect. A* **239** (1985) 19. https://doi.org/10.1016/0168-9002(85)90693-x

[2] E.T. Scharlemann, A.M. Sessler and J.S. Wurtele, *Phys. Rev. Lett.* **54** (1985) 1925. https://doi.org/10.1103/physrevlett.54.1925

[3] L.-H. Yu, S. Krinsky and R.L. Gluckstern, *Phys. Rev. Lett.* **64** (1990) 3011. https://doi.org/10.1103/physrevlett.64.3011

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Temporal Coherence of Radiation from a Collection of Electrons[1]

K.-J. Kim[1], L. Ryan[2] and Z. Huang[3]
[1]The University of Chicago, Illinois, USA
[2]Argonne National Lab, Illinois, USA
[3]Stanford University, California, USA

Abstract
We review [1] the temporal characteristics of the radiation produced by an electron beam, in time-domain as well as in the frequency domain. For synchrotron radiation, the radiation is chaotic, while it is coherent when the beam is micro-bunched as in a free-electron laser.

Keywords
Radiation; electron beam; time domain; frequency domain.

1. Time-domain picture

Temporal coherence of a radiation specifies the extent to which the radiation maintains a definite phase relationship at two different times. Temporal coherence is characterized by the coherence time, which can be experimentally determined by measuring the path length difference over which fringes can be observed in a Michelson interferometer. A simple representation of a coherent wave in time is given by

$$E_0(t) = e_0 \exp\left(-\frac{t^2}{4\sigma_\tau^2} - \mathrm{i}\omega_1 t\right). \tag{1}$$

Here σ_τ is the root mean square (RMS) temporal width of the intensity profile $|E_0(t)^2|$. The coherence time t_{coh} can be defined as

$$t_{\mathrm{coh}} \equiv \int \mathrm{d}\tau |C(\tau)|^2 , \tag{2}$$

where $C(\tau)$ is the normalized, first-order correlation function (or complex degree of temporal coherence) given by

$$C(\tau) \equiv \frac{\langle \int \mathrm{d}t\, E(t)E^*(t+\tau)\rangle}{\langle \int \mathrm{d}t\, |E(t)|^2\rangle} , \tag{3}$$

and the brackets denote ensemble averaging. In the simple Gaussian model of Eq. (1), the coherence time $t_{\mathrm{coh}} = 2\sqrt{\pi}\sigma_\tau$.

In the frequency domain, we have

$$E_\omega^0 = \int \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} E_0(t) = \frac{e_0\sqrt{\pi}}{\sigma_\omega} \exp\left[-\frac{(\omega-\omega_1)^2}{4\sigma_\omega^2}\right], \tag{4}$$

where $\sigma_\omega = (2\sigma_\tau)^{-1}$ is the RMS width of the frequency profile $|E_\omega|^2$. Let us introduce the temporal (longitudinal) phase space variables ct and $(\omega-\omega_1)/\omega_1 = \Delta\omega/\omega_1$.The Gaussian wave packet then satisfies

[1]from Kwang-Je Kim, Zhirong Huang, Ryan Lindberg, Synchrotron Radiation and Free-Electron Lasers (2017), ©Cambridge University Press 2017, reproduced with permission. This material cannot be reproduced, shared, altered, or exploited commercially in any way without the permission of Cambridge University Press, as it is copyrighted material and therefore not subject to the allowances permitted by a CC licence.

2519-8041 – © Cambridge University Press 2017, reproduced with permission.

https://doi.org/10.23730/CYRSP-2018-001.489

$$c\sigma_\tau \cdot \frac{\sigma_\omega}{\omega_1} = \frac{\lambda_1}{4\pi}\,. \tag{5}$$

Most radiation observed in nature, however, is temporally incoherent. Sunlight, fluorescent light bulbs, black-body radiation, and undulator radiation are all temporally incoherent, and are often referred to as chaotic light or as a partially coherent wave. As a mathematical model of such chaotic light, we consider a collection of coherent Gaussian pulses that are displaced randomly in time with respect to each other:

$$E(t) = \sum_{j=1}^{N_\mathrm{e}} E_0\big(t - t_j\big) = e_0 \sum_{j=1}^{N_\mathrm{e}} \exp\left[-\frac{\big(t - t_j\big)^2}{4\sigma_\tau^2} - \mathrm{i}\omega_1\big(t - t_j\big)\right]. \tag{6}$$

In Eq. (6), t_j is a random number, and the sum extends to N_e to suggest that these wave packets have been created by electrons. We illustrate this partially coherent wave (chaotic light) in Fig. 1, which we obtained by using $N_\mathrm{e} = 100$ wave packets with $\lambda_1 = 2\pi/\omega_1 = 1$ and $\sigma_\tau = 2$ ($\sigma_\omega = 0.25$), assuming that the t_j are randomly distributed with equal probability over the bunch length duration $T = 100$. Panel (a) shows 10 randomly chosen such wave packets; plotting many more than this results in a jumbled disarray. Figure 1(b) shows the $E(t)$ that results by summing over all 100 waves.

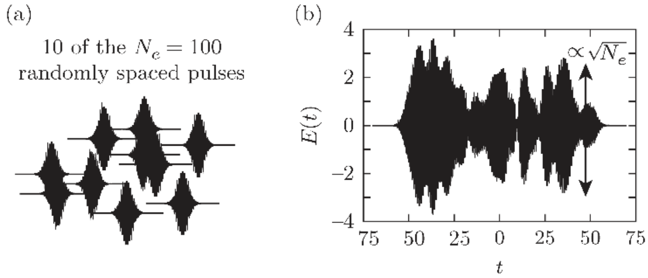


Fig. 1: (a) Representation of the randomly phased wave packets that chooses 10 out of the 100 total waves. The individual waves are shown transversely displaced for illustrative purposes only. (b) Total electric field, given by the incoherent sum of the 100 wave packets. The field consists of order $T/4\sigma_\tau \approx 10$ regular regions (i.e., $M_\mathrm{L} \approx 10$ longitudinal modes).

The remarkable feature of this plot is that the resultant wave is a relatively regular oscillation that is interrupted only a few times, much fewer than one might have naively guessed based on the fact that it is a random superposition of 100 wave packets. In fact, the duration of each regular region is independent of the number of wave packets, and is instead governed by the time over which the wave maintains a definite phase relationship, namely, the coherence time. Note that the coherence time of a random collection of Gaussian waves Eq. (6) equals that of the single mode Eq. (1). Thus, each regular region can be identified with a coherent mode whose temporal width is of order the coherence time t_coh. The number of regular regions equals the number of coherent longitudinal modes M_L, which is roughly the ratio of the bunch length to the coherence length. Approximately, we have

$$M_\mathrm{L} \approx \frac{T}{t_\mathrm{coh}} = \frac{T}{2\sqrt{\pi}\sigma_\tau} \approx \frac{T}{4\sigma_\tau}\,. \tag{7}$$

The average field intensity scales linearly with the number of sources, while the instantaneous intensity fluctuates as a function of time. Associated with this intensity variation will be a fluctuation in the observed number of photons $\mathcal{N}_\mathrm{ph}$ over a given time. Denoting the average photon number by $\langle\mathcal{N}_\mathrm{ph}\rangle$, the RMS squared fluctuation in the number of photons observed is

$$\sigma^2_{\mathcal{N}_{\mathrm{ph}}} = \frac{\langle \mathcal{N}_{\mathrm{ph}} \rangle^2}{M_{\mathrm{L}}}\,, \tag{8}$$

where M_{L} is the number of longitudinal modes in the observation time T.

The formula Eq. (8) for the photon number variation can be generalized in two respects. First, the mode counting must include the number of transverse modes M_{T} in both the x and y directions, so that the total number of modes

$$M = M_{\mathrm{L}} M_{\mathrm{T}}^2. \tag{9}$$

Second, there are inherent intensity fluctuations arising from quantum mechanical uncertainty in the form of photon shot noise. This number uncertainty is attributable to the discrete quantum nature of electromagnetic radiation, and, like any shot noise, it adds a contribution to $\sigma^2_{N_{\mathrm{ph}}}$ equal to the average number $\mathcal{N}_{\mathrm{ph}}$. Thus, the RMS squared photon number fluctuation is

$$\sigma^2_{\mathcal{N}_{\mathrm{ph}}} = \frac{\langle \mathcal{N}_{\mathrm{ph}} \rangle^2}{M} + \langle \mathcal{N}_{\mathrm{ph}} \rangle = \frac{\langle \mathcal{N}_{\mathrm{ph}} \rangle^2}{M}\left(1 + \frac{1}{\delta_{\mathrm{degen}}}\right). \tag{10}$$

The second term in parentheses is the inverse of the number of photons per mode, which is also known as the degeneracy parameter. In the classical devices that we consider there are many photons per mode, $\langle \mathcal{N}_{\mathrm{ph}} \rangle / M \equiv \delta_{\mathrm{degen}} \gg 1$, and the fluctuations due to quantum uncertainty are negligible. In this classical limit the length of the radiation pulse can be determined by measuring its intensity fluctuations, from which the source electron beam length may be deduced, see Ref. [2].

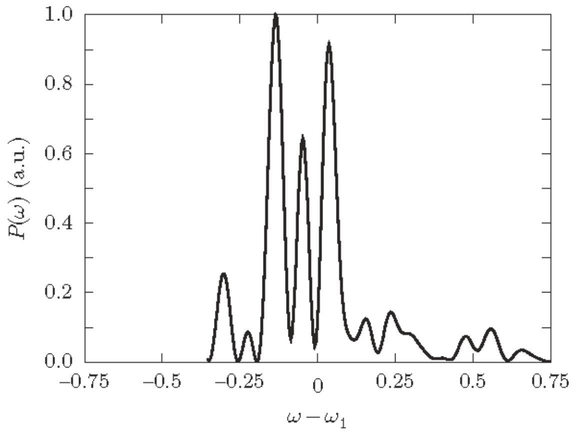


Fig. 2: Intensity spectrum of Eq. (11) using identical parameters as Fig. 1(b). The spectrum consists of $M \sim 10$ sharp frequency spikes of approximate width $2/T \approx 0.02$, which are distributed within a Gaussian envelope of RMS width $\sigma_\omega \sim 0.25$. The height and placement of the spectral peaks fluctuate by 100 per cent for different sets of random numbers.

2. Frequency-domain picture

It is interesting to note that the mode counting we performed in the time domain can also be done in the frequency domain. Figure 2 shows the intensity spectrum $P(\omega) \propto |E_\omega|^2$, where

$$E_\omega = \frac{e_0\sqrt{\pi}}{\sigma_\omega} \sum_{j=1}^{N_{\mathrm{e}}} \exp\left[-\frac{(\omega - \omega_1)^2}{4\sigma_\omega^2} + \mathrm{i}\omega t_j\right] \tag{11}$$

using the same wave parameters as in Fig. 1. The spectrum consists of sharp peaks of width $\Delta\omega \sim 2/T$ that are randomly distributed within the radiation bandwidth $\sigma_\omega = (2\sigma_\tau)^{-1}$. In other words, the frequency bandwidth $\Delta\omega$ of each mode is set by the duration of the entire radiation pulse T, while the frequency range over which the modes exist is given by the inverse coherence time. Thus, the total number of spectral peaks is the same as the number of the coherent modes in the time domain.

References

[1] K.-J. Kim, "Temporal and Transverse Coherence of Self-Amplified Spontaneous Emission", in AIP Conf. Proc. 413, "*Towards X-ray Free Electron Lasers*," R. Bonifacio and W. Barletta, eds., (AIP, 1997)

[2] P. Catravas *et al.*, *Phys. Rev. Lett.* **82** (1999) 5261. https://doi.org/10.1103/physrevlett.82.5261

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

FEL Oscillator Principles[1]

K.-J. Kim[1], L. Ryan[2] and Z. Huang[3]
[1]The University of Chicago, Illinois, USA
[2]Argonne National Lab, Illinois, USA
[3]Stanford University, California, USA

Abstract
In this lecture we discuss the principles of an FEL oscillator, in which a radiation pulse is trapped in an optical cavity but receives repeated amplification as the pulse meets an electron bunch as it come to the entrance of a low-gain FEL. We provide a qualitative picture of how the power and the longitudinal and transverse modes of the pulse develop.

Keywords
FEL oscillator; basic principles.

1. Introduction

The basic schematic of an FEL oscillator is illustrated in Fig. 1. Electron bunches from a (usually radiofrequency) accelerator pass through an undulator that is located inside a low-loss optical cavity. Starting from an empty cavity, in the first pass the electron beam emits spontaneous undulator radiation that is reflected back into the undulator by the cavity mirrors. In the second pass, the pulse of spontaneous emission meets and overlaps with a second electron bunch at the entrance of the undulator. The radiation and the e-beam interact in the undulator, after which the output field is composed of the spontaneous emission from both the first and second pass, along with an amplified signal due to FEL gain. This process repeats, so that the amplified radiation signal will eventually dominate the output if the gain is larger than the round-trip loss in the cavity.

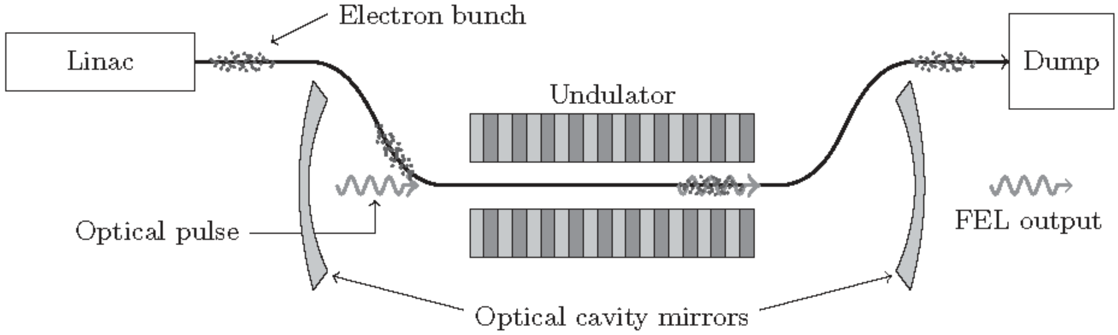


Fig. 1: Schematic of an FEL oscillator showing its basic operating principle

1 Power evolution and saturation

For a simple mathematical description of the power evolution in an oscillator, let P_n be the power of the optical pulse at the undulator exit after its nth pass, and P_s be the power of spontaneous emission. Then

[1]from Kwang-Je Kim, Zhirong Huang, Ryan Lindberg, Synchrotron Radiation and Free-Electron Lasers (2017), © Cambridge University Press 2017, reproduced with permission. This material cannot be reproduced, shared, altered, or exploited commercially in any way without the permission of Cambridge University Press, as it is copyrighted material and therefore not subject to the allowances permitted by a CC licence.

$$P_1 = P_s\,,$$
$$P_n = R(1+G)P_{n-1} + P_s \quad \text{for } n \geq 2\,, \tag{1}$$

where G is the FEL gain and R is the reflectivity of the optical transport line. The net single pass power amplification is $R(1 + G)$, and evidently the power increases if the single pass gain overcomes the losses such that

$$R(1 + G) > 1\,. \tag{2}$$

This is the 'lasing' condition for an FEL oscillator. The power after the nth pass is governed by Eq. (1), whose solution is

$$P_n = \frac{[R(1+G)]^n - 1}{R(1+G) - 1} P_s\,. \tag{3}$$

Assuming that $R(1 + G) > 1$, we see that the power increases exponentially with n after sufficiently many passes of amplification.

The exponential growth of the intracavity radiation power does not continue indefinitely. Rather, the optical power eventually becomes large enough to trap electrons in the ponderomotive potential and then rotate them to an absorptive phase where they extract energy from the field. This in turn reduces the gain from its small signal value, and the system reaches a steady state or 'saturates' when the gain decreases to the value G_{sat} given by

$$R(1 + G_{\text{sat}}) = 1\,. \tag{4}$$

Furthermore, at saturation the power generated during one pass ΔP equals the total losses, so that if the power inside the cavity is P_{sat} we have $\Delta P = (1 - R)P_{\text{sat}}$. It can be shown that $\Delta P \approx P_{\text{beam}}/2N_u$, which in turn implies that the intracavity optical power at saturation is

$$P_{\text{sat}} \approx \frac{1}{2N_u(1-R)} P_{\text{beam}}\,. \tag{5}$$

The optical elements in the cavity, and in particular the mirrors, must be able to withstand the power P_{sat} for the oscillator to operate stably.

At saturation the power decreases by an amount $(1 - R)\Delta P$ during any complete round-trip cycle; this energy loss can be due to many different mechanisms, including radiation absorption in the mirror material, diffraction at the edges of the optical elements, and transmission out of the cavity for useful purposes. If one had an ideal optical line with no losses, the cavity transmission would equal $(1 - R)$ so that the maximum power that can be coupled out of the oscillator is $(1 - R)P_{\text{sat}} \approx P_{\text{beam}}/2N_u$.

Useful output radiation from an FEL oscillator requires it to operate for some time at saturation. Hence, an oscillator can be driven by a pulsed accelerator only if the number of bunches within each macro-pulse is more than that required to reach saturation. With a CW accelerator, on the other hand, the oscillator can be maintained at a steady state indefinitely. This is a desirable mode of operation, since the FEL then provides a stable source with a higher average photon flux.

2 Qualitative description of longitudinal mode development

There is much more physics at work in addition to the power evolution just described. One subtle but important phenomenon is lethargy (Ref. [1])—the fact that the trailing part of the optical pulse (the tail) is more strongly amplified than the front (the head). This is because the initially unmodulated electron beam must propagate some distance through the undulator to develop the density modulation that provides FEL gain, during which time the electron beam and its gain slips behind the field envelope. As a consequence, the FEL gain is maximized when the cavity length is slightly shorter than that given by the exact synchronism condition (the synchronism condition is when the cavity length equals the distance between successive bunches).

The lethargy effect causes the round-trip time of the pulse envelope to be in general different from the round-trip time of the phase, since the latter is determined essentially by the cavity length. In other words, the phase fronts return to the undulator after a time approximately equal to the round-trip time in the cavity, while the peak of the pulse envelope arrives a time of order the slippage time $N_u\lambda_1/c$ after. To be more precise, any delay of the phase fronts is given by the imaginary part of the complex gain, which is small at peak gain.

Temporal coherence in an FEL oscillator is achieved by gain narrowing due to the FEL itself and also through spectral filtering provided by the cavity mirrors if their reflectivity is wavelength-dependent. The FEL-induced spectral gain narrowing occurs because the FEL gain is frequency dependent; alternatively, it can be understood as the slow increase in the coherence length from $N_u\lambda_1$ due to many passes through the undulator. Hence, when the mirror reflectivity is independent of wavelength, we expect that the FEL spectral bandwidth σ_ω decreases with pass number n as

$$\left(\frac{\sigma_\omega}{\omega_1}\right)_n \sim \frac{1}{N_u}\frac{1}{\sqrt{n}}\,. \tag{6}$$

For short electron bunches, gain narrowing stops when $(\sigma_\omega/\omega_1)_n$ becomes the transform limited bandwidth $\lambda_1/(4\pi\sigma_z)$ associated with the root mean square (RMS) length of the electron bunch σ_z. For longer electron bunches that have a current maximum in the centre, the non-uniform gain causes the optical pulse profile to also narrow in length/duration, with $(\Delta z)_n^{\mathrm{rms}} \sim \sigma_z/\sqrt{n}$. The spectral and temporal narrowing will stop when the pulse is determined by Fourier transform-limited, i.e., at the pass number $n \sim N_{\mathrm{FT}}$ determined by

$$\left(\frac{\sigma_\omega}{\omega_1}\right)_{N_{\mathrm{FT}}} (\Delta z)_{N_{\mathrm{FT}}}^{\mathrm{rms}} \sim \frac{\lambda_1}{4\pi}\,, \tag{7}$$

from which we determine that the steady state is reached after approximately $N_{\mathrm{FT}} \sim 4\pi\sigma_z/\lambda_1 N_u$ passes, and that the limiting bandwidth is, as seen in Ref. [2]

$$\frac{\sigma_\omega}{\omega_1} \sim \sqrt{\frac{\lambda_1}{4\pi N_u\sigma_z}}\,. \tag{8}$$

This limiting mode is known as the dominant supermode, see Ref. [3].

In what follows we will show how the longitudinal supermodes arise from the dynamic interplay between amplification, gain narrowing, FEL lethargy, and spectral filtering from the mirrors. Hence, we will further extend the physics above to include the possibility that the spectral narrowing comes about not only through slippage, but also because the mirrors have a limited bandpass.

3 Longitudinal supermodes of the FEL oscillator

In this section we use the simple low-gain model developed by Elleaume [4] to more fully investigate the supermode longitudinal dynamics. This model divides the evolution during a single round trip into its various components: gain that depends on the current and the propagation/slippage in the undulator, reflection by the mirrors, and propagation in the cavity. Assuming that all of these effects result in small perturbations to the radiation (as is true in the low-gain regime), then we can approximate each as acting individually and in succession on the electric field $E(t)$. We discuss these longitudinal effects in turn, and then combine them into a single equation describing the linear dynamics of a low-gain oscillator.

We assume that the FEL gain transforms the field via the amplification operator $E \to E + \mathcal{G}[E]$. To develop a simple model for, we recall that FEL gain depends linearly on the current and that the field interacts with the electron beam within one slippage length $N_u\lambda_1$. In terms of the light-cone coordinate $\tau \equiv z - ct$, this means that the amplification of $E(\tau)$ depends on the interaction between the current and field amplitude for points τ' satisfying $\tau \le \tau' \le \tau + N_u\lambda_1$ (see, for example, [5]). We will use a very simple description of this process in which we model the gain operation $\mathcal{G}[E]$ as increasing the field by an amount depending on the e-beam current and E-field amplitude at the point one-half the slippage distance $N_u\lambda_1/2$ ahead. Hence, we approximate the amplitude gain from an electron beam with RMS length σ_z as acting via

$$
\begin{aligned}
E(\tau) \to E(\tau) + \mathcal{G}[E] &\approx E(\tau) + \frac{G}{2}\mathrm{e}^{-(\tau+N_u\lambda_1/2)^2/2\sigma_z^2}E\left(\tau + \frac{1}{2}N_u\lambda_1\right) \\
&\approx \left[1 + \frac{G}{2}\left(1 - \frac{t^2}{2\sigma_z^2}\right)\right]E(\tau) \\
&\quad + \frac{G}{4}N_u\lambda_1\frac{\partial E}{\partial \tau} + \frac{G}{16}(N_u\lambda_1)^2\frac{\partial^2 E}{\partial \tau^2}\,,
\end{aligned}
\tag{9}
$$

where for simplicity we assume that $\sigma_z \gg N_u\lambda_1$ and that the amplitude gain $G/2$ is real[2].

After the FEL interaction, the mirror reduces the field amplitude by the multiplicative factor $\sqrt{R} \equiv \sqrt{1-\alpha} \approx 1 - \alpha/2$, where α is the (assumed real) power loss. In addition, we include the possibility that the reflectivity depends on frequency by modelling it as a Gaussian filter in ω with RMS power bandwidth σ_{refl}. Since we model the mirror filtering as acting on the slowly-varying field envelope, it is centred near $\omega = 0$ and results in the transformation

$$
\begin{aligned}
E(\tau) \to \int \mathrm{d}\omega\, \mathrm{e}^{-\mathrm{i}\omega\tau/c}\, R(\omega)E(\omega) &\approx (1 - \alpha/2)\int \mathrm{d}\omega\, \mathrm{e}^{-\mathrm{i}\omega\tau/c}\mathrm{e}^{-\omega^2/4\sigma_{\text{refl}}^2}E(\omega) \\
&\approx \left(1 - \frac{\alpha}{2}\right)\int \mathrm{d}\omega\, \mathrm{e}^{-\mathrm{i}\omega\tau/c}\left[1 - \frac{\omega^2}{4\sigma_{\text{refl}}^2}\right]E(\omega) \\
&= \left(1 - \frac{\alpha}{2}\right)E(\tau) + \frac{c^2}{4\sigma_{\text{refl}}^2}\frac{\partial^2}{\partial \tau^2}E(\tau).
\end{aligned}
\tag{10}
$$

Finally, we include the possibility that after one round trip through the cavity the arrival time of the radiation pulse and the next electron bunch may differ by an amount ℓ/c; this timing difference, called *detuning* in the FEL community, could be due to adjustments to the cavity length or timing jitter of the electrons; we model it by

$$
E(\tau) \to E(\tau + \ell) \approx E(\tau) + \ell\frac{\partial}{\partial \tau}\,E(\tau)\,.
\tag{11}
$$

A full pass through the oscillator is composed of the transformation Eqs. (9)–(11) due to the gain including slippage, the mirror, and the cavity length detuning. Every transformation is written as a sum of the initial field $E(\tau)$ and a perturbation. If each of these perturbing effects is small, then the field at pass $(n + 1)$ can be written a sum of the various perturbations acting on the field E_n as follows:

$$
\begin{aligned}
E_{n+1}(\tau) &\approx E_n(\tau) + \frac{G-\alpha}{2}E_n(\tau) - \frac{G\tau^2}{4\sigma_z^2}E_n(\tau) \\
&\quad + \left(\ell + \frac{GN_u\lambda_1}{4}\right)\frac{\partial E_n}{\partial \tau} + \left[\frac{c^2}{4\sigma_{\text{refl}}^2} + \frac{G(N_u\lambda_1)^2}{16}\right]\frac{\partial^2 E_n}{\partial \tau^2}\,.
\end{aligned}
\tag{12}
$$

[2] The generalization to complex G and $\sqrt{R}$ is straightforward but messy. For example, the change in power is $|1 + G/2|^2 \approx 1 + (G + G^*)/2$ if G is complex.

Moving E_n to the left-hand side and setting $E_{n+1} - E_n \approx \partial E_n/\partial n$ leads to a linear partial differential equation for the field $E_n(\tau)$. This PDE can be solved by the separation of variables technique, which leads to exponential dependence on *n*, while the temporal variation is described by Hermite–Gauss functions. We index these linear modes by p and find that the general solution can be written as a sum over the 'supermodes'

$$E_n^p(\tau) = \exp\left[\left(\frac{G-\alpha}{2}\right)n - \left(\frac{2D^2\sigma_{\text{filter}}^2}{c^2} + \frac{c(1+2p)\sqrt{G}}{2\sigma_z\sigma_{\text{filter}}}\right)n\right]$$
$$\times\, \mathrm{e}^{-2\sigma_{\text{filter}}^2 D\tau/c^2} \exp\left[-\frac{\sqrt{G}c\sigma_{\text{filter}}}{\sigma_z}\tau\right] H_p\left(G^{1/4}\sqrt{\frac{c\sigma_{\text{filter}}}{\sigma_z}}\tau\right), \tag{13}$$

where we have defined the net detuning length $D \equiv \ell + GN_u\lambda_1/4$ and the effective filtering bandwidth σ_{filter} via

$$\frac{1}{\sigma_{\text{filter}}^2} \equiv \frac{1}{\sigma_{\text{refl}}^2} + \frac{GN_u^2\lambda_1^2}{16}. \tag{14}$$

The first line in Eq. (13) indicates that the exponential power growth is reduced from its nominal value $G - \alpha$ (gain minus loss) if the total detuning length $D \neq 0$; this condition shows one effect of lethargy since maximum gain is achieved when the cavity length is reduced slightly from its nominal synchronous length (i.e., $D = 0$ implies that $\ell < 0$). Significant FEL gain requires the total detuning to be within the effective oscillator bandwidth such that $D\sigma_{\text{filter}} \ll 1$. Additionally, setting $D = 0$ shows that the gain approaches the infinite beam limit only if the electron beam is also significantly longer than the inverse bandwidth $1/\sigma_{\text{filter}}$. For shorter electron bunches, only the fraction of current whose spectral content lies within the effective bandpass set by either the mirror σ_{refl} or the slippage $4/(N_u\lambda_1\sqrt{G})$ contributes to the gain.

The RMS width of the *p*th mode is proportional to the geometric mean of the e-beam size and $1/\sigma_{\text{filter}}$, with temporal width $\sim\sqrt{(1+2p)\sigma_z/c\sigma_{\text{filter}}}$. When the electron bunch is long there are many longitudinal modes with comparable growth rates, and the oscillator output is comprised of a superposition of supermodes whose total bandwidth $\sim 1/N_u$ and temporal duration $\sim \sigma_z$. As the evolution proceeds through many passes, however, the lowest-order ($p = 0$) Gaussian mode with largest gain will eventually become dominant. If the mirror is essentially wavelength independent, $\sigma_{\text{refl}} \gg 1/N_u\lambda_1$, our discussion in the beginning of this chapter applies and the output bandwidth will approach the limiting value Eq. (8), albeit slowly.

On the other hand, we will see that the crystal mirrors that enable FEL oscillators in the X-ray spectral region have $\sigma_{\text{refl}} \ll 1/N_u\lambda_1$ (typically $N_u \lesssim 3\times10^3$ while $\sigma_{\text{refl}}/\omega_1 \sim 10^{-5}$ to 10^{-7}). In this case, $\sigma_{\text{filter}} \to \sigma_{\text{refl}}$ which simplifies some of the preceding discussion. For example, the lowest-order (Gaussian) supermode simplifies to

$$E_n^0(\tau) = \exp\left[\frac{1}{2}\left(G - R - \frac{4\ell^2\sigma_{\text{refl}}^2}{c^2} - \frac{c\sqrt{G}}{\sigma_z\sigma_{\text{refl}}}\right)n\right]\mathrm{e}^{-2\sigma_{\text{refl}}^2 l\tau/c^2}\mathrm{e}^{-\tau^2/2\sigma_0^2}, \tag{15}$$

where the mean square temporal width $\sigma_0^2 \equiv \sigma_z/(\sqrt{G}c\sigma_{\text{refl}})$. Hence, in this case the gain is reduced from its nominal value if either the cavity length shift ℓ or the electron beam width σ_z is smaller than the inverse bandwidth of the mirror c/σ_{refl}. The steady state temporal width is given by $\sigma_0 \propto \sqrt{\sigma_z/c\sigma_{\text{refl}}}$ with final bandwidth $\propto \sqrt{c\sigma_{\text{refl}}/\sigma_z}$, and it now requires $N_{\text{FT}} \sim 2\sigma_{\text{refl}}\sigma_z/c \ll 4\pi\sigma_z/(\lambda_1 N_u)$ passes to reach this steady state.

In addition to modifying the supermode behaviour, the additional spectral filtering provided by the mirrors also completely suppresses the sideband/synchrotron instability, eliminating the unstable and chaotic ‘spiking mode’ of operation observed at lower wavelengths, example [6, 7]. This is because the sideband instability amplifies frequency content near that associated with the synchrotron period, i.e., with frequencies at $= \omega_1 \pm \omega_s$, where

$$\omega_s \sim \frac{\lambda_u}{\lambda_1} c k_s = \frac{\lambda_u}{\lambda_1}\frac{c}{L_u}\sqrt{\epsilon} = \frac{c}{N_u \lambda_1}\sqrt{\epsilon}\,, \tag{16}$$

where ϵ is the normalized field strength. At saturation $\epsilon \sim 1$, so that the characteristic frequency of the sideband/‘spiking’ mode is

$$\omega_s \sim \frac{c}{N_u \lambda_1} \gg \sigma_{\text{refl}}\,, \tag{17}$$

and the narrow bandwidth of the XFELO crystal mirrors effectively filters out the sideband instability.

4 Transverse physics of the optical cavity

When the gain is small, the transverse mode is typically well described by the vacuum resonator modes of the cavity. We will briefly describe some of the transverse cavity physics in the limit of Gaussian optics, which assumes that angles from the optical axis are small (paraxial) and that optical elements can be treated as producing linear transformations to the field. The linear transformations propagate the radiation brightness/Wigner function along rays, which implies that we can analyse the cavity modes using the same matrix formulation for particle beams. Under these limiting circumstances, the transformations act on the (pseudo)-distribution of rays in the position–angle phase space $(\boldsymbol{x}, \boldsymbol{\phi})$, and the wave behaviour can be described by referencing only the propagation of rays[3].

In the laser community the matrix approach is referred to as the ABCD-matrix method, see Ref. [8], and typically these matrix elements are used to derive the stable Rayleigh range and wavefront curvature for Hermite–Gaussian cavity modes. The lowest-order mode is analogous to a Gaussian particle beam with emittance $\lambda/4\pi$, while its Rayleigh length at the waist $Z_R = \sigma_r^2/(\lambda/4\pi) = (\lambda/4\pi)\sigma_{r'}^2$ is equivalent to the Courant–Snyder beta function from particle optics.

In order to understand the transverse X-ray profile, we first consider the simple two-mirror resonator. We model this optical cavity as containing one ideal mirror of focal length f, such that the round-trip distance in the cavity is L_c. We restrict our discussion to the two-dimensional phase space (x, ϕ), and using the fact that the matrices associated with a drift length ℓ and a focusing mirror are, respectively,

$$\mathsf{L}(\ell) = \begin{bmatrix} 1 & \ell \\ 0 & 1 \end{bmatrix}, \qquad \mathsf{F}(f) = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}. \tag{18}$$

Stable resonator modes exist when the RMS size, divergence, and correlation are periodic over one round trip through the cavity. These mode sizes can be determined from the matrix map that starts and ends in the middle of the undulator; for the two-mirror resonator this is given by $\mathsf{M}_{2\text{res}} = \mathsf{L}(L_c/2)\mathsf{F}(f)\mathsf{L}(L_c/2)$. The matrix $\mathsf{M}_{2\text{res}}$ maps (x, ϕ) from and back to the centre of the cavity such that

$$\begin{bmatrix} x \\ \phi \end{bmatrix}_{\text{out}} = \mathsf{M}_{1\text{res}} \begin{bmatrix} x \\ \phi \end{bmatrix}_{\text{in}}, \tag{19}$$

[3] Non-ideal elements, apertures, and nonlinear transformations introduce interference effects that may not be well described by the methods presented here.

while the second-order moment matrix Σ_{out} at the output plane is related to the initial Σ_{in} via

$$\Sigma_{\text{out}} \equiv \begin{bmatrix} \langle x^2 \rangle & \langle x\phi \rangle \\ \langle \phi x \rangle & \langle \Phi^2 \rangle \end{bmatrix}_{\text{out}} = \mathsf{M}_{2\text{res}} \begin{bmatrix} \langle x^2 \rangle & \langle x\phi \rangle \\ \langle \phi x \rangle & \langle \phi^2 \rangle \end{bmatrix}_{\text{in}} \mathsf{M}_{2\text{res}}^T$$

$$= \mathsf{M}_{2\text{res}} \Sigma_{\text{in}} \mathsf{M}_{2\text{res}}^T \,. \tag{20}$$

Equating Σ_{out} and Σ_{in} implies that at the cavity middle the correlation vanishes (the radiation has a waist), and that the cavity round trip length L_c and mirror focal length f are related to the trapped mode Rayleigh length through the following relation:

$$f = \frac{L_c}{4} + \frac{1}{L_c} \frac{\langle x^2 \rangle_{\text{in}}}{\langle \phi \rangle_{\text{in}}} = \frac{L_c}{4} + \frac{z_R^2}{L_c} \,. \tag{21}$$

Note that stable operation requires $f > L_c/4$, which in terms of the mirror's radius of curvature is $2f = r > L_c/2$. This inequality can be violated if there is sufficient FEL amplification, but for low-gain devices it provides a good starting point for optical cavity design.

References

[1] H. Al-Abawi, F.A. Hoff, G.T. Moore and M.O. Scully, *Opt. Comm.* **30** (1979) 235. https://doi.org/10.1016/0030-4018(79)90085-3

[2] K.-J. Kim, *Phys. Rev. Lett.* **66** (1991) 2746. https://doi.org/10.1103/physrevlett.66.2746

[3] G. Dattoli, G. Marino, A. Renieri and F. Romanelli, *IEEE J. Quantum Electron.* **17** (1981) 1371. https://doi.org/10.1109/jqe.1981.1071268

[4] P. Elleaume, *IEEE J. Quantum Electron.* **21** (1985) 1012. https://doi.org/10.1109/jqe.1985.1072746

[5] T.M. Antonsen and B. Levush, *Phys. Fluids B: Plasma Phys.* **1** (1989) 1097. https://doi.org/10.1063/1.858980

[6] R.W. Warren, J.E. Sollid, D.W. Feldman, W.E. Stein, W.J. Johnson, A.H. Lumpkin and J.C. Goldstein, *Nucl. Instrum. Methods Phys. Res., Sect. A* **285** (1989) 1. https://doi.org/10.1016/0168-9002(89)90417-8

[7] R. Hajima, N. Nishimori, R. Nagai and E.J. Minehara, *Nucl. Instrum. Methods Phys. Res., Sect. A* **475** (2001) 270. https://doi.org/10.1016/s0168-9002(01)01621-7

[8] A.E. Siegman, *Lasers* (University Science Book, Sausalito, CA 1986).

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

LLRF Controls and Feedback: Free-Electron Lasers and Energy Recovery Linacs

S. Pfeiffer
DESY, Hamburg, Germany

Abstract
The low-level radio frequency (LLRF) system generates the drive sent to the high-power components. In this paper, we give the basics of LLRF controls and feedback. This includes a brief introduction to the two concepts for this school, i.e., the energy recovery linac and the free-electron laser, and their main differences, from a LLRF perspective. The physical behaviour of the subcomponents (detector, amplifier, cavity, etc.) within an LLRF feedback loop is outlined, together with its system modelling in the time and frequency domains. System identification using special input–output signals is introduced. Furthermore, stability checks for the closed-loop operation and the feedback controller design in the time and frequency domains are briefly discussed. This paper concludes with some examples for system identification, its feedback design, and the achieved stability at different facilities, showing the need for LLRF control and feedback operation.

Keywords
Low-level radio frequency; LLRF; feedback; system identification; controller design.

1 Introduction

Low-level RF (LLRF) field control is used to stabilize the accelerating field inside normal- and superconducting RF cavities, mainly using digital feedback loops. Optimal feedback regulation is part of this contribution and discussed in this paper. First, a brief introduction to energy recovery linacs and free-electron lasers is given. Here, the main focus is on the differences between both concepts. At the end of this section, a short introduction to the need for LLRF controls is discussed. The mathematical modelling of certain components within the feedback loop (Section 2) is often necessary to optimize the feedback controller. An alternative to this mathematical modelling using differential equations is system identification using special excitation signals and relating them to the corresponding output signal (Section 3). Based on such a system model, the optimal feedback controller can be designed in the time or frequency domain, as shown in Section 4. Examples for the system identification and the feedback controller design are discussed in Section 5, justifying the need of LLRF controls and feedback.

1.1 Energy recovery linac

An energy recovery linac concept is given in the CAS 2016 contribution from Dr. A. Jankowiak (HZB) *Concept of ERL*. At bERLinPro, the beam current will be 100 mA and the accelerating structures are superconducting RF cavities operated in continuous wave mode at an RF frequency of 1.3 GHz [1–3]. Another energy recovery linac, the compact energy recovery linac at KEK, is described in the following [4]. The compact energy recovery linac is an ongoing project at KEK, intended to demonstrate excellent energy recovery linac performance towards a future light source. The compact energy recovery linac is illustrated in Fig. 1. The injector consists of a photocathode d.c. gun, a normal-conducting buncher cavity with a loaded quality factor (Q_L) of 1.1×10^5 and three two-cell superconducting cavities operated with $Q_\mathrm{L} = 5 \times 10^5$. The recirculation loop hosts several magnets (dipoles and quadrupoles), the

2519-8041 – © the Author/s, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.501

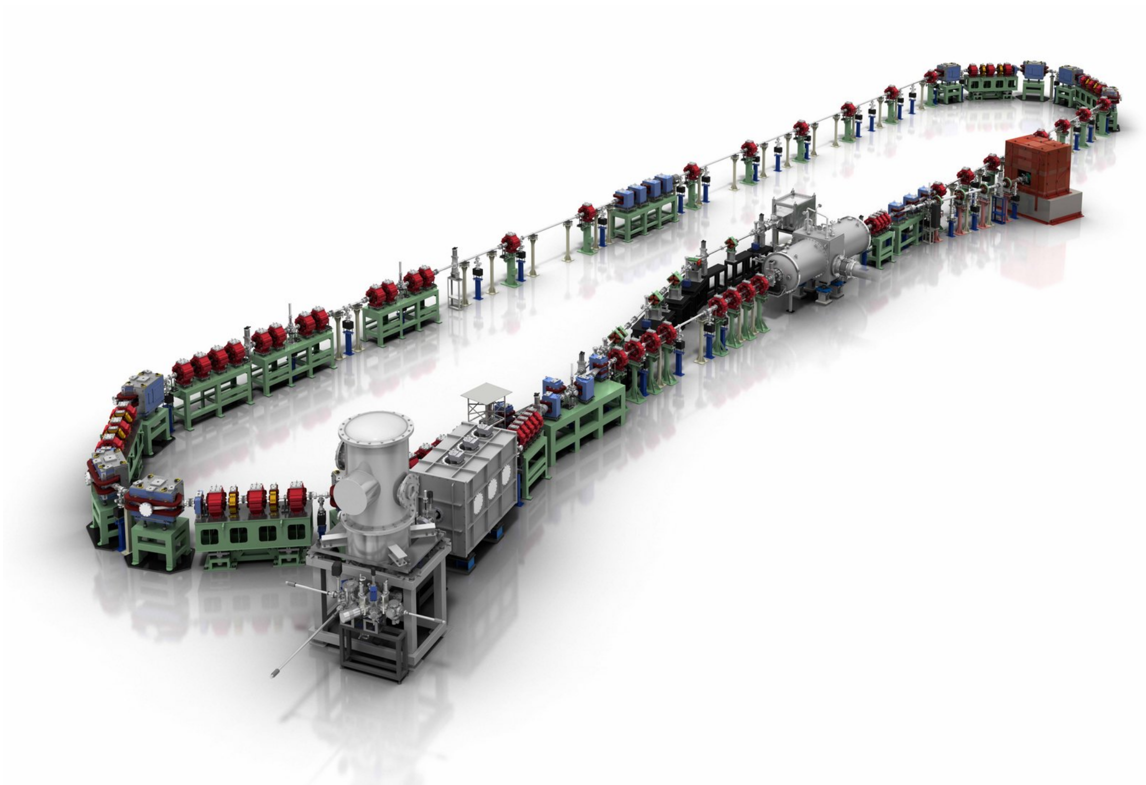


Fig. 1: Compact energy recovery linac ©Rey.Hori/KEK [6]: the injector part (located in the lower left) and the recirculation loop are shown.

Table 1: Example RF and beam parameters for energy recovery linac and free-electron laser

	Compact energy recovery linac	bERLinPro	FLASH
Desired RF stability (r.m.s.)	$\Delta A/A \leq 0.1\%$	$\Delta A/A \leq 0.05\%$	$\Delta A/A \leq 0.01\%$
	$\Delta\varphi \leq 0.1°$	$\Delta\varphi \leq 0.02°$–$0.1°$	$\Delta\varphi \leq 0.01°$
Loaded quality factor Q_{L}	1.1×10^5–1×10^7	1×10^5–5×10^7	1.2×10^5–3×10^6
Nominal beam energy	35 MeV	50 MeV	max. 1.25 GeV
Beam current	$\sim$100 mA	$\sim$100 mA	$\sim$1 mA
Mode of operation	Continuous wave	Continuous wave	Pulsed

beam diagnostics (e.g., beam position monitors), and two nine-cell superconducting cavities operated with $Q_{\mathrm{L}} = 1 \times 10^7$. The RF frequency for all cavities is 1.3 GHz. For the compact energy recovery linac, a RF field stability of 0.1% in amplitude and $0.1°$ in phase is required [5]. The compact energy recovery linac is operated in continuous wave mode with a beam current of up to 100 mA; normal- and superconducting RF structures, powered with klystrons and solid state amplifiers, are used to accelerate the beam to its target energy.

Energy recovery linacs are widely used in research facilities. The beam current is of the order of 100 mA; it is foreseen to increase to the ampere regime. Normal- and superconducting RF cavities are used to reach the final target energy. The cavities are driven by klystrons, solid state amplifiers, or inductive output tubes. The key parameters of this compact energy recovery linac at KEK and the bERLinPro at HZB are given in Table 1.

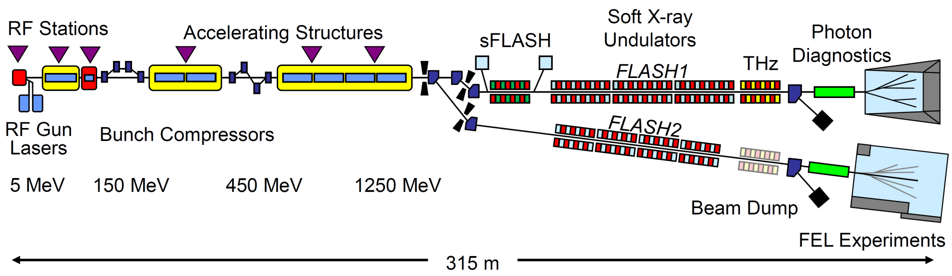


Fig. 2: FLASH: beam direction is from left to right. Both beamlines are shown, the old FLASH1 beamline and the new FLASH2 beamline [7].

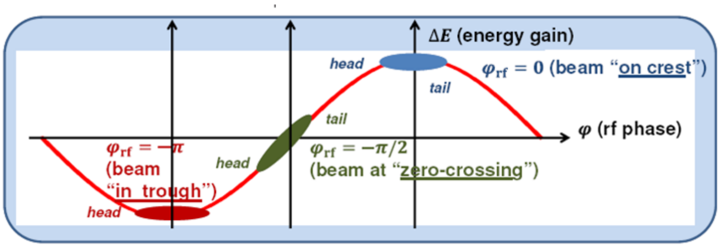


Fig. 3: Location of electron bunch for different phase settings [8]

1.2 Free-electron laser

The free-electron laser (FLASH) is the world's first soft X-ray free-electron laser (XFEL). It is a linear accelerator with a total length of about 315 m [7]. Electron bunches are generated using an external laser and a photocathode inside the normal-conducting RF-gun ($Q_\mathrm{L} \approx 1.2 \times 10^5$). They are further accelerated by 56 superconducting RF cavities operated at 1.3 GHz with $Q_\mathrm{L} = 3 \times 10^6$ in seven cryomodules. Each cryomodule is depicted in Fig. 2 as a blue box hosting eight superconducting RF cavities. Furthermore, four superconducting RF cavities (red box with one cryomodule) operated at 3.9 GHz, i.e., the third harmonic of 1.3 GHz, are used for phase space linearization. The cavities at FLASH are all driven by klystrons. Up to a maximum of 16 cavities are driven by one klystron, for cost efficiency. FLASH is operated in short-pulse mode with a repetition rate of 10 Hz and 800 µs maximum beam time. Each RF pulse is active for about 1.4 ms (including fill time and beam time), corresponding to a duty factor of about 1%. During the beam time a maximum of 800 electron bunches at 1 MHz repetition rate can be used for experiments. The maximum final bunch energy is 1.25 GeV, corresponding to a fundamental laser light wavelength of 4.12 nm using fixed gap undulators at the FLASH1 beamline. The maximum bunch charge of each electron bunch is of the order of 1 nC; therefore, the maximum beam current is of the order of 1 mA (1 nC bunch at 1 MHz repetition rate), instead of 100 mA for an energy recovery linac, see Table 1.

1.3 LLRF in general—control goal

The LLRF system is used to supply the accelerating cavities with high-power RF signals. Furthermore, it is used to modify or adjust the accelerating field in amplitude and phase (A/P). The desired A/P value depends on the mode of operation, i.e., acceleration, deceleration, or an imposed energy chirp, see Fig. 3. The typical RF stability of an accelerating section is of the order of $\Delta A/A \leq 0.1\%$–0.01% and $\Delta\varphi \leq 0.1°$–$0.01°$.

Consider, e.g., a free-electron laser with a bunch compressor to achieve high peak currents. Its injector laser stability, cavity amplitude, phase stability are directly related to the timing jitter after a bunch compressor defining the achievable free-electron laser performance, see contribution from Dr. M. Divall (PSI) *Lasers in FEL Facilities*. This requires a precise timing and synchronization of the master

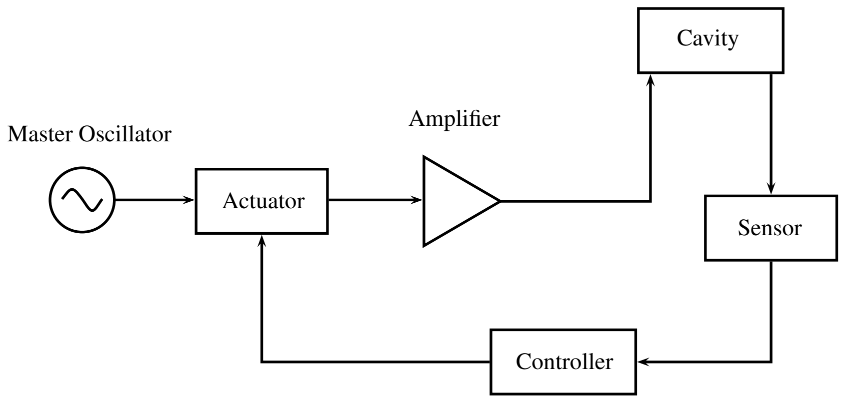


Fig. 4: The LLRF control loop with master oscillator, actuator, amplifier, cavity, and sensor. The controller closes the feedback loop.

clock and all attached subdevices, such as the accelerating modules. For now, we will assume that the timing and synchronization are perfect.

We will start with the description of the closed-loop system to be controlled. An example of a LLRF feedback control loop is depicted in Fig. 4. The master oscillator provides a constant sinusoidal signal with the desired frequency, e.g., 1.3 GHz. This master oscillator signal is amplified from low power (a few volts) to high power to drive the cavity. A sensor, e.g., a pick-up or magnetic loop, is placed inside the cavity to detect a fraction of the RF field. This information is used by a controller to control the amplitude and phase inside the cavity by modulating the amplitude and phase of the master oscillator signal. This closes the feedback loop. The LLRF regulates on a low power signal, while only the part after the amplifier and in the cavity are operated at high power (e.g., in the megavolt regime).

In the following we will discuss the individual blocks separately before modelling the system behaviour to optimize the controller.

2 System description

In this section, we will focus on the individual blocks within a feedback loop for the LLRF system. A brief overview is given in Fig. 4.

2.1 RF detection

The principles of RF field detection are based on mixing a reference signal (LO) with the RF signal to be detected. Here, the goal is to measure the amplitude and phase, or the in-phase (I) and quadrature (Q) part of an RF signal with respect to a reference signal. The relation between A/P and I/Q is:

$$\begin{aligned} I &= A \cdot \cos\varphi & \qquad A &= \sqrt{I^2 + Q^2} \\ Q &= A \cdot \sin\varphi & \qquad \varphi &= \text{atan2}(Q, I)\,, \end{aligned}$$

with

$$\text{atan2}(Q, I) = \begin{cases} \arctan\left(\frac{Q}{I}\right) & \text{if } I > 0\,, \\ \arctan\left(\frac{Q}{I}\right) + \pi & \text{if } I < 0 \text{ and } Q \geq 0\,, \\ \arctan\left(\frac{Q}{I}\right) - \pi & \text{if } I < 0 \text{ and } Q < 0\,, \\ +\frac{\pi}{2} & \text{if } I = 0 \text{ and } Q > 0\,, \\ -\frac{\pi}{2} & \text{if } I = 0 \text{ and } Q < 0\,, \\ \text{undefined} & \text{if } I = 0 \text{ and } Q = 0\,. \end{cases}$$

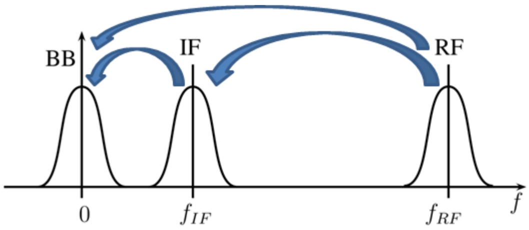


Fig. 5: RF down-sampling directly to baseband (BB) or via an intermediate frequency (IF)

The high-frequency signal is down-converted to the baseband. Different amplitude and phase conditions between the RF and the LO are responsible for the corresponding output signal (A/P or I/Q). Only a short overview of different detection schemes is given; detailed information can be found in Ref. [9].

The briefly discussed detection schemes are:

1. direct amplitude and phase detection:
   - no down-conversion;
   - analogue or digital (up to 800 MHz analogue-to-digital converters);
2. baseband sampling (analogue I/Q detector);
3. digital I/Q sampling;
4. intermediate-frequency sampling (non-I/Q sampling).

Scheme (1) is a direct sampling scheme, while schemes (2)–(4) are based on mixing a reference signal (LO) with the RF signal, meaning that the RF signal is down-converted into baseband directly or using an intermediate frequency, see Fig. 5.

2.1.1 Analogue direct amplitude and phase detection

The analogue amplitude and phase detection are achieved by splitting the RF signal into two parts, see Fig. 6. The first part is used to detect the amplitude using a diode. The second part is mixed with a frequency- and phase-stable LO signal (the same frequency as the expected RF signal), leading, for small angle variations, to a signal proportional to the phase difference between the RF and LO signal. The mathematical description for the phase signal is given by:
Input signals:

$$V_{\mathrm{RF}}(t) = A_{\mathrm{RF}} \sin(\omega t + \varphi_0)\,,$$
$$V_{\mathrm{LO}}(t) = A_{\mathrm{LO}} \cos(\omega t)\,,$$

Output signal:

$$V_{\mathrm{mixer}} = A_{\mathrm{RF}} \sin(\omega t + \varphi_0) \cdot A_{\mathrm{LO}} \cos(\omega t) = \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \left(\sin(\varphi_0) + \sin(2\omega t + \varphi_0)\right)\,,$$

Output signal after low-pass filtering:

$$V_{\mathrm{LPF}} = \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \sin(\varphi_0) \approx \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \varphi_0 \quad \text{(for small } \varphi_0) \quad .$$

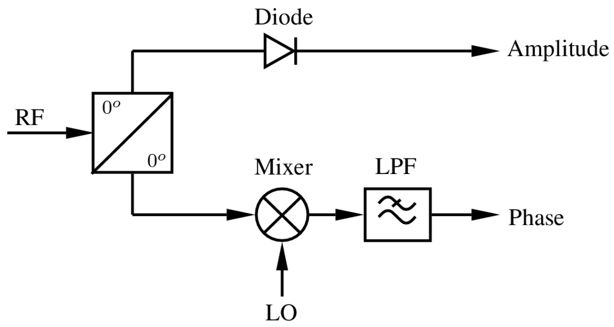


Fig. 6: Basic principle of analogue direct amplitude and phase detection

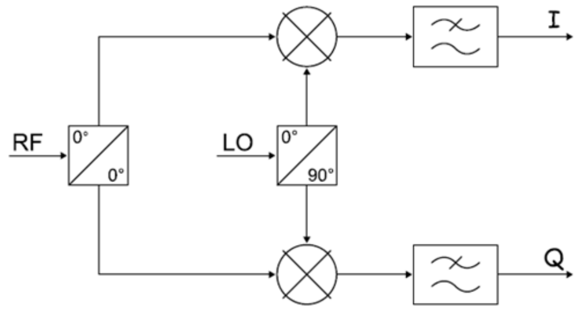


Fig. 7: Analogue I/Q detection [9]

2.1.2 Digital direct amplitude and phase detection

Nowadays, high-speed analogue-to-digital converters are available with sampling rates of more than 500 megasamples per second and analogue input bandwidth of up to 1.5 GHz. These analogue-to-digital converters enable the RF signal to be sampled directly without any RF converter or intermediate-frequency scheme, with a signal-to-noise ratio of more than 50 dB. The advantage is the reduction of the number of input stages for pre-configuration (down-conversion, filter, amplifier, etc.), reducing possible errors and noise sources and increasing reliability. The disadvantage is the increasing impact of the clock jitter on the signal-to-noise ratio for higher input frequencies. This effect is linear and can be reduced by the process gain due to digital band-limitation and averaging as a function of the square root. If digital filtering is used to filter out noise components outside the bandwidth, then the process gain must be included in the equation to account for the resulting increase in the signal-to-noise ratio using an N-bit analogue-to-digital converter, as shown in Eq. (1):

$$\text{signal-to-noise ratio} = 6.02 \times N + 1.76\ \text{dB} + 10 \log_{10} \frac{f_\text{s}}{2 \times f_\text{BW}}\ . \tag{1}$$

See, in addition, Ref. [10] for the noise characteristics and measurements performed at FLASH.

2.1.3 Baseband sampling (analogue I/Q detector)

This concept is an analogue I/Q detector, in which a direct conversion from RF to baseband is performed. The RF signal is multiplied by the LO signal resulting in analogue I/Q signals. Here, the LO signal is split by a hybrid and one signal is given a phase difference of 90°, see Fig. 7. Two analogue-to-digital converters are required for the digitization of an I/Q pair; hence, it requires more space and leads to higher cost than using only one analogue-to-digital converter, and the reliability is reduced. The main problem of this scheme is the phase-dependent amplitude detection due to I/Q imbalances and offsets, see Fig. 8.

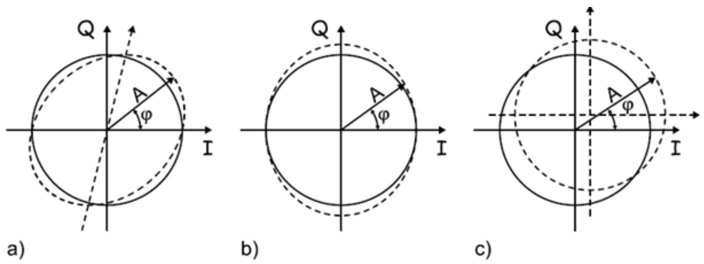


Fig. 8: Constellation diagram with different errors in I and Q. (a) The output signals of the mixers are not exactly I and Q, i.e., the phase difference is not $90°$. (b) Gain mismatch between I and Q mixer branch. (c) I/Q offset at the mixer outputs [9].

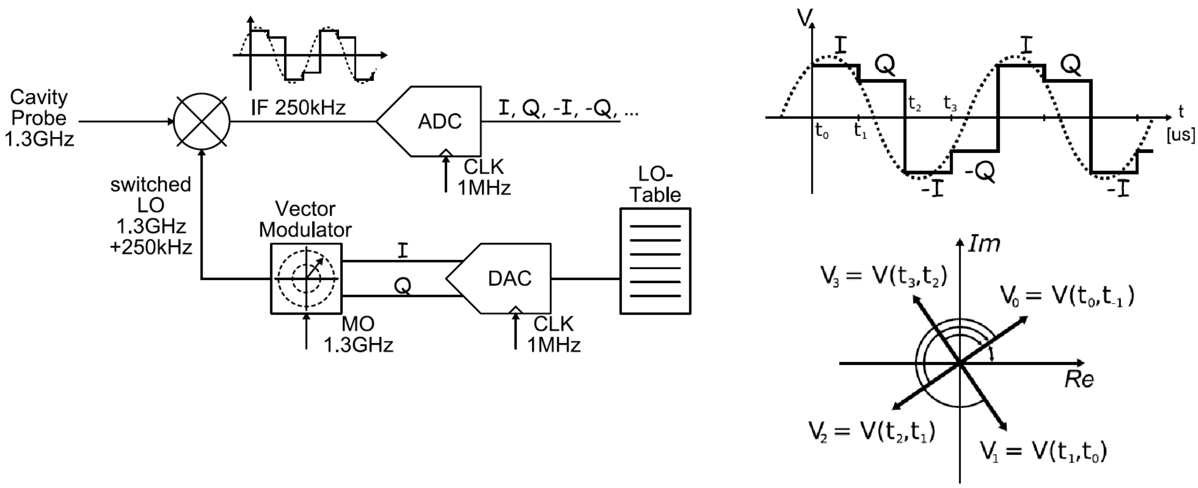


Fig. 9: Left-hand side: digital I/Q sampling scheme. Right-hand side: corresponding output sequence in time and complex domains [9].

2.1.4 Digital I/Q sampling

This detection scheme is similar to baseband sampling, but here only one analogue-to-digital converter and a switched LO are required, see Fig. 9. The LO signal after each sampling is switched in phase by $90°$ for a fixed time interval, leading to an output signal that represents a series of $I, Q, -I, -Q, \ldots$ and so on. If the signal to be detected is of low bandwidth, or if it can be assumed that the signal does not change significantly between two samples, the field vector can be computed by two consecutive samples (I/Q value). To do this, the signal needs to be shifted by $n \times 90°$ ($n = 0, 1, 2, 3$). The problem for this scheme is that the Nyquist frequency is $f_s/4$. Furthermore, owing to the switching with rectangular output signal, the entire signal chain needs a relatively high bandwidth.

2.1.5 Intermediate-frequency sampling (non-I/Q sampling)

In contrast with the conversion directly from RF to baseband, the RF can be first mixed down to an intermediate frequency and sampled. The detection scheme is depicted in Fig. 10. Based on the sampled analogue-to-digital converter signals, the actual I/Q detection is achieved by multiplication of the input signal by a sine and a cosine function on a digital level. The advantage of this scheme is that, e.g., phase and gain imbalance of the two multipliers of an analogue IQ-detector do not exist. For a series of samples, the series of $I[i]$ and $Q[i]$ are calculated by:

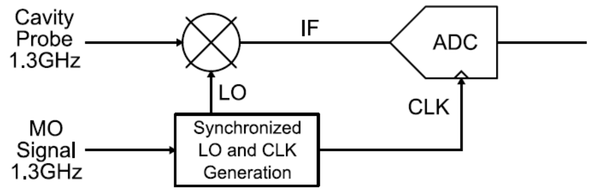


Fig. 10: Intermediate-frequency (IF) sampling principle with LO and clock (CLK) generation synchronized to the master oscillator (MO) [9]

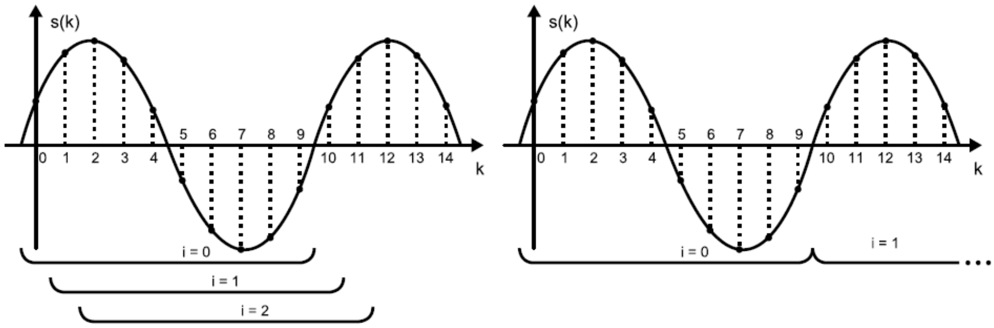


Fig. 11: Intermediate-frequency sampling: left-hand side, with sliding window detection; right-hand side, with step window detection [9]

(1) sliding window detection (left-hand plot in Fig. 11):

$$I(i) = \frac{2}{M} \sum_{k=0}^{M-1} s(k+i) \cos((k+i)\Delta\varphi) \ ,$$
$$Q(i) = \frac{2}{M} \sum_{k=0}^{M-1} s(k+i) \sin((k+i)\Delta\varphi) \ ;$$

(2) step window detection (right plot in Fig. 11):

$$I(i) = \frac{2}{M} \sum_{k=0}^{M-1} s(k+iM) \cos((k+iM)\Delta\varphi) \ ,$$
$$Q(i) = \frac{2}{M} \sum_{k=0}^{M-1} s(k+iM) \sin((k+iM)\Delta\varphi) \ .$$

For further details, see, e.g., Ref. [9].

2.2 RF manipulation using a vector modulator

A vector modulator is often used to up-convert a baseband signal in I/Q coordinates back to an RF signal. The basic operation of a vector modulator is shown in Fig. 12. The signal from the master oscillator $y_{\mathrm{in}}(t)$ is split into two parts, i.e., a 0° and a 90° part. The baseband I/Q values are multiplied with the corresponding sine and cosine parts. The sum of both signals leads to the RF output signal, i.e., the signal to the klystron, which is modulated by the I/Q components. In this way, the amplitude and phase of the RF output signal can be modulated by an I/Q pair with respect to the RF input signal. The bandwidth of a vector modulator is usually tens of megahertz [11].

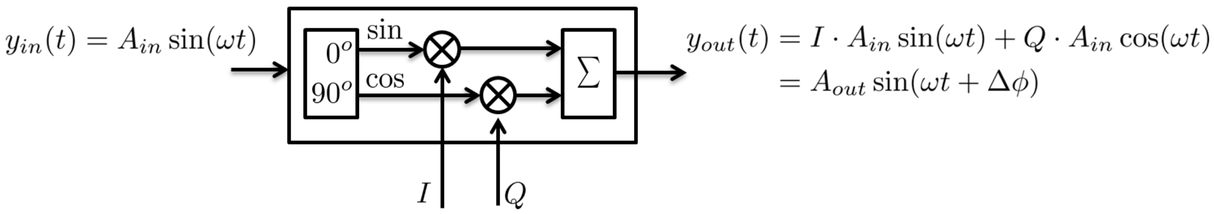


Fig. 12: Principle of vector modulator

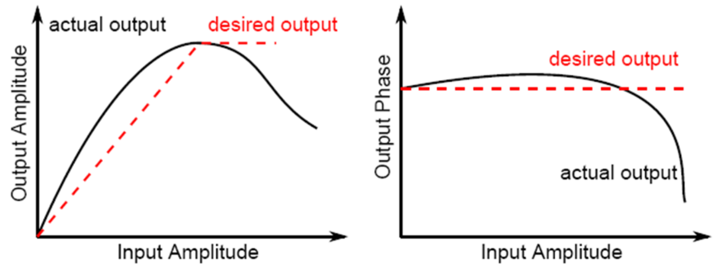


Fig. 13: Desired and real amplifier characteristics [12]

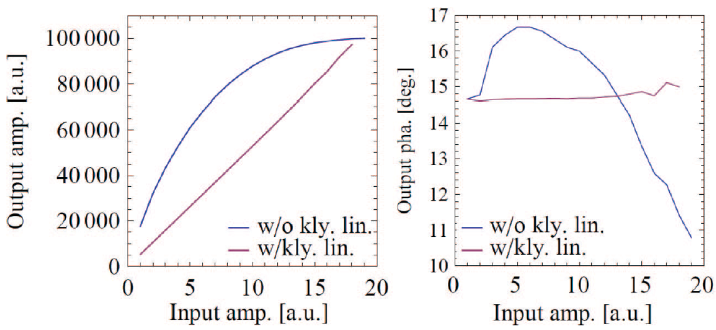


Fig. 14: Result of iSim simulation without klystron linearization (blue) and with linearization (purple). Left-hand side: klystron linearization algorithm output amplitude [arbitrary units] versus input amplitude [arbitrary units]. Right-hand side: klystron linearization algorithm output phase [degrees] versus input amplitude [arbitrary units] [12].

2.3 Amplifier

Each amplifier, e.g., a klystron, solid state amplifier, or inductive output tube, is itself a non-linear device. The non-linearity is caused by, e.g., saturation at maximum output power. Furthermore, non-linearities in terms of changes of the output amplitude and phase are mostly present if only the input amplitude is changed, see, for example, Fig. 13. For most applications, such as superconducting RF cavity control, it is sufficient to linearize the amplifier input–output (I/O) behaviour using a static I/O characteristic. An example is shown in Fig. 14, where a simulation of a real and corrected klystron characteristic is shown. The bandwidth of a klystron is of the order of 10 MHz [11]. Refer to, e.g., Ref. [12] for further details of the klystron linearization.

2.4 Cavity

This section gives a brief overview of the mathematical modelling for a TESLA-type RF cavity; detailed information can be found in Ref. [11].

Fig. 15: TESLA-type superconducting RF cavity [13]

Table 2: Parameters for superconducting RF Tesla-type cavity

Operating frequency	1.3 GHz
Length	1.036 m
Aperture diameter	70 mm
Cell to cell coupling	$\approx 2\%$
Quality factor Q_0	$\approx 10^{10}$
$r/Q := r_{\mathrm{sh}}/Q_0$	1036 Ω

2.4.1 Introduction

A typical nine-cell TESLA-type superconducting RF cavity is shown in Fig. 15. The cavity has a length of about 1 m and is operated with a driving frequency of 1.3 GHz; its typical parameters are shown in Table 2.

2.4.2 Cavity modelling by RCL circuit

The TESLA-type superconducting RF cavity consists of nine electrically coupled cells. An equivalent RCL circuit model with nine magnetically coupled cells is shown in Fig. 16. Such a structure with nine coupled cells has nine normal modes, the so-called passband modes. The fundamental operating mode is called the π-mode, meaning that the RF fields in adjacent cells are at a phase difference of π. The closest mode to the operating mode is the $8\pi/9$-mode, with a frequency separation of about 800 kHz. A plot of all the modes is shown in Fig. 17. Every passband mode can be modelled by an RCL circuit, which is in parallel with the others. To simplify the modelling, let's first consider only the π-mode and neglect all the other modes. Then the nine-cell structure simplifies to a single cell, displayed in Fig. 18, with generator driving term $\tilde{I}_{\mathrm{g}}$, beam-induced driving term I_{b}, and an external load Z_{ext}. Here, the transmission line, circulator, and coupler are already part of the model and can be neglected. The total current going into the RCL circuit must be the same as the current inside the RCL circuit, i.e.,

$$I_C + I_{R_{\mathrm{L}}} + I_L = I = \tilde{I}_{\mathrm{g}} + I_{\mathrm{b}} \qquad (I_{R_{\mathrm{L}}} = I_R + I_{Z_{\mathrm{ext}}}, \text{ see Eq. (5)})\,, \tag{2}$$

using Kirchhoff's rule. Inserting the well-known relationships

$$\frac{\mathrm{d}I_L}{\mathrm{d}t} = \dot{I}_L = \frac{V}{L}\,, \qquad \dot{I}_{R_{\mathrm{L}}} = \frac{\dot{V}}{R_{\mathrm{L}}}\,, \qquad \text{and} \qquad \dot{I}_C = C\ddot{V} \tag{3}$$

leads to the differential equation for a driven RCL circuit:

$$\ddot{V}(t) + \frac{1}{R_{\mathrm{L}}C}\dot{V}(t) + \frac{1}{LC}V(t) = \frac{1}{C}\dot{I}(t)\,. \tag{4}$$

Here, the external load Z_{ext} and the cavity resistor R are combined to give the so-called loaded shunt impedance

$$R_{\mathrm{L}} = \left(\frac{1}{R} + \frac{1}{Z_{\mathrm{ext}}}\right)^{-1} = \frac{R}{1+\beta} \quad \text{, with coupling factor} \quad \beta = \frac{R}{Z_{\mathrm{ext}}}\,. \tag{5}$$

The loaded quality factor is defined similarly to the loaded shunt impedance as

$$Q_{\mathrm{L}} = \left(\frac{1}{Q_0} + \frac{1}{Q_{\mathrm{ext}}}\right)^{-1} = \frac{Q_0}{1+\beta} \quad \left(\beta = \frac{Q_0}{Q_{\mathrm{ext}}}\right). \tag{6}$$

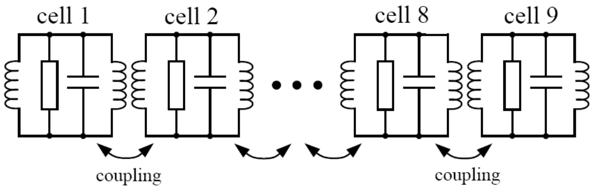


Fig. 16: Equivalent RCL circuit for nine-cell cavity modelled with nine magnetically coupled loops [11]

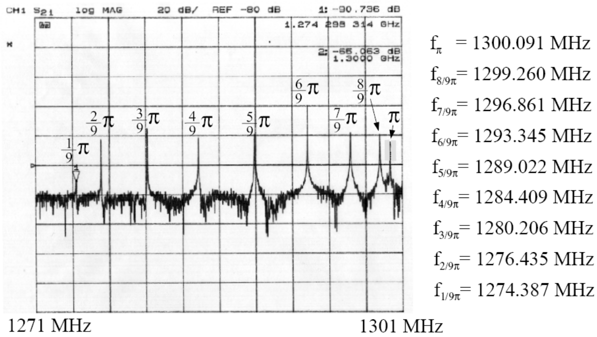


Fig. 17: Passband modes of TESLA-type nine-cell cavity [11]

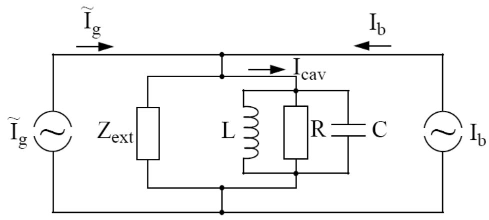


Fig. 18: Simplified equivalent RCL circuit for single-cell cavity with external load and, as driving source, generator and beam [11].

The inductance L and capacitance C can be described by the measurable cavity quantities Q_L and ω_0 with

$$\frac{1}{R_\mathrm{L}C} = \frac{\omega_0}{Q_\mathrm{L}} \qquad \text{and} \qquad \frac{1}{LC} = \omega_0^2\,. \tag{7}$$

Using these relations in Eq. (4) gives the well-known cavity differential equation,

$$\ddot{V}(t) + \frac{\omega_0}{Q_\mathrm{L}}\dot{V}(t) + \omega_0^2 V(t) = \frac{\omega_0 R_\mathrm{L}}{Q_\mathrm{L}}\dot{I}(t)\,, \tag{8}$$

with harmonic RF driving term $I(t) = \hat{I}_0 \sin(\omega t)$, e.g., with driving frequency $\omega = 2\pi \times 1.3$ GHz.

The solution of the cavity differential equation (Eq. (8)) for a driving term $I(t) = \hat{I}_0 \sin(\omega t)$ will be proportional to $\cos(\omega t)$. Its stationary solution, i.e., the particular solution of the cavity differential equation, is given by

$$V(t) = \hat{V}(t) \cdot \sin(\omega t + \psi) \tag{9}$$

with

$$\hat{V} = \frac{R_\mathrm{L}\hat{I}_0}{\sqrt{1+\tan^2(\psi)}} \approx \frac{R_\mathrm{L}\hat{I}_0}{\sqrt{1+\left(2Q_\mathrm{L}\frac{\Delta\omega}{\omega}\right)^2}}\,, \tag{10}$$

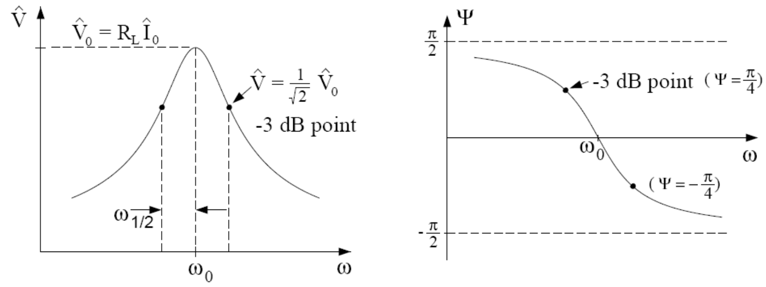


Fig. 19: Resonance curves for amplitude and phase [11]

and ψ as angle between the driving current and cavity voltage; $\psi(t) = \angle(I(t), V(t))$. Furthermore, it is shown in Ref. [11] that the tuning angle can be approximated if the generator frequency ω is very close to the cavity resonance frequency ω_0 ($\omega \approx \omega_0$) by $\tan\psi \approx 2Q_\mathrm{L}\frac{\Delta\omega}{\omega}$, with $\Delta\omega = \omega_0 - \omega \ll \omega$. The frequency dependence of the amplitude is known as the Lorentz curve, shown in Fig. 19. The half bandwidth $\omega_{1/2}$ is defined as the frequency offset where the voltage drops to $1/\sqrt{2}$ (-3 dB) from its maximum. The corresponding phase at the -3 dB point will have an offset of $45°$ $\left(\left|\psi(\omega_{1/2})\right| = \pi/4\right)$. The half bandwidth of a cavity is inversely proportional to the loaded quality factor Q_L and is given by:

$$\omega_{1/2} = \frac{\omega_0}{2Q_\mathrm{L}} = \frac{1}{\tau}(\ll \omega), \quad \text{with time constant } \tau.$$

The definition of half bandwidth can be used to rewrite Eq. (10) as:

$$\ddot{V}(t) + 2\omega_{1/2}\dot{V}(t) + \omega_0^2 V(t) = \omega_{1/2} R_\mathrm{L} \dot{I}(t)\,. \tag{11}$$

2.4.3 *Baseband model*

The high-frequency part from the generator, i.e., $\sin(\omega t)$ or, more generally, $\mathrm{e}^{\mathrm{j}\omega t}$, is not of interest in the case of a down-conversion regulation scheme. For this reason, the high-frequency part can be separated from the slow variations using

$$\begin{aligned} V &= \vec{V}(t)\mathrm{e}^{\mathrm{j}\omega t}; \qquad \vec{V}(t) = V_I + \mathrm{j}V_Q\,, \\ I &= \vec{I}(t)\mathrm{e}^{\mathrm{j}\omega t}; \qquad \vec{I}(t) = I_I + \mathrm{j}I_Q\,. \end{aligned} \tag{12}$$

Consider the phasor diagram in Fig. 20. Both the driving current $\vec{I}(t)$ and the cavity voltage $\vec{V}(t)$ have the same driving frequency component ω. The cavity voltage and phase with respect to the driving current depend on the tuning angle $\psi(t)$, see Eq. (10). Inserting Eq. (12) in Eq. (11) and neglecting the second-order time derivatives of $V(t)$ leads to the first-order differential cavity equation for the envelope (cavity baseband equation) in the complex domain, as

$$\dot{\vec{V}} + (\omega_{1/2} - \mathrm{j}\Delta\omega)\,\vec{V} = \omega_{1/2} R_\mathrm{L}\,\vec{I}\;\;. \tag{13}$$

Separating its real and imaginary parts gives the two first-order equations:

$$\begin{aligned} \dot{V}_I + \omega_{1/2}V_I + \Delta\omega V_Q &= R_\mathrm{L}\omega_{1/2}I_I\,, \\ \dot{V}_Q + \omega_{1/2}V_Q - \Delta\omega V_I &= R_\mathrm{L}\omega_{1/2}I_Q\,. \end{aligned} \tag{14}$$

The latter is well known in accelerator physics and can be rewritten in a so-called state space form,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} V_I(t) \\ V_Q(t) \end{bmatrix} = \begin{bmatrix} -\omega_{1/2} & -\Delta\omega \\ \Delta\omega & -\omega_{1/2} \end{bmatrix}\begin{bmatrix} V_I(t) \\ V_Q(t) \end{bmatrix} + R_\mathrm{L}\cdot\omega_{1/2}\begin{bmatrix} I_I(t) \\ I_Q(t) \end{bmatrix}\,, \tag{15}$$

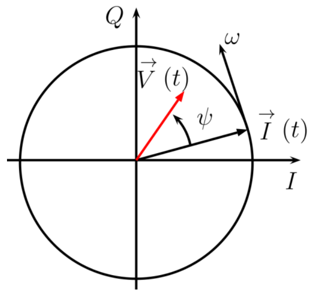


Fig. 20: Phasor diagram of driving current and cavity voltage

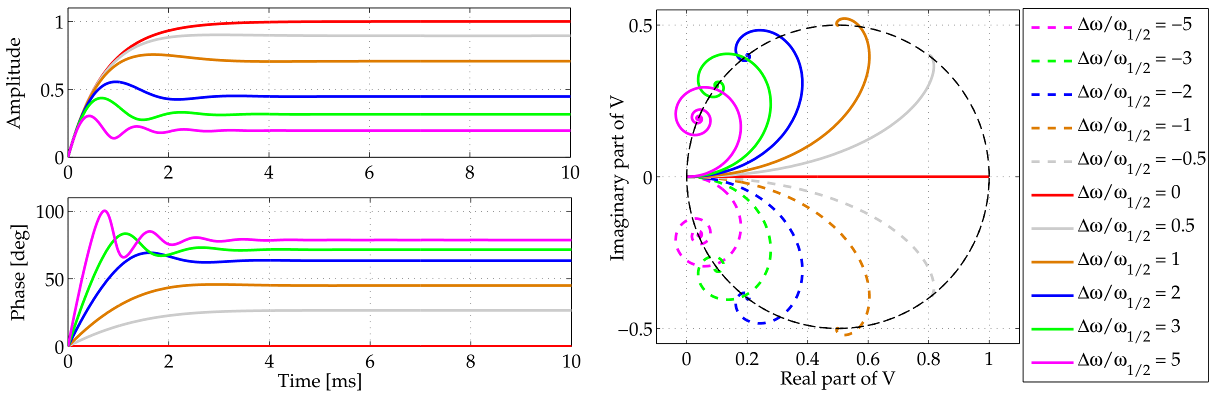


Fig. 21: Transient step responses of a cavity for different normalized detuning values: left-hand side, shown in amplitude and phase; right-hand side, shown in complex plane.

in which V is the complex cavity voltage, I the complex driving current, $\omega_{1/2}$ the half bandwidth, $\Delta\omega = \omega_0 - \omega$ the detuning and R_L the shunt impedance of the cavity. This cavity equation has two inputs (the driving current in I and Q) and two outputs (the cavity voltage in I and Q). We will call this a multi-input, multi-output (MIMO) system.

2.4.4 Step response

Drawing the step response of Eq. (15) for different normalized detuning values $\Delta\omega/\omega_{1/2}$ for an input of which the real part is normalized to one and the imaginary part is zero, i.e., $R_\mathrm{L} \cdot [I_I(t) \quad I_Q(t)]^\mathrm{T} = [1 \quad 0]^\mathrm{T}$, we have the graphs shown in Fig. 21. Here, the left panel of Fig. 21 shows the amplitude and phase of the resulting cavity voltage. The right panel shows the step response as a time-dependent field vector in the complex plane with the steady-state value located on the resonance circle (black dashed line), corresponding to the stationary solution in Eq. (8).

2.4.5 Additional passband modes

The latter cavity equation is given around the baseband frequency, meaning that the operating π-mode is located at frequency zero; Eq. (15) holds only for this mode. Eight other passband modes (see Fig. 17) are present for a nine-cell cavity. Thus, all the differential equations for each mode have to be added to get the overall cavity model; remember that all cavity mode models are in parallel. The field probe signal is measured at the end cell of the cavity; thus, each even mode, e.g., the $8/9\pi$-mode, is sign-inverted such that each fundamental mode is represented in the complex domain—compare this with Eq. (13) for the π-mode—as

$$\dot{\vec{V}} + \left((\omega_{1/2})_{\frac{n}{9}\pi} - \mathrm{j}(\Delta\omega)_{\frac{n}{9}\pi}\right)\vec{V} = (-1)^{n+1} K_{\frac{n}{9}\pi}(\omega_{1/2})_{\frac{n}{9}\pi} \cdot R_\mathrm{L}\,\vec{I}\,, \qquad n = 1, \ldots, 9\,.$$

The extended state space form, see Eq. (15), is given as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} V_I(t) \\ V_Q(t) \end{bmatrix} = \begin{bmatrix} -(\omega_{1/2})_{\frac{n}{9}\pi} & -\Delta\omega_{\frac{n}{9}\pi} \\ \Delta\omega_{\frac{n}{9}\pi} & -(\omega_{1/2})_{\frac{n}{9}\pi} \end{bmatrix} \begin{bmatrix} V_I(t) \\ V_Q(t) \end{bmatrix} + (-1)^{n+1} K_{\frac{n}{9}\pi}(\omega_{1/2})_{\frac{n}{9}\pi} \cdot R_\mathrm{L} \begin{bmatrix} I_I(t) \\ I_Q(t) \end{bmatrix} . \quad (16)$$

This equation is used to compute the transfer function matrix of the π-mode ($n = 9$) with $K_\pi = 1$. The variation in the coupling and loaded quality factor for the remaining modes ($n \neq 9$) is described in Ref. [14] and requires an adjustment of static gain, bandwidth, and detuning by

$$K_{\frac{n}{9}\pi} \overset{n\neq 9}{=} 2\sin^2\left(\frac{n\pi}{18}\right), \quad (\omega_{1/2})_{\frac{n}{9}\pi} \overset{n\neq 9}{=} 2\sin^2\left(\frac{n\pi}{18}\right)\frac{\pi f_{\frac{n}{9}\pi}}{Q_\mathrm{L}} \quad \text{and} \quad \Delta\omega_{\frac{n}{9}\pi} = 2\pi\left(f_{\frac{n}{9}\pi} - f_\pi\right), \quad (17)$$

where Q_L is the loaded quality factor and f_π is the resonance frequency of the π-mode.

2.5 Summary

In this section, we have seen the different components acting in an LLRF feedback loop. For a feedback controller design, one can follow and identify the mathematical description of each component separately. However, setting up an optimal feedback regulation of each LLRF system, e.g., 27 LLRF feedback loops at the European XFEL, requires the mathematical description and parameter estimation of a huge amount of components. To avoid this, a system modelling for all components within one step can be performed, as described in the next section.

3 System modelling

The last section gave an overview of the main components within an LLRF feedback loop, such as the vector modulator, klystron, and cavity. However, additional components are required, e.g., to transport the signal using a cable or waveguide. Mathematical modelling using, e.g., differential equations of a complete signal chain is a huge amount of work and very complex. Furthermore, consider the European XFEL project with its 27 RF stations. Each RF station will differ in its mechanical design, e.g., waveguide distribution, but also in the mathematical description of the individual components, e.g., klystron. Here, separate modelling for 27 feedback systems is required, but very time consuming. Furthermore, ageing and, e.g., temperature changes may require this modelling to optimize the feedback controller. Another way to overcome this mathematical description of each feedback loop is system modelling using a special input signal, discussed next. Such system modelling can be done in the time or frequency domain. We will see in this section why the frequency domain is the preferable approach. Furthermore, we will assume that only linear time-invariant systems are considered. These are linear systems because they are describable by linear differential equations. Time-invariance of a system implies that for any time delay $d > 0$ the response to the delayed input $u(t-d)$ is a time-delayed output $y(t-d)$. Moreover, the coefficients of the transfer function are time-invariant and their values are constant.

3.0.1 Example RCL circuit

The electrical circuit in this example (Fig. 22) contains an inductance L, a resistance R, and a capacitance C. Applying Kirchhoff's voltage law to the system, we obtain the equations:

$$L\frac{\mathrm{d}i}{\mathrm{d}t} + Ri(t) + V_\mathrm{o}(t) = V_\mathrm{i}(t),$$
$$C\frac{\mathrm{d}V_\mathrm{o}}{\mathrm{d}t} = i(t).$$

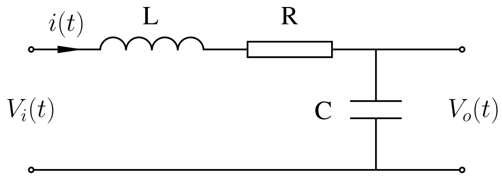


Fig. 22: Electrical RCL circuit

Eliminating the current $i(t)$ between both equations yields

$$LC\frac{\mathrm{d}^2V_\mathrm{o}}{\mathrm{d}t^2} + RC\frac{\mathrm{d}V_\mathrm{o}}{\mathrm{d}t} + V_\mathrm{o}(t) = V_\mathrm{i}(t)\,,$$

leading to

$$\ddot{V}_\mathrm{o}(t) + \frac{R}{L}\dot{V}_\mathrm{o}(t) + \frac{1}{LC}V_\mathrm{o}(t) = \frac{1}{LC}V_\mathrm{i}(t)\,.$$

We can identify $V_\mathrm{i}(t)$ with the input signal $u(t)$ and $V_\mathrm{o}(t)$ with the output signal $y(t)$.

3.1 State space model

The previous example illustrates how dynamic systems can be modelled by a second-order linear ordinary differential equation. It may be more convenient to rewrite the differential equation in a more compact form, called a state space model, when dealing with systems that are described by higher-order differential equations. Each nth-order linear differential equation model can be transformed into a first-order vector differential equation model. Let's use the previous example and introduce a second variable as $V_m(t) = \dot{V}_\mathrm{o}(t)$. Then we can write

$$\ddot{V}_\mathrm{o}(t) + \frac{R}{L}\dot{V}_\mathrm{o}(t) + \frac{1}{LC}V_\mathrm{o}(t) = \frac{1}{LC}V_\mathrm{i}(t)$$

as

$$\begin{aligned}\dot{V}_\mathrm{o}(t) &= V_m(t)\,,\\ \dot{V}_m(t) &= -\frac{R}{L}V_m(t) - \frac{1}{LC}V_\mathrm{o}(t) + \frac{1}{LC}V_\mathrm{i}(t)\,.\end{aligned}$$

Defining a state vector

$$x(t) = \begin{bmatrix} V_\mathrm{o}(t) \\ V_m(t) \end{bmatrix}\,,$$

we can express this second-order ordinary differential equation in a more compact form as

$$\begin{bmatrix} \dot{V}_\mathrm{o}(t) \\ \dot{V}_m(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R}{L} \end{bmatrix} \begin{bmatrix} V_\mathrm{o}(t) \\ V_m(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{LC} \end{bmatrix} V_\mathrm{i}(t)\,. \tag{18}$$

Furthermore, the output is

$$V_\mathrm{o}(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} V_\mathrm{o}(t) \\ V_m(t) \end{bmatrix}\,. \tag{19}$$

Equations (18) and (19) are an example of a state space model, which in general has the form

$$\dot{x}(t) = Ax(t) + Bu(t)\,, \tag{20}$$

$$y(t) = Cx(t) + Du(t)\,, \tag{21}$$

with input $u(t)$, output $y(t)$, state $x(t)$, system matrix A, input matrix B, output matrix C, and feedthrough matrix D.

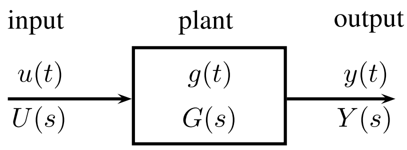


Fig. 23: System or plant $g(t)$ with input $u(t)$ and output $y(t)$; with signal notation in frequency domain using Laplace operator s.

3.2 System description in the time and frequency domains

3.2.1 Time domain

The system output $y(t)$, see Fig. 23, can be computed by a convolution of the input signal $u(t)$ with the impulse response of the system $g(t)$ as

$$y(t) = g(t) * u(t)\,. \tag{22}$$

Solving a convolution in the time domain and, by this, analysing the system behaviour is complicated, making the system analysis very complex. Therefore, system analysis is typically done in the frequency domain.

3.2.2 Frequency domain

The system output $Y(s)$, see again Fig. 23, in the frequency domain is given as a multiplication of the transformed input signal $U(s)$ and the impulse response $G(s)$:

$$Y(s) = G(s) \cdot U(s)\,, \tag{23}$$

with $s := \sigma + \mathrm{j}\omega$, as operator in the frequency domain.

3.3 Transformation into the frequency domain

The transformation into the frequency domain using *Fourier transformation* is well known and given by

$$\mathcal{F}\{f(t)\} = F(\omega) = \int_{t=-\infty}^{\infty} f(t) \cdot \mathrm{e}^{-\mathrm{j}\omega t} \mathrm{d}t\,.$$

This transformation is defined for all times, even negative times, which are not needed for system modelling. The Laplace transform, often used by system engineers, is very similar to the Fourier transform. The Fourier transform of a function is a complex function of a real variable (frequency), while the Laplace transform of a function is a complex function of a complex variable.

3.3.1 Laplace transform

Laplace transforms are usually restricted to functions of t with $t \geq 0$, with $f(t) = 0, \forall t < 0$. A consequence of this restriction is that the Laplace transform of a function is a holomorphic function of the variable $s := \sigma + \mathrm{j}\omega$. The Laplace transform from the time to the frequency domain is given as

$$\mathcal{L}\{f(t)\} = F(s) = \int_{t=0}^{\infty} f(t) \cdot \mathrm{e}^{-st} \mathrm{d}t\,. \tag{24}$$

The Laplace transform of a signal $f(t)$ as defined in Eq. (24) exists if the integral converges. In practice, the question of existence and the region of convergence are not usually issues of concern. The *inverse Laplace transform*, mapping from the frequency to the time domain is defined as:

$$f(t) = \mathcal{L}^{-1}\{F(s)\} = \frac{1}{2\pi \mathrm{j}} \lim_{T\to\infty} \int_{s=\alpha-\mathrm{j}T}^{\alpha+\mathrm{j}T} F(s) \cdot \mathrm{e}^{st}\, \mathrm{d}s\,.$$

We will not go into detail with the Laplace transform, since many properties and examples of using the Laplace transform can be found in the literature, e.g., Ref. [15].

3.3.1.1 Example cavity equation

The cavity equation introduced in Eq. (13) is given in the time domain as

$$\dot{\vec{V}} + (\omega_{1/2} - \mathrm{j}\Delta\omega)\,\vec{V} = \omega_{1/2} R_\mathrm{L}\,\vec{I}\ .$$

Mapping this equation into the frequency domain using the Laplace transform is done by solving Eq. (24) or by using the relations given in Table 3. Here, the cavity equation in the frequency domain is given as

$$\begin{aligned} s\vec{V} + (\omega_{1/2} - \mathrm{j}\Delta\omega)\,\vec{V} &= \omega_{1/2} R_\mathrm{L}\,\vec{I}\,, \\ (s + (\omega_{1/2} - \mathrm{j}\Delta\omega))\,\vec{V} &= \omega_{1/2} R_\mathrm{L}\,\vec{I}, \end{aligned}$$

where only the first derivative in time is transformed using No 5 from Table 3 and setting $f(0) = 0$. Sorting the input $(\vec{I})$ and output $(\vec{V})$ of the system leads to the transfer function $G(s)$, defined as the ratio of the output to the input signal, as:

$$\begin{aligned} \vec{V}\,(s) &= \frac{\omega_{1/2} R_\mathrm{L}}{s + (\omega_{1/2} - \mathrm{j}\Delta\omega)} \cdot \vec{I}\,(s) \\ G(s) &= \frac{\vec{V}\,(s)}{\vec{I}\,(s)} = \frac{\omega_{1/2} R_\mathrm{L}}{s + (\omega_{1/2} - \mathrm{j}\Delta\omega)}\ . \end{aligned}$$

This concept is shown in Fig. 23. If we consider the non-complex MIMO case from Eq. (14), with

$$\begin{aligned} \dot{V}_I + \omega_{1/2} V_I + \Delta\omega V_Q &= R_\mathrm{L}\omega_{1/2} I_I\,, \\ \dot{V}_Q + \omega_{1/2} V_Q - \Delta\omega V_I &= R_\mathrm{L}\omega_{1/2} I_Q\,, \end{aligned}$$

and if we assume that $\Delta\omega = 0$, i.e., there is no coupling between the I and Q channels, we get

$$\begin{aligned} \dot{V}_I + \omega_{1/2} V_I &= R_\mathrm{L}\omega_{1/2} I_I\,, \\ \dot{V}_Q + \omega_{1/2} V_Q &= R_\mathrm{L}\omega_{1/2} I_Q\,, \end{aligned}$$

leading to two decoupled first-order single-input, single-output (SISO) systems as

$$G_{II}(s) = G_{QQ}(s) = \frac{V_x}{I_x} = \frac{R_\mathrm{L}\omega_{1/2}}{s + \omega_{1/2}}\,,$$

with x as $*_{II}$ or $*_{QQ}$. Based on this transformation, we can directly find several properties of a first-order transfer function given as

$$G(s) = \frac{b_0}{s + a_0}\,,$$

such as the *static gain*, given as $K_\mathrm{P} = b_0/a_0$ for $s \to 0$, the *time constant*, as $\tau = 1/a_0$, and the *step response in time*, as $y(t) = K_\mathrm{P}(1 - \mathrm{e}^{-t/\tau})$; these properties are displayed in Fig. 24.

3.3.1.2 Time delay

A positive time delay within the system shifts the output signal by the delay. Such a time-delayed system can be described by a system free of time delay $G(s)$ and the delay. The overall system including the time delay T_d is hereby given as

$$G_\mathrm{d}(s) = G(s) \cdot \mathrm{e}^{-T_\mathrm{d} \cdot s}\,.$$

Table 3: Examples of Laplace transforms

No	Time domain $f(t)$	Frequency domain $F(s)$
1	Unit impulse $\delta(t)$	1
2	Unit step $\sigma(t)$	$\frac{1}{s}$
3	t	$\frac{1}{s^2}$
4	t^n	$\frac{n!}{s^{n+1}}$
5	$\frac{\mathrm{d}f}{\mathrm{d}t} = \dot{f}(t)$	$sF(s) - f(0)$
6	$\ddot{f}(t)$	$s^2F(s) - sf(0) - f'(0)$
7	e^{at}	$\frac{1}{s-a}; s > a$
8	$t^n\mathrm{e}^{at}$	$\frac{n!}{(s-a)^{n+1}}; s > a$
9	$\sin at$	$\frac{a}{s^2+a^2}; s > 0$
10	$\cos at$	$\frac{s}{s^2+a^2}; s > 0$
...	...	...

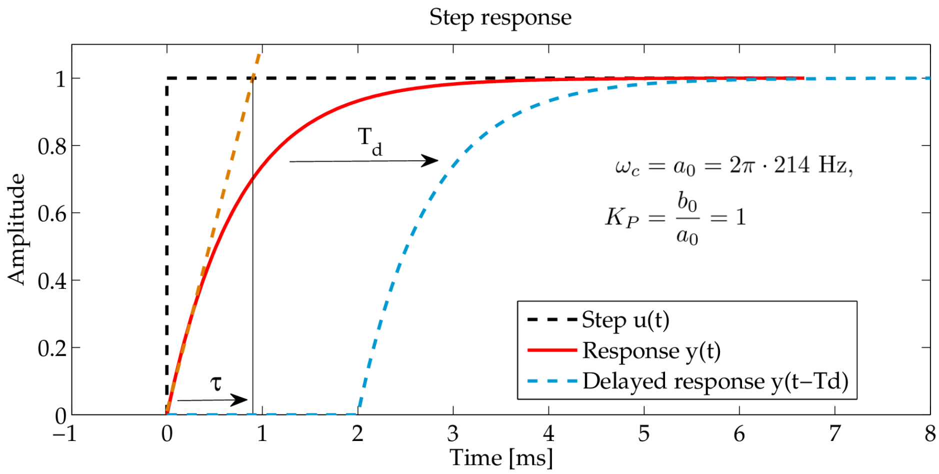


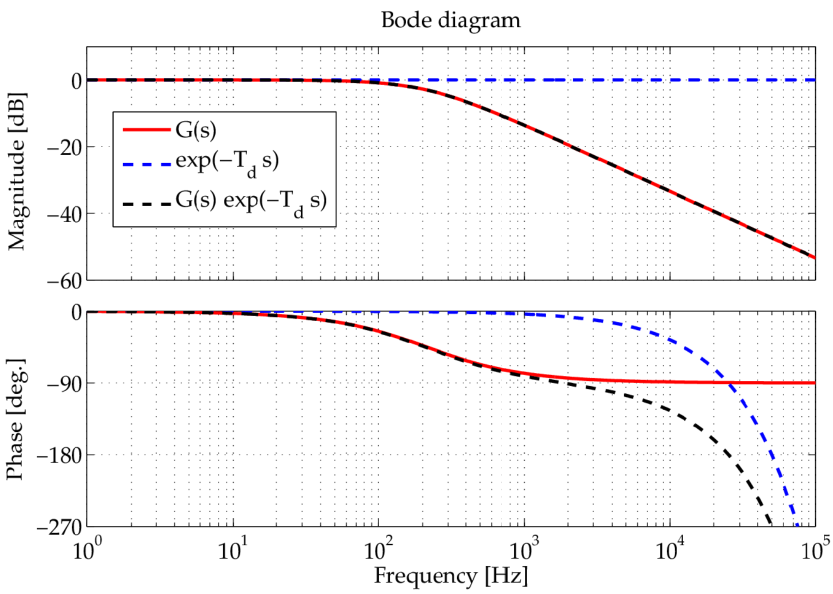
Fig. 24: Step response of first-order system: time delay, $T_\mathrm{d} = 2$ ms

3.3.1.3 Bode diagram

The information about the frequency response can be displayed graphically. The most widely used graphical representation of a frequency response is the Bode diagram. The magnitude and phase are plotted versus frequency in two separate plots, where a log scale is used for magnitude and frequency and a linear scale for the phase. The magnitude and phase are defined as

$$|G(s)|_{\mathrm{dB}} = 20\log_{10}(|G(s = \mathrm{j}\omega)|) \quad \text{and} \quad \angle(G(s)) = \arg(G(s = \mathrm{j}\omega))\,.$$

Consider a system given with a measurable input $u(t)$ and output $y(t)$. Apply to the input a sinusoidal signal with amplitude set as 1 and frequency f_i as $u(t) = \sin(2\pi f_i t)$. This leads to an output signal $y(t) = K{\cdot}\sin(2\pi f_i t{+}\varphi)$, where the amplitude and phase of the sinusoidal frequency may change. The input/output gain and phase relation between input and output signal for all frequencies f_i results in a Bode diagram. An example of a Bode diagram is given in Fig. 25. This Bode diagram shows the delay-free system $G(s)$, the pure time delay $\exp(-T_\mathrm{d}s)$, and its combination as serial connection, $G(s) \cdot \exp(-T_\mathrm{d}s)$. Here, it can be seen that the log scale is useful because the transfer function is composed of pole and zero factors that can be added graphically, i.e., adding $G(s)$ and $\exp(-T_\mathrm{d}s)$ leads to the combined transfer function. Again, the static gain ($f \to 0$) and the corner frequency (-3 dB drop at 214 Hz) can be read out. Furthermore, it is shown that a pure time delay has no effect on the magnitude given as $|\mathrm{e}^{-T_\mathrm{d}\cdot s}| = 1$, i.e., only the output is delayed without change in the gain. The phase of a pure time delay shows a roll-off



Example:

$\omega_c = a_0 = 2\pi \cdot 214$ Hz,

$K_P = \dfrac{b_0}{a_0} = 1$ and

$T_d = 10\mu s$

Fig. 25: Bode diagram of example system without and with time delay

of

$$
\begin{aligned}
\angle(\mathrm{e}^{-T_\mathrm{d} \cdot s}) = \varphi(\omega) &= -T_\mathrm{d}\omega \qquad [\mathrm{rad}] \\
&= -T_\mathrm{d}\omega \frac{180}{\pi} \; [\mathrm{deg}] \; .
\end{aligned}
$$

3.3.1.4 Example Bode diagram for cavity model

Figure 26 shows the resulting continuous-time white-box model for the cavity with all fundamental modes. Here, it is assumed that the cavity equation includes an initial detuning of $\Delta\omega_\pi = -2\pi \cdot 10$ Hz and a loaded quality factor of $Q_\mathrm{L} = 3 \times 10^6$. The Bode diagram is computed by Eqs. (16) and (17).

3.4 System connection

If not just one system but several systems are considered, it may required to connect them using a serial and parallel connection or in a feedback scheme.

3.4.1 Serial connection

The serial connection of two systems is simply the multiplication of the individual systems, shown in Fig. 27.

3.4.2 Parallel connection

The parallel connection of two systems is the sum of the systems, resulting in a combined system, Fig. 28.

3.4.3 Feedback

In contrast with a serial and parallel connection, a positive feedback loop, indicated by a summation of the feedback path from $G_2(s)$ to the input of the system $G_1(s)$, leads to a combined system with a forward path divided by one minus the loop path, i.e., in this example the serial connection of $G_1(s)$ and $G_2(s)$, Fig. 29. For a negative feedback loop, used for most of the feedback controller, the sign of the denominator needs to be changed from minus to plus.

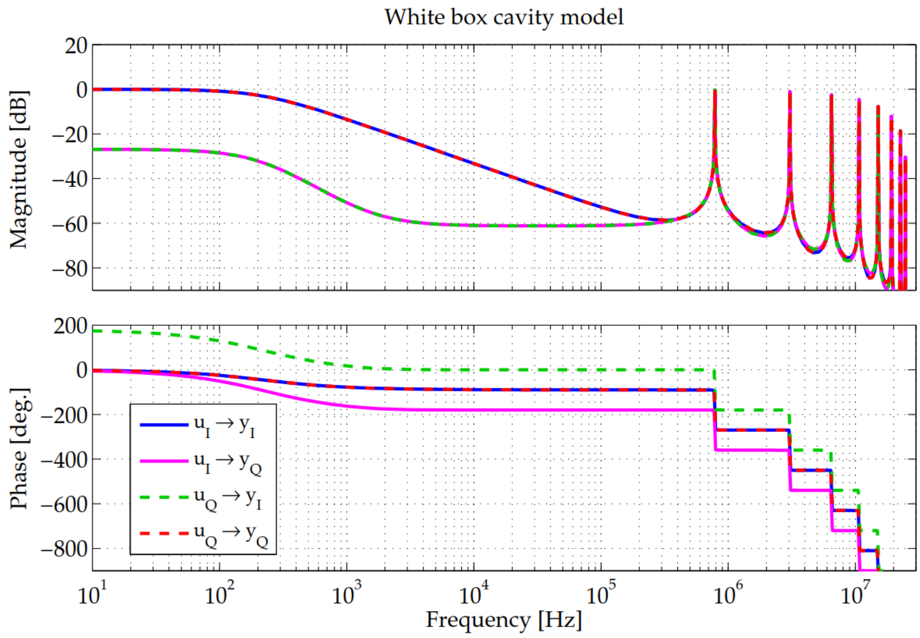


Fig. 26: Continuous-time cavity baseband white-box model without time delay. The fundamental π-mode (f_π = 1.3 GHz) is located at frequency zero. The other peaks are the eight additional fundamental cavity modes, i.e., from $8\pi/9$-mode to $1\pi/9$-mode.

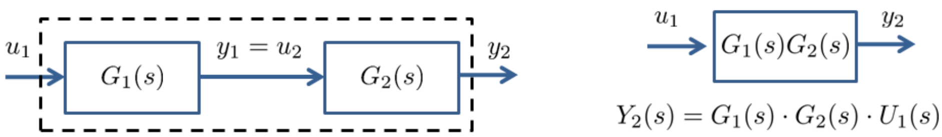


Fig. 27: Serial connection of two systems

Fig. 28: Parallel connection of two systems

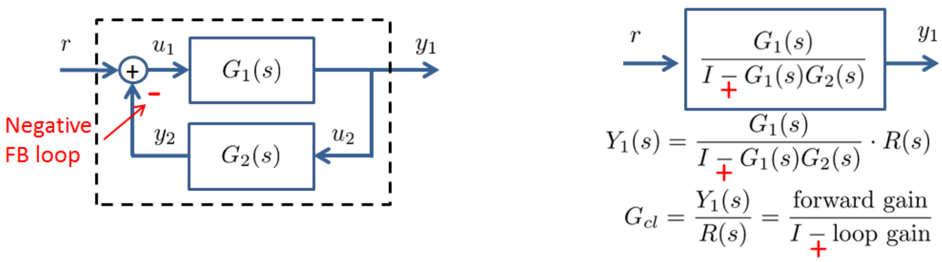


Fig. 29: Feedback of two systems

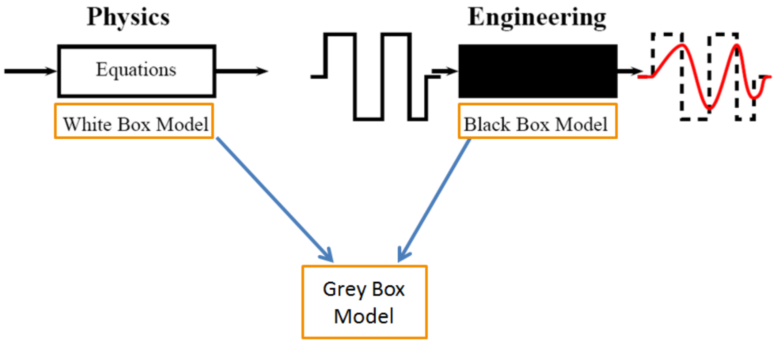


Fig. 30: Overview of system modelling

3.4.4 Summary

We have considered the system description using input/output signals of individual subsystems. The combination of different systems with different paths is shown in Section 3.4. In the next section, we will focus on the identification of a system model by exciting a system with special input signals and observing its output response.

3.5 System identification

A system model is a simplified representation or abstraction of a reality. Modelling the reality is generally too complex to produce an exact copy. Furthermore, much of the complexity is actually irrelevant in problem solving, e.g., controller design. This can be overcome to define a robust controller that can cope with model uncertainties. In general, one can describe a system using different approaches. On the one hand, a mathematical description can be created, using differential equations. This is a so-called *white-box model*, where all necessary components are described, e.g., the cavity modelling described in Section 2.4, which is already an example of the *white-box modelling* approach. On the other hand, system engineers often use so-called *black-box models*, where no assumption about the system is made, see Fig. 30. Applying special input signals, e.g., pseudo-random binary signals, and observing the output signal leads to the possibility of estimating the system impulse response, i.e., the transfer function in the frequency domain or the state space model in the time domain. A combination of white-box and black-box modelling is often used to define the main physical equations describing a system. A *grey-box model* can be identified by leaving important parameters within those main equations of the white-box model as free parameters to be identified. The main possibilities of a system identification and its different approaches to identify a system are described in Ref. [16]. An example of a grey-box modelling approach to identify a superconducting RF feedback loop can be found in Ref. [17].

In the following, a short example of a black-box modelling using MATLAB is given. For the modelling, MATLAB provides a graphical tool within the *System Identification Toolbox*™. This tool allows the transfer function, state space models, etc., to be identified only for SISO systems. Estimation of MIMO systems requires use of the command line.
Let's define a transfer function (`tf` command) of a system $G(s)$, its input signal $u(t)$, which is a step, and an output signal, which is the simulation of the input with the system using the MATLAB `lsim` command, see Fig. 31. To this output signal, 1% noise is added by a random signal generator. The MATLAB `ident` command opens a graphical interface for SISO system identification, see Fig. 32. Within this graphical tool, the following steps need to be taken.

1. Import data.
   - Choose data format and select the input and output data.
     - *Hint: An example data-set should be available.*

```
Command Window
>> s = tf('s');
>> w12 = 2*pi*214;
>> G = w12/(s+w12);
>> u = ones(1000,1);
>> y = lsim(G,u,[0:1:length(u)-1]*10e-6);
>> yn = y + 0.01*(rand(length(y),1)-0.5);
>> ident
fx >>
```

Fig. 31: MATLAB command line to define system identification

   - Adjust the sampling interval.
2. Estimate the system model.
   - Select one of the proposed methods, e.g., transfer function, state space model.
   - Define the number of poles and zeros.
   - Choose continuous or discrete time for the resulting model.
   - If necessary, set the I/O delay.
3. Compare the model (`tf1`) with the plant `G`.
   - *Hint: You can import the initial plant model* `G` *for comparison.*
4. Export the system model to the workspace for further use of the dataset, e.g., controller design.

This graphical tool gives an idea of how system identification using MATLAB works. Without going into detail, the identification requires some knowledge of the plant, to avoid the identification of unnecessary dynamics of the system using an overestimated model order. As a rule of thumb, start with a low system order and increase the model order (number of poles and zeros) until the system response fits the simulated model response sufficiently well. Another method to estimate the minimum model order is to use the state space model as the resulting model and choose the option *Pick best value in the range*. This option guessing the optimal model order based on the I/O data.

The interested reader is referred to other books and publications for detailed information of the systems identification.

4 Feedback controller design

We will start in this section with an overview of the different regulation or control schemes and their usage before defining the objective of a feedback control problem. Since each feedback scheme changes the transfer function (poles and zeros) of the closed loop, it is further necessary to introduce some checks for the stability of the resulting system.

4.1 Ways to control

There are different ways to control a plant or system. On the one hand, there is the open-loop case, Fig. 33, and on the other hand, regulation using a feedback loop, Fig. 34. If the disturbance acting on the plant is known, using only a feedforward adaptation scheme, Fig. 35, or using a feedforward control scheme for the reference and disturbance together with feedback control, Fig. 36, are further ways to minimize disturbances.

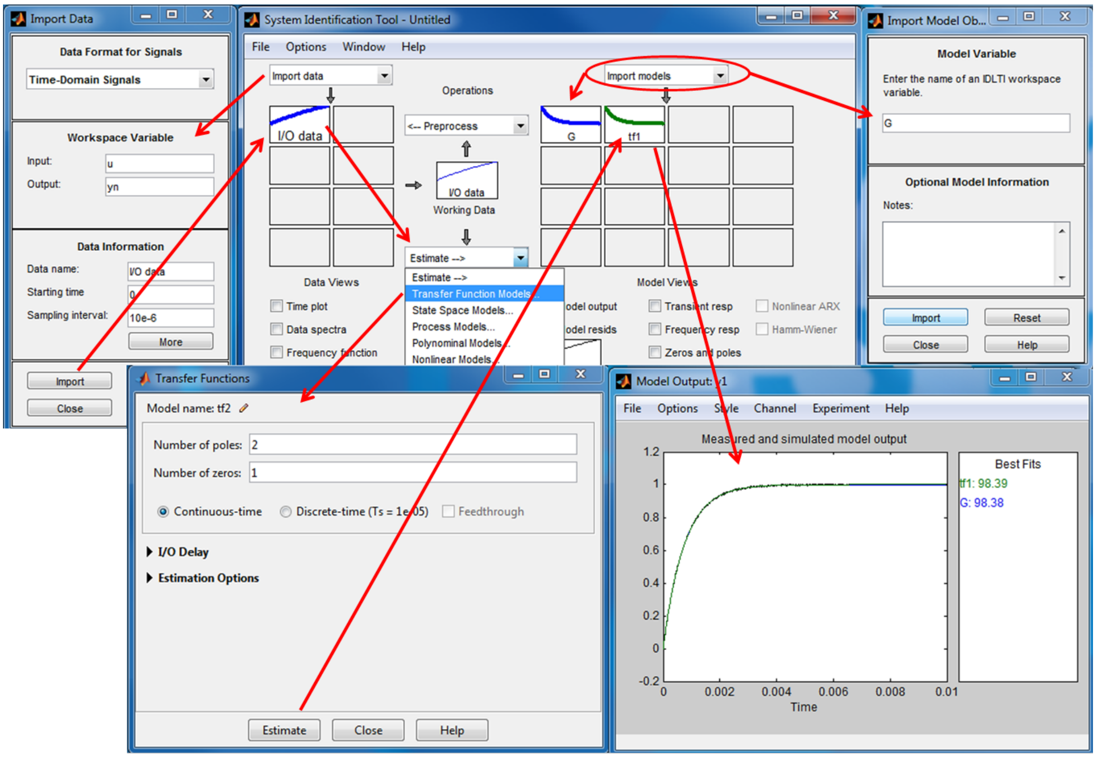


Fig. 32: System identification tool in MATLAB

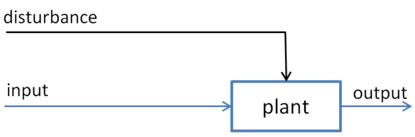


Fig. 33: Open-loop scheme

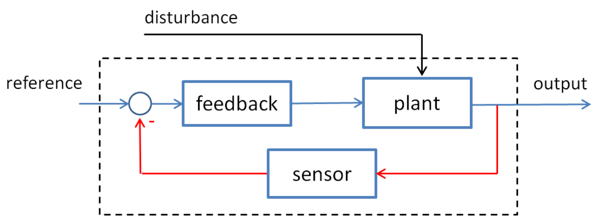


Fig. 34: Feedback scheme

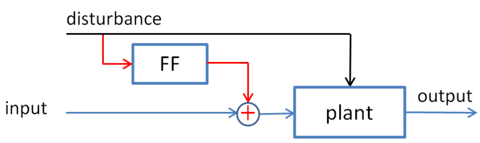


Fig. 35: Feedforward scheme

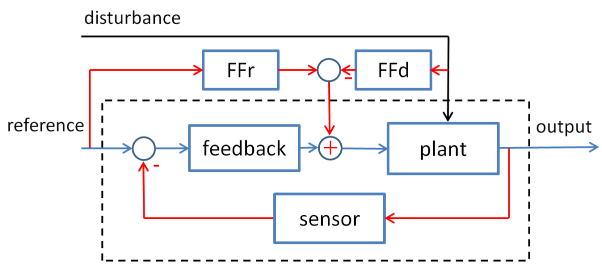


Fig. 36: Feedback scheme with feedforward compensation

4.1.1 Open loop

The input/output (I/O) behaviour of a system driven in open loop needs to be known exactly, since there is no feedback to adjust the input signal to compensate for output variations e.g., caused by disturbances. Such a scheme can be used if the output signal is not critical and may change by several per cent without losing system performance.

4.1.2 Feedforward scheme

Like the open-loop case, a precise knowledge of I/O behaviour is required when operating a system in a feedforward scheme. The disturbances must be measurable and therefore they can be compensated for by adjusting the drive signal. However, in the feedforward scheme, there is no process for automatically adjusting the output signal to be controlled.

4.1.3 Feedback scheme

The signal to be controlled (output) is fed back using a feedback regulation scheme. Therefore, the output signal acts directly on the input. In contrast with the feedforward scheme, this interconnection of signals changes the system transfer function, as seen in Section 3.4.

4.1.4 Feedback controller together with feedforward scheme

This control structure is often called a two-degrees-of-freedom controller, where two signals, i.e., the output and reference, are treated independently by introducing, in addition to a feedback controller, a feedforward controller to the reference (FFr). Furthermore, a feedforward controller for the measurable disturbance is used. Hence, this concept copes with a combination of the feedback and the two feed-forward controllers.

4.1.5 Short summary

The optimal regulation approach depends on the application for which it is used. Here, several questions must be addressed, e.g., are the signals of interest of even all signals measurable or observable? Further information on the different regulation schemes can be found, e.g., in Ref. [18] or control theory lectures from universities.

4.1.5.1 Example feedback loop for cavities

Figure 37 depicts an example of a feedback loop in LLRF controls. The plant consists in this case of a serial connection of its main components, i.e., the pre-amplifier, klystron, and cavity. The signal to be controlled is depicted as an ideal signal $y_\mathrm{i}(t)$, while, in practice, disturbances to the output $d_y(t)$ need to be added. The measurable signal $y_\mathrm{r}(r)$ is overlapped with noise $n(t)$ from the detection scheme. The feedback controller with the error signal as input, i.e., the difference between the reference $r(t)$ and the measured output signal $y_\mathrm{r}(t) + n(t)$, optimizes the ideal drive signal $u_\mathrm{i}(t)$. Again, we can assume that the drive signal to the real plant is modulated by a disturbance $d_u(t)$ leading to the real drive signal $u_\mathrm{r}(t)$, which is the input to the plant. To complete the entire feedback loop, several time delays in the detection path, the controller, and plant must be added; these are not negligible and affect the feedback operation.

4.2 Objective of a feedback control problem

The main objective of a feedback regulation scheme is to make the output $y(t)$ behave in a desired way by manipulating the plant input $u(t)$. Two different scenarios can be considered.

1. Regulator problem (output disturbance rejection with constant reference);
2. Servo problem (reference tracking without disturbance).

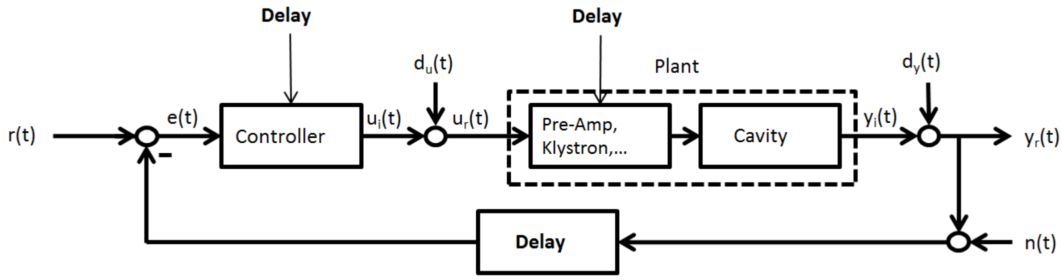


Fig. 37: Simplified block diagram of cavity feedback loop with signals and subsystems to be considered

On the one hand, in the *regulator problem*, the main goal is to suppress output disturbances. Here, it is assumed that the reference signal is constant and the feedback loop counteracts only the effect of disturbances. On the other hand, in the *servo problem*, the reference signal must be tracked optimally. In this case, it is assumed that no disturbance acts on the plant. Therefore, the goal is to keep the output close to the reference by manipulating the plant input.

Mostly, the regulator problem is addressed in practice. However, for both cases, the error signal, which is the difference between reference and output, should be minimal.

4.3 Stability

A system is stable if, for a given bounded input signal, the output signal is bounded and finite (bounded input, bounded output stable); otherwise, the system is called unstable. Stability can be checked for stable and unstable linear systems in open or closed feedback loops. For an unstable open-loop system, the main goal is to stabilize the closed-loop system behaviour using a feedback controller. The stability can be checked in the time or frequency domain.

4.3.1 Time domain

The system is stable if its impulse response $g(t)$ is absolutely integrable and bounded:

$$\int_{t=-\infty}^{\infty} |g(t)|\mathrm{d}t < \infty \,.$$

This stability check in the time domain is often not preferred and frequency domain checks are recommended.

4.3.2 Frequency domain

The stability can be checked in the frequency domain using:

1. pole location (all poles in left half plane);
2. Bode diagram;
3. Nyquist plot;
4. H-infinity norm (mostly for MIMO systems);
5. harmonic balance (for non-linear systems).

In the following, we will restrict the stability check to SISO systems, i.e., cases (1) to (3). The interested reader is referred to, e.g., Ref. [18] to learn about MIMO stability checks using the H-infinity norm. Furthermore, special methods, e.g., the harmonic balance, exist to check the stability for non-linear systems, which are beyond the scope of this paper.

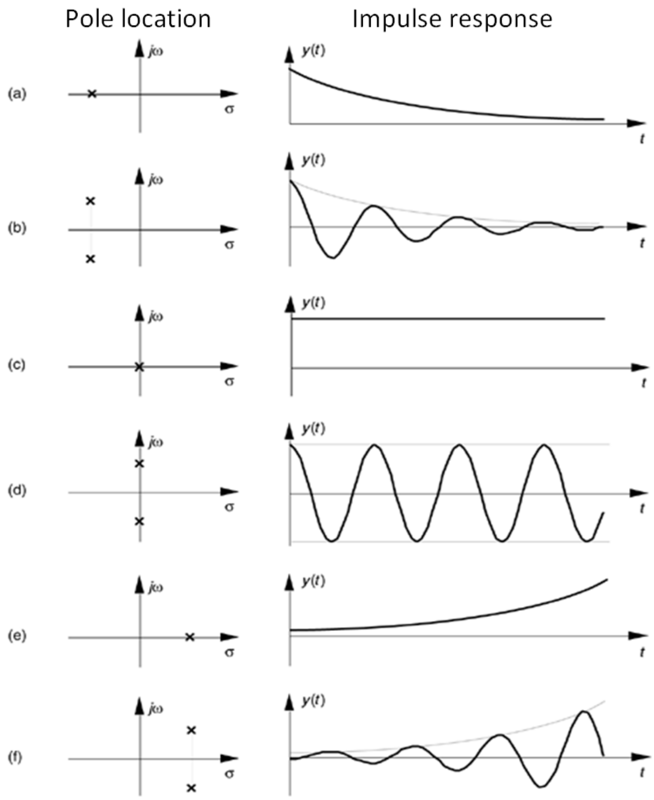


Fig. 38: Time-domain impulse response for different pole locations [19]

4.3.2.1 Pole location

Assume a SISO continuous-time transfer function of the system given as

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0} \,.$$

This system has n poles and m zeros, and if it is physically realizable, we have $n \geq m$. The values of s at which the denominator of $G(s)$ takes the value zero, and therefore at which $G(s)$ becomes infinite, are called the poles of the transfer function $G(s)$. The locations of the poles in the complex plane determine the dynamic behaviour of the system; for this reason the denominator polynomial of the transfer function is called the characteristic polynomial. To check the stability, the poles of the system must be computed and its pole location must be checked. If all poles have negative real part and therefore lie in the left half plane of the complex plane, the system is stable. A system is called semi-stable if the real part of a pole is zero. A visualization is given in Fig. 38 as an example.

4.3.2.2 Bode diagram and Nyquist plot

Another way of checking the stability of the system is by its Bode magnitude and phase plot or by its corresponding Nyquist representation. The Nyquist plot is simply the mapping from the Bode amplitude and phase plot to the complex plane; here, the frequency information is lost. The following description is a very simplified check; for further details, the reader is referred to, e.g., Ref. [18].

The system is plotted in open loop and the so-called gain margin and phase margin can be read from the graph. This gives an easy stability check for the closed-loop system using a proportional controller. Hence, both checks are made in open loop, and the closed-loop stability can be predicted.

Example: Assume a proportional controller connected serially with the system and plot the system as a Bode diagram or a Nyquist plot, as shown in Figs. 39 and 40, respectively. Read out the *phase margin*

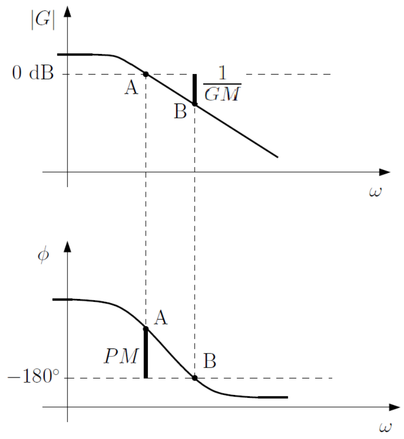


Fig. 39: Bode: gain and phase margin

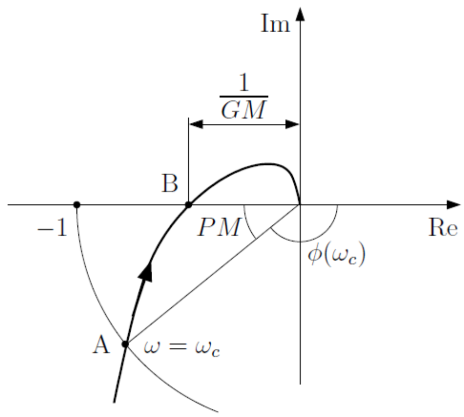


Fig. 40: Nyquist: gain and phase margin

at point A, where the magnitude crosses 0 dB and the *gain margin* at point B, where the phase crosses $-180°$. Increasing the proportional controller gain from one shifts the point A towards higher frequencies for the Bode representation. This obviously reduces the phase margin. If points A and B are equal, i.e., if the controller gain is equal to the reciprocal of the gain margin, from the system Bode plot, the closed-loop system becomes unstable, i.e., with a phase shift of $180°$, the negative feedback loop becomes a positive feedback loop. The same can be observed using the Nyquist plot. Here, the complex system line is blown up and the point B moves towards the critical point -1, i.e., with magnitude 0 dB and phase $-180°$ in the Bode plot. Furthermore, point A moves on the circle with radius 1 towards -1 causing a decrease of the phase margin.

Thus, both representations help to identify the maximum possible proportional feedback gain. Furthermore, one can easily see whether the loop is stable or not. To err on the side of caution in real applications, it is often recommended to choose a phase margin of $60°$ to cope with system variations and to keep the closed-loop operation more robust.

4.3.2.3 MIMO system stability check

Consider a MIMO system with transfer matrix $G_{\mathrm{cl}}(s)$ with input $r(t)$ and output $y(t)$ operated in a closed feedback loop, see e.g., the example given in Fig. 29. This feedback system is stable if and only if the transfer function matrix $G_{\mathrm{cl}}(s)$ is stable. Furthermore, it is mandatory that each transfer function itself within the transfer matrix is stable. Furthermore, the MIMO stability can be checked using the generalized Nyquist plot or H-infinity based methods, see, e.g., Ref. [18].

4.4 Gang of four

Often it is not sufficient to check the stability or closed-loop behaviour for only one I/O pair. It is highly recommended to check it for the so-called *gang of four*. An example of a feedback loop is depicted in Fig. 41. Here, the system has one output ($y(t)$), three inputs ($r(t)$, $d(t)$, $n(t)$), and intermediate signals (e.g., $e(t)$ and $u(t)$). First check the response of $y(t)$ to disturbance $d(t)$ and the response of $u(t)$ to measurement noise $n(t)$ by

$$G_{yd}(s) = \frac{G(s)}{1+G(s)C(s)} \quad \text{and} \quad G_{un}(s) = -\frac{C(s)}{1+G(s)C(s)}\,.$$

This is followed by a check of the robustness to process variations by

$$S(s) = \frac{1}{1+G(s)C(s)} \quad \text{and} \quad T(s) = \frac{G(s)C(s)}{1+G(s)C(s)}\,,$$

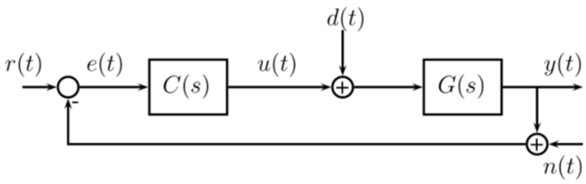


Fig. 41: Closed loop with feedback controller $C(s)$ and plant $G(s)$ operated in negative feedback loop. The main signals are given: reference $r(t)$, output $y(t)$, error $e(t)$, controller output $u(t)$, disturbance $d(t)$, and noise $n(t)$.

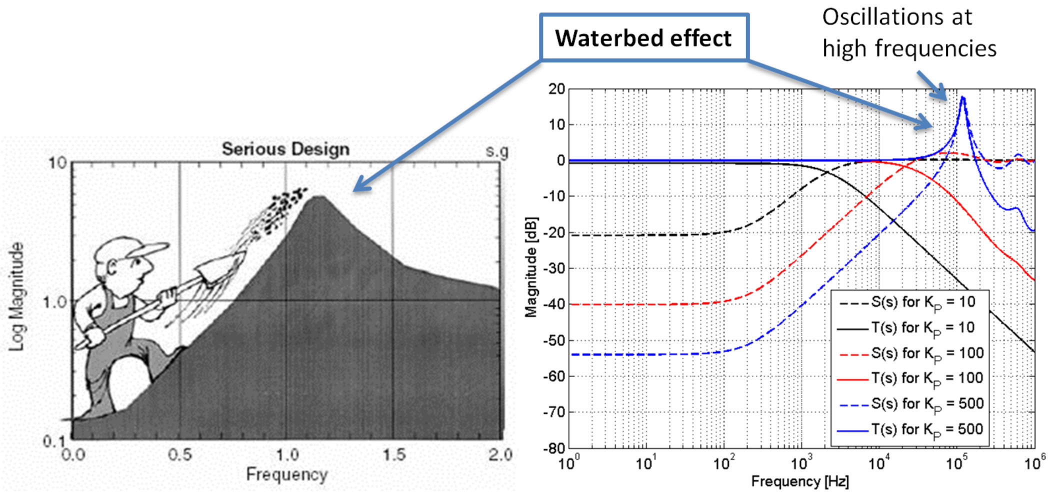


Fig. 42: Increase of feedback gain and the *waterbed effect* [20]

where $S(s)$ is the so-called sensitivity function and $T(s)$ is the so-called complementary sensitivity function. Both transfer functions depend on the loop transfer function $L(s) = G(s) \cdot C(s)$. The coupling between the sensitivity and complementary sensitivity function is given for the SISO case as $S(s) + T(s) = 1$. In a typical feedback operation $S(s)$, for frequencies towards zero should be small ($S(0) \rightarrow 0$) and $S(\infty) \rightarrow 1$, i.e., often referred to as disturbance rejection, and $T(0) \rightarrow 1$ and $T(\infty) \rightarrow 0$, i.e., referred to as reference tracking. By changing the controller transfer function $C(s)$, all four transfer functions to be checked for stability and its behaviour are changing. The effect of changing the controller gain is often visualized by $S(s)$ and $T(s)$; an example is given in Fig. 42. Pushing down the sensitivity function $S(s)$ for low frequencies pushes up $S(s)$ for higher frequencies caused by the so-called waterbed effect. In this way, high-frequency oscillations occur if the controller gain is increased. Hence, the optimal feedback strategy is often not reaching high feedback gain. The optimal controller is often a trade-off between reference tracking and disturbance rejection.

Hint: By using, in addition, a filter function for the reference, the so-called *gang of six* is formed, so, in total, six transfer functions need to be checked.

4.5 Types of feedback controller design

A feedback controller can be designed in two different ways, i.e., in the frequency or time domain. The *classical feedback controller design* is normally implemented in the frequency domain for a SISO controller design using loop-shaping techniques. Here, the open-loop controller and system analysis is performed using e.g., Bode or Nyquist plots, see Section 4.3.2. A simple example is the PID controller with transfer function

$$\frac{U(s)}{E(s)} = C(s)$$

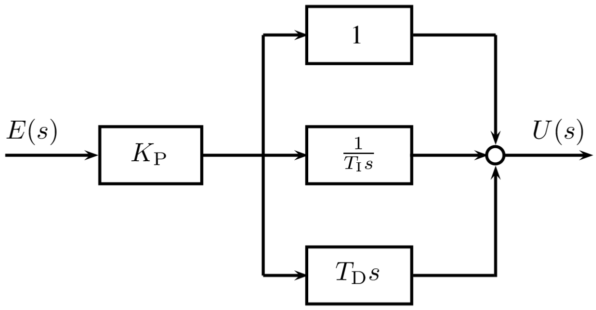


Fig. 43: PID controller

and the error signal as input, i.e., the difference between the set-point and the signal to be controlled, and the output signal, i.e., the signal driving the system or plant. Its time and frequency response is given as

$$\begin{aligned} u(t) &= K_{\mathrm{P}}\left(e(t) + \frac{1}{T_{\mathrm{I}}}\int_{t_0}^{t} e(\tau)\mathrm{d}\tau + T_{\mathrm{D}}\frac{\mathrm{d}e(t)}{\mathrm{d}t}\right) , \\ &= K_{\mathrm{P}}e(t) + K_{\mathrm{I}}\int_{t_0}^{t} e(\tau)\mathrm{d}\tau + K_{\mathrm{D}}\dot{e}(t) , \\ U(s) &= K_{\mathrm{P}}E(s) + \frac{1}{s}K_{\mathrm{I}}E(s) + sK_{\mathrm{D}}E(s) , \\ &= K_{\mathrm{P}}\left[1 + \frac{1}{T_{\mathrm{I}}s} + T_{\mathrm{D}}s\right]E(s) , \end{aligned}$$

with a block diagram for the frequency domain depicted in Fig. 43. The proportional part accounts for present values of the error, i.e., for a positive proportional gain, and if the error is positive, the control output will also be positive. The integral part accounts for past values of the error. Therefore, the integral of the error will accumulate over time and the controller will respond by applying a stronger action. The derivative part accounts for possible future trends of the error, based on its current rate of change.

A PID controller relies only on the measured process variable, not on knowledge of the underlying process. The use of the PID algorithm does not guarantee optimal control of the system, or even its stability. Normally PI controllers are in use, since the derivative action is sensitive to measurement noise (for digital signal processing this involves analogue-to-digital converter noise), whereas the absence of an integral term may prevent the system from reaching its target value. Here, the goal is to tune two parameters (K_{P} and K_{I}) to achieve a kind of optimal closed-loop behaviour. The tuning of a PID controller in general can be implemented in different ways, e.g., loop-shaping techniques, manual tuning by observing the closed-loop behaviour, or the Ziegler–Nichols tuning rules. The interested reader is referred to the literature for more detail on such tuning rules.

In contrast with classical feedback control, the *modern feedback controller design* is usually implemented as time-domain analysis, i.e., the system is presented as a state space model, given as

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) , \\ y(t) &= Cx(t) + Du(t) . \end{aligned}$$

Such a system representation and, therefore, the modern feedback controller design is not restricted to SISO systems. Hence, it is more general for MIMO systems and often based on signals directly, e.g., the linear quadratic regulator, or closed-loop shaping methods, e.g., considering the H-infinity norm of the weighted sensitivity function. The *linear quadratic regulator control* approach assumes that the plant dynamics are linear and known. The quadratic cost function is given as

$$J = \int_0^{\infty} \left[x(t)^{\mathrm{T}}Qx(t) + u(t)^{\mathrm{T}}Ru(t)\right] \mathrm{d}t , \tag{25}$$

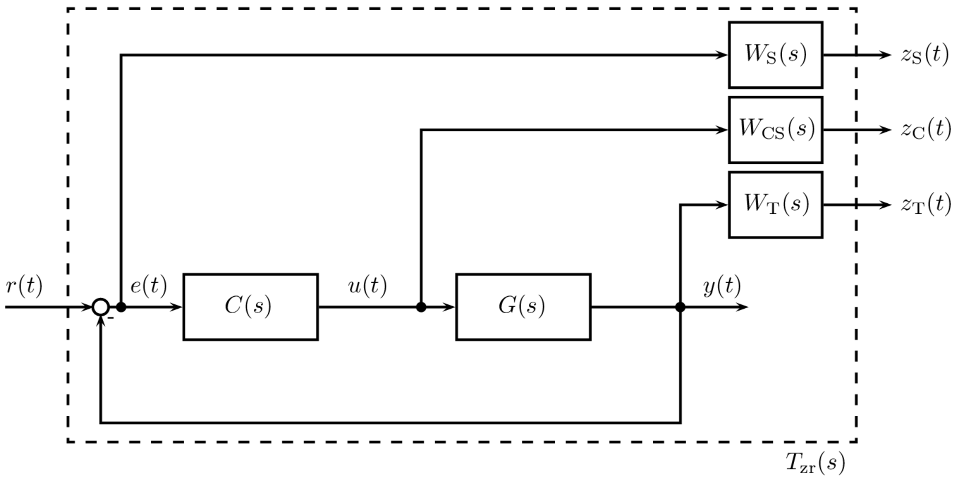


Fig. 44: H_∞ design problem with shaping filter $W_\mathrm{S}(s)$, $W_\mathrm{CS}(s)$, and $W_\mathrm{T}(s)$

with weighted state $x(t)$ and weighted controller output $u(t)$; the constant weighting matrices Q and R are the design parameters. The optimal solution to this problem is

$$u(t) = -K \cdot x(t)\,,$$

with $K = R^{-1}B^\mathrm{T}X$ and X as the unique positive semi-definite solution of the Riccati equation

$$A^\mathrm{T}X + XA - XBR^{-1}B^\mathrm{T}X + Q = 0\,.$$

Often not all states $x(t)$ are directly measurable, a prerequisite for linear quadratic regulator state feedback. This requires a Kalman filter as state estimator and its regulation on the estimated state. The reader is referred to Ref. [18].

In contrast with state feedback using a linear quadratic regulator, direct feedback to the measurable output signal $y(t)$ can be designed, e.g., using *H-infinity* (H_∞) based optimization methods and weighting filter functions to shape the sensitivity and complementary sensitivity function, see Section 4.4. A block diagram is given in Fig. 44. Here, the goal is to find the optimal controller minimizing the H_∞-norm

$$\|T_\mathrm{zr}(s)\|_\infty = \left\| \begin{bmatrix} W_\mathrm{S}(s) \cdot S(s) \\ W_\mathrm{CS}(s) \cdot C(s)S(s) \\ W_\mathrm{T}(s) \cdot T(s) \end{bmatrix} \right\|_\infty < 1 \tag{26}$$

from the reference $r(t)$ to the fictitious output signals $z(t)$, which are connected by shaping filters to different signals within the feedback loop. Keeping the H_∞-norm below 1 one guarantees a stable closed-loop system, see the small-gain theorem. A complete description of the H_∞-norm optimization would be beyond the scope of this paper and the interested reader is referred to Ref. [18].

5 Examples

Next, some examples for feedback controller design at different facilities are given. First let's start with the feedback controller design for an RF field regulation, where the system is operated in pulsed mode. The transition from pulsed to continuous wave mode often requires an increase in Q_L, putting additional effort in microphonics suppression or rejection acting to the cavities, described in the second example. If additional disturbances are measurable, a feedback scheme based on known disturbance is recommended, discussed in the last example.

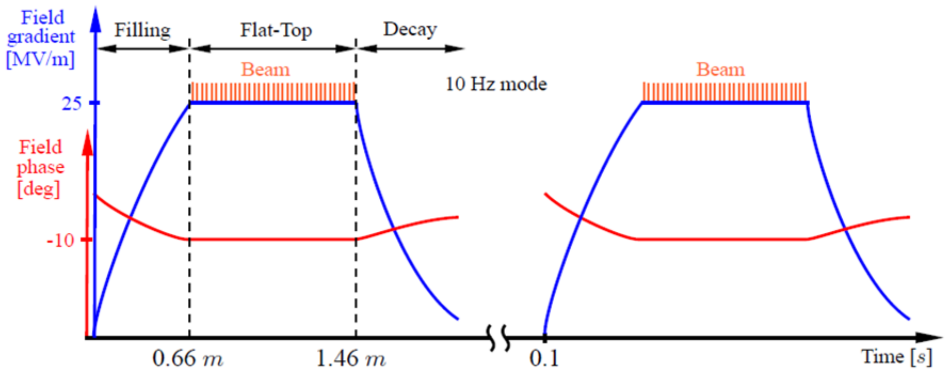


Fig. 45: Pulsed mode at FLASH with filling, flattop, and decay. The electron bunches are injected during the flattop time, where the amplitude and phase of the RF field needs to be kept constant.

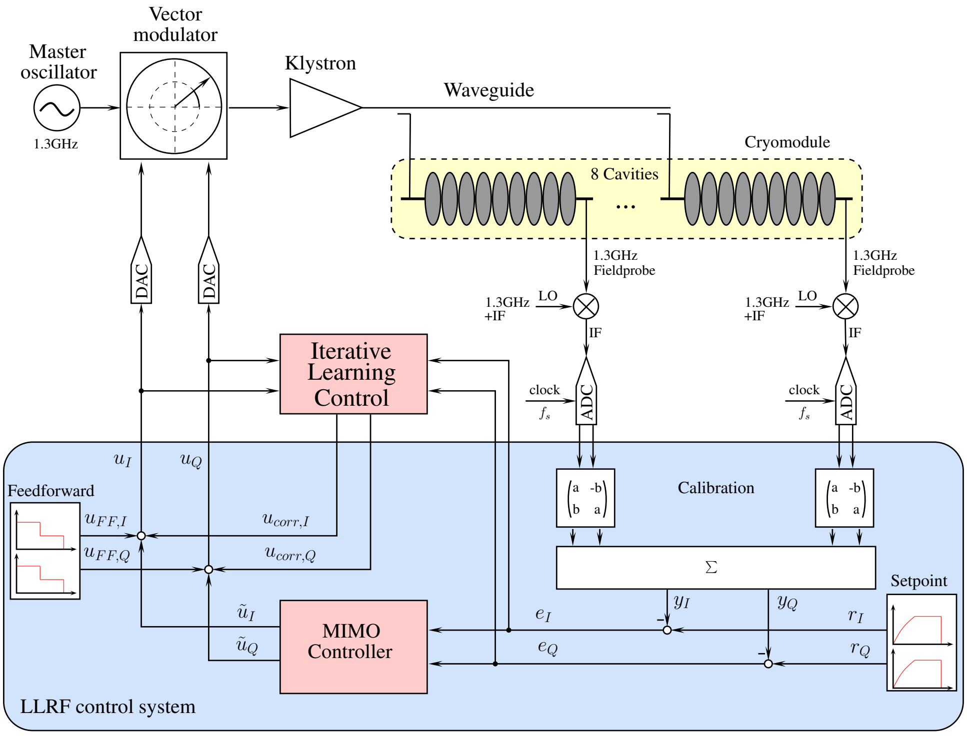


Fig. 46: LLRF system overview with plant and LLRF control system

5.1 RF field feedback loop

This example is performed at the FLASH based on the description in Ref. [17]. FLASH is operated in a pulsed mode with 10 Hz repetition rate and a pulse length of about 1 ms, see Fig. 45.

Each RF pulse can be divided into three parts: the filling; the flattop, where the electron bunches are accelerated; and the decay. The regulation goal is to achieve a stability of $\Delta A/A \leq 0.01\%$ and $\Delta\varphi \leq 0.01°$. Two main concepts are used; on the one hand, adaptation by learning from pulse to pulse, i.e., the optimization of the drive signal at 10 Hz, and, on the other hand, the feedback controller acting within an RF pulse, i.e., at 9 MHz sampling frequency. The system is controlled in the in-phase (I) and quadrature (Q) plane. Hence, the signals for regulation are mapped from the amplitude and phase representation. With this, the system has two inputs and two outputs, i.e., it is a MIMO system (Fig. 46).

The feedback controller is designed using H_∞-norm optimization. To do this, a system model is required. This is created by adding, in addition to the nominal drive signal (the input to the system u),

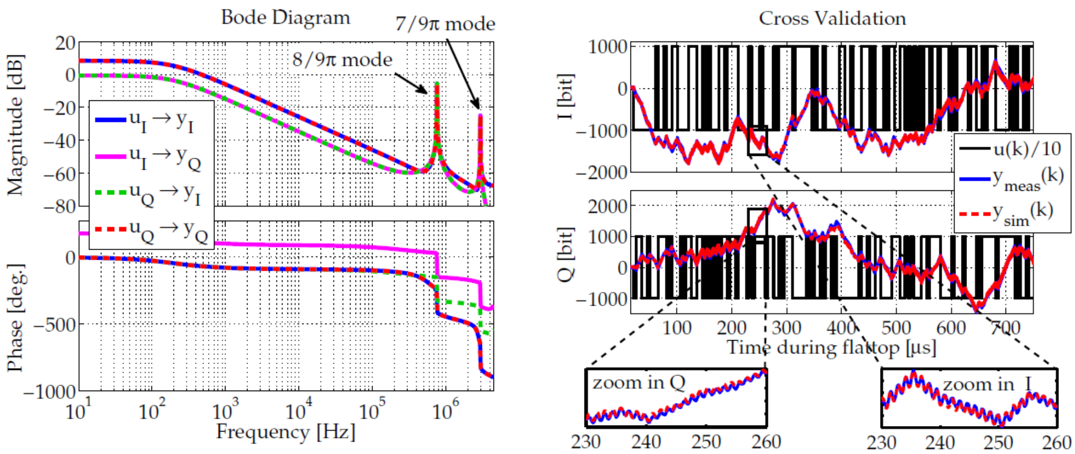


Fig. 47: Bode diagram of model and cross-validation, confirming system modelling

a small excitation signal. The change in output and its corresponding input signal are used to estimate a small signal model around the nominal operating point. In this way, the transfer matrix is estimated; its transfer functions (four in total for a 2×2 MIMO system) are given as a Bode diagram in Fig. 47. Furthermore, a cross-validation (the new excitation signal differs from the signal used for the system identification) shows the accuracy of the system identification.

With this model, the optimal MIMO controller coefficients based on an H_∞ approach are computed and follow the introduction in Section 4.5. Here, the weighting filters for closed-loop shaping are used to define the resulting closed-loop behaviour. The filter function for the controller sensitivity $W_{\mathrm{CS}}(s)$ is chosen such that the controller must suppress the $7\pi/9$ resonance mode at 3 MHz; a notch filter at the analogue-to-digital converter for the $8\pi/9$ resonance mode is set such that this mode is not present in the error signal. The optimization is performed and leads to a MIMO controller with the control objectives of plant decoupling (no coupling from channel I to Q and vice versa), avoidance of passband mode excitation, and optimal closed-loop behaviour.

The idea of iterative learning control [21] for pulse-to-pulse feedforward adaptation is described in Ref. [22]. It is similar to the MIMO controller design for a model-based regulation scheme. The time-dependence within an RF pulse is given by $t = 0, \ldots, T$, and the indices k and $k+1$ represent the previous and current RF pulses, respectively. Given is a discrete linear time-invariant state space model,

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t)\,, \qquad x(0) = x_0\,, \qquad 0 \le t \le T\,, \\ y(t) &= Cx(t)\,, \end{aligned}$$

of the closed-loop system $G(z)$, with the discrete time shift operator z. Mapping from continuous ($G_c = G(s)$) to discrete time ($G_d = G(z)$) can be done in MATLAB, using the command `Gd = c2d(Gc,Ts,'tustin')`; T_{s} is the sampling time in seconds and 'tustin' the bilinear discretization method. Here, the closed-loop transfer function is given with the error as input ($u(t) \to e(t)$) and the controller output as output ($y(t) \to u(t)$), since the LLRF system is operating with MIMO feedback and the iterative learning control is acting on $u(t)$ with the signal to be minimized from pulse to pulse $e(t)$, see Fig. 41. The iterative learning control performance index is given by

$$J_{k+1}(u_{k+1}) = \frac{1}{2}\left(\sum_{t=1}^{T} e_{k+1}(t)^T Q(t) e_{k+1}(t) + \sum_{t=0}^{T-1} \left(u_{k+1}(t) - u_k(t)\right)^T R(t) \left(u_{k+1}(t) - u_k(t)\right)\right),$$

where the weighting matrices $Q(t)$ and $R(t)$ are of compatible dimensions. This is the familiar performance criterion from linear quadratic optimal control theory. The optimal correction signal for the next RF

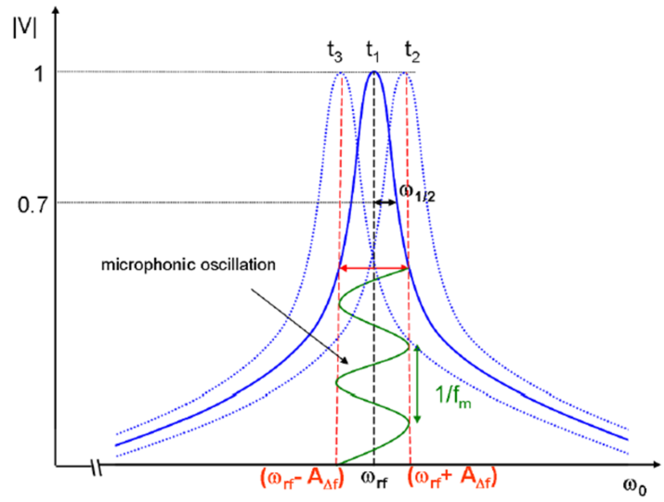


Fig. 48: Resonance frequency change by microphonics (courtesy: C. Schmidt)

pulse $u_{k+1}(t)$ is computed by

$$\xi_{k+1}(t) = \beta(t)\xi_{k+1}(t+1) + \gamma(t)e_k(t+1),$$
$$u_{k+1}(t) = u_k(t) + \omega\xi_{k+1}(t),$$

where the required matrices $\beta(t)$, $\gamma(t)$, $\lambda(t)$, and ω are calculated by solving a Riccati equation backwards with the corresponding closed-loop model. For further information see, e.g., Refs. [17,21,22].

Using both feedback schemes, the MIMO controller, together with iterative learning control, the regulation goal in amplitude and phase stability are met and typically below the specification, i.e., $\Delta A/A \approx 0.008\%$ and $\Delta\varphi \approx 0.008°$.

5.2 Microphonics suppression

Microphonics is one of the main disturbances acting to a cavity leading to small deformations of the cavity body. Such small deformations lead to an amplitude and phase error of a driven cavity caused by a detuning of the cavity from its resonance frequency, see Fig. 48.

The effect is mainly dominant for high Q_L ($Q_\mathrm{L} > 10^7$) cavities used in continuous wave operation. However, even at FLASH, microphonics was observed during the construction of the European XFEL. FLASH is typically operated with $Q_\mathrm{L} = 3\times10^6$, where microphonics is not the main disturbance source. This changes, by using e.g., a compacting machine at the XFEL injector (distance to FLASH is approx. 400 m) and its strong ground shaking, see Fig. 49. Here, the system performance gets worse (up to a factor of four during the work). Hence, a decoupling from the ground or all external noise sources is often the first approach. This could be confirmed for a continuous wave driven cavity by placing an anti-vibration mat below a vacuum pump, see Fig. 50. Such *passive noise reduction* helps improve the system performance. However, not all disturbance and noise sources can be avoided.

An *active microphonics reduction* for high Q_L ($Q_\mathrm{L} > 10^7$) superconducting RF cavities was studied [23]. In this thesis, the main disturbances acting on a superconducting RF cavity were studied, see Fig. 51.

Furthermore, a system model is estimated, which consists of different mechanical cavity modes:

$$\Delta\ddot{\omega}_{\mathrm{cav},k}(t) + 2\xi_k\omega_{m,k}\cdot\Delta\dot{\omega}_{\mathrm{cav},k}(t) + \omega_{m,k}^2\cdot\Delta\omega_{\mathrm{cav},k}(t) = \pm k_{\mathrm{LF},k}2\pi\omega_{m,k}^2 V_{\mathrm{Piezo}}(t)\,,$$
$$\Delta\omega_{\mathrm{cav}}(t) = \sum_k \Delta\omega_{\mathrm{cav},k}(t)\,,$$

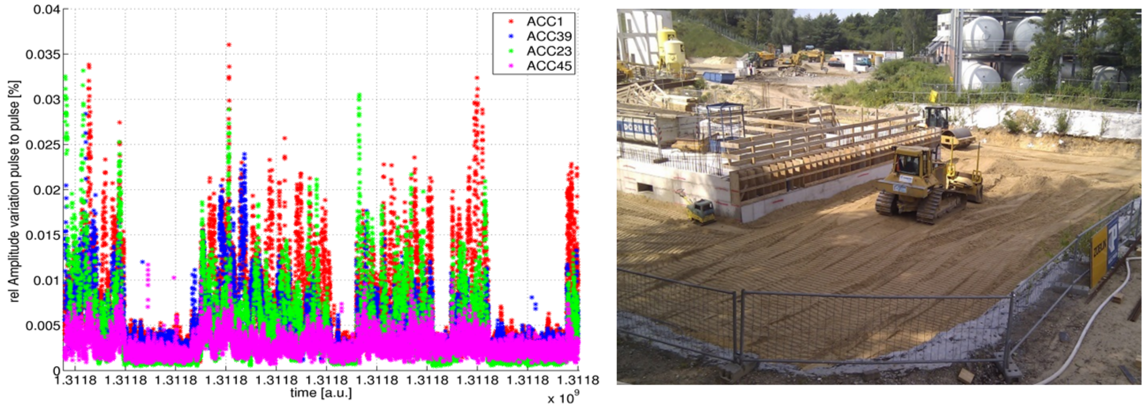


Fig. 49: Left-hand side, amplitude stability at the FLASH superconducting RF modules; right-hand side, compacting machine at XFEL injector (noise source).

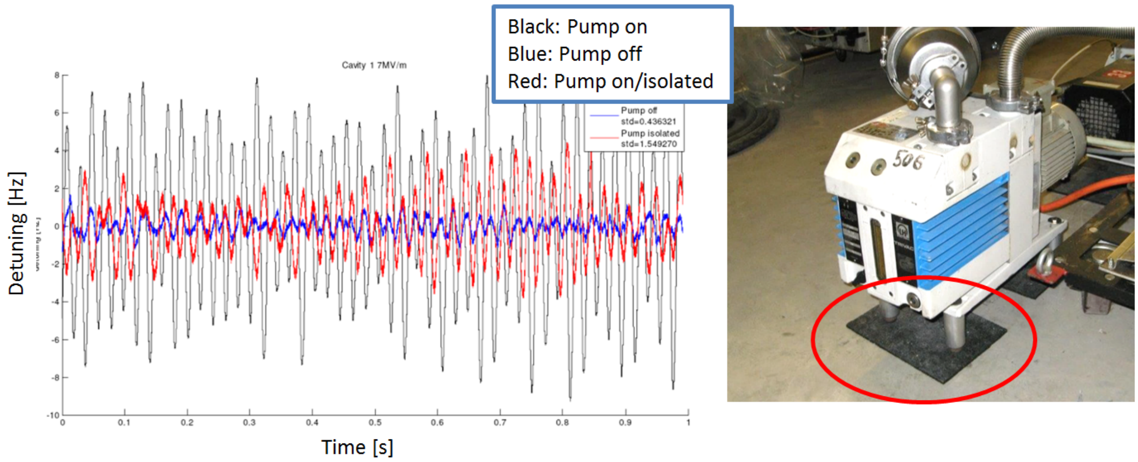


Fig. 50: Left-hand side, superconducting RF cavity detuning (courtesy: R. Rybaniec); right-hand side, vacuum pump as noise source (courtesy: J. Eschke).

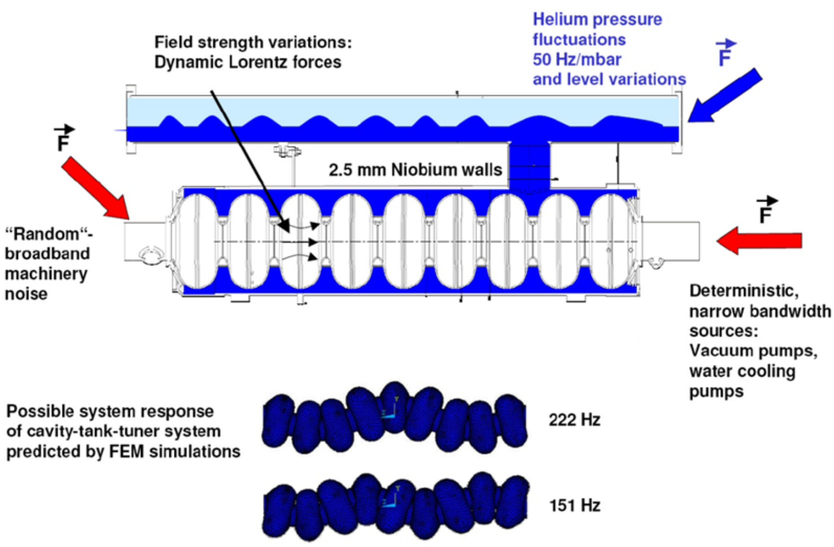


Fig. 51: TESLA cavity welded in cryo-unit; possible detuning sources are depicted [23]

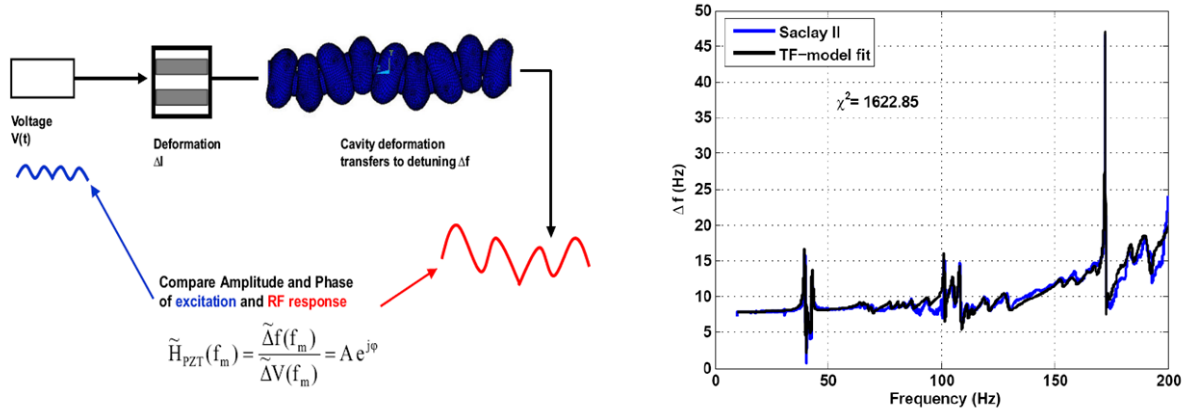


Fig. 52: Left-hand side, principle of transfer function measurement; right-hand side, model transfer function in frequency range of 10–200 Hz [23].

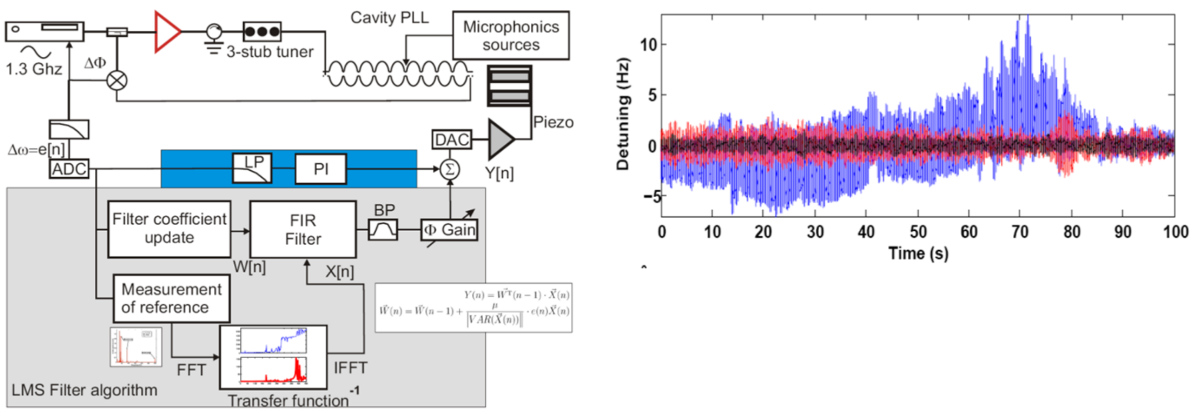


Fig. 53: Left-hand side: detuning compensation scheme given by a feedback and an adaptive feedforward signal path. Right-hand side: time-domain detuning data in open loop (blue), PI controller (red), and combined feedback and feedforward control (black).

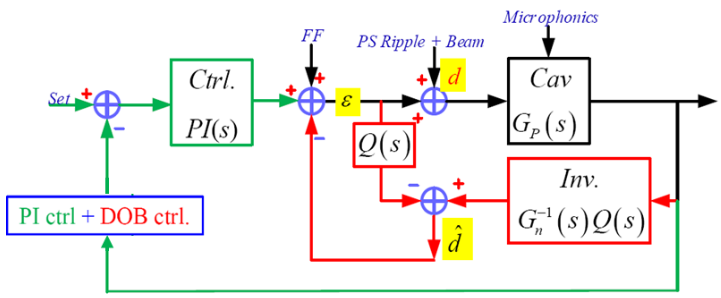


Fig. 54: Disturbance observer-based control scheme in combination with PI feedback loop [26]

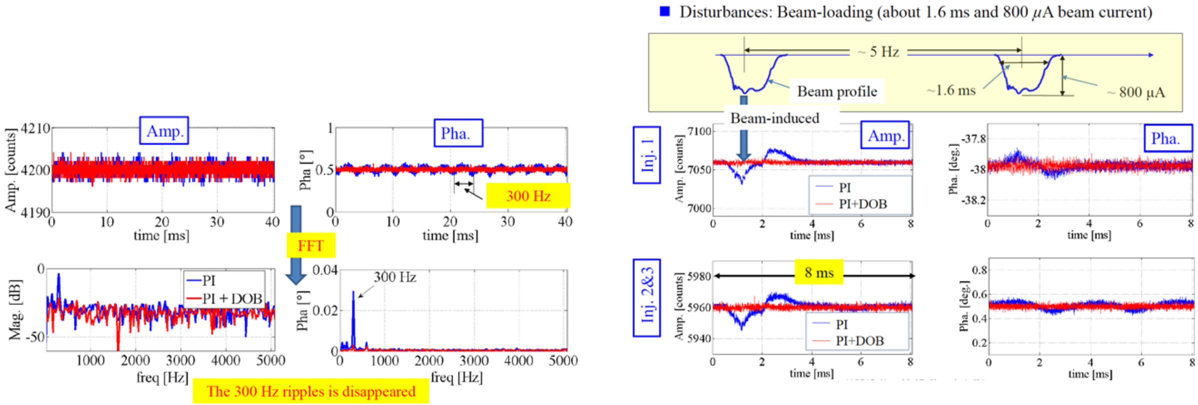


Fig. 55: Left-hand side, power supply ripple reduction using disturbance observer-based control; right-hand side, beam-induced disturbance rejection [26].

with the applied piezo voltage V_{piezo} as input and the corresponding eigenmode detuning $\Delta\omega_{\text{cav},k}$ as output. Thus, the model for each mode k in frequency domain using Laplace transform is given as

$$H_k(s) = \frac{\omega_k^2 M_k}{s^2 + 2\xi_k \omega_k s + \omega_k^2} \,.$$

Together with a low-pass behaviour and a time delay, the total system model is

$$H(s) = \left(H_0(s) + \sum_{k=1}^{N} H_k(s) \right) \cdot H_{\text{delay}}(s)\,,$$

with

$$H_0(s) = \frac{M_0}{\tau s + 1} \text{ (low-pass)}$$

and

$$H_{\text{delay}}(s) = \exp(-T_{\text{delay}} \cdot s) \text{ (time delay)}\,.$$

A model is identified for the Saclay II tuner and its measured and estimated transfer function is shown in Fig. 52. Based on this model a controller consisting of a feedforward and feedback scheme is designed, as shown in the left panel of Fig. 53. The resulting detuning stability for a superconducting RF cavity operated at 1.8 K with $Q_{\text{L}} = 6.4 \times 10^7$ is shown on the right panel of Fig. 53. For further information, see Refs. [23] and [24].

5.3 Disturbance rejection

Reference [25] describes the application using a disturbance observer-based control scheme for LLRF systems. It is in use at the compact energy recovery linac at KEK. In the following, a brief outline of this topic is given. Assume, a disturbance acting on the system, and that, for this example, the disturbance is not precisely measurable. However, a disturbance observer can be used to estimate the disturbance. This requires a sufficiently precise system model $G_P(s)$. The system output is filtered by the inverse system transfer function $G_n^{-1}(s)$ together with a filter function $Q(s)$, see Fig. 54. This filter function is needed to low-pass filter high-frequency components, to set the frequency region of interest for the expected occurring disturbance and to cope with the inversion of causal systems. Furthermore, it is necessary to filter the system input (known from the drive signal) by the same filter function. The difference between the filtered drive signal and the estimated drive signal computed from the output signal is an estimate of the disturbance. This estimated disturbance signal is subtracted from the drive, i.e., in the best case, the subtracted estimated disturbance and the real disturbance cancel.

Beside this feedforward disturbance cancellation, a PI feedback loop is added as an outer loop. This further reduces the occurrence of disturbances that are not compensated by the disturbance observer-based control scheme. Two examples are given in Fig. 55: the suppression of power supply ripples and the beam-induced voltage decrease. Both examples show that using advanced control schemes helps to improve the system performance. For further details, see Ref. [25].

Acknowledgements

I warmly thank J. Branlard, M. Hoffmann, G. Lichtenberg, H. Schlarb, and C. Schmidt for their support and helpful comments during the preparation of this lecture and for proofreading this contribution. I wish to acknowledge in particular, for the material provided: J. Eschke, M. Hoffmann, T. Miura, A. Neumann, M. Omet, F. Qiu, R. Rybaniec, T. Schilcher, and C. Schmidt.

References

[1] W. Anders *et al.*, Status report of the Berlin Energy Recovery Linac Project bERLinPro, in IPAC2016, Busan, Korea, 2016, p. 1827.

[2] P. Echevarria *et al.*, First LLRF tests of bERLinPro gun cavity prototype, in IPAC 2016, Busan, Korea, 2016, p. 1831.

[3] *bERLinPro—SC RF Accelerating Structures*, 2016. https://www.helmholtz-berlin.de/projects/berlinpro/bpro-groups/srf_en.html

[4] N. Nakamura *et al.*, Present status of the compact ERL at KEK, in Proc. 5th International Particle Accelerator Conf. (IPAC 2014): Dresden, Germany, 2014, p. MOPRO110. http://jacow.org/IPAC2014/papers/mopro110.pdf

[5] T. Miura *et al.*, Low-level RF system for cERL, IPAC 2010, Kyoto, Japan, 2010, p. 1440.

[6] *IMS—Image Archive*, 2016. https://www2.kek.jp/imss/eng/library/

[7] *Free-Electron Laser FLASH*, 2016. http://flash.desy.de/

[8] S. Di Mitri, Bunch length compressors, *CAS: FELs and ERLs*, Hamburg, Germany, 2016.

[9] M. Hoffmann, Ph.D. thesis, Hamburg-Harburg University of Technology, 2008. http://www-library.desy.de/cgi-bin/showprep.pl?desy-thesis-08-028

[10] Z. Geng and S. Simrock, Evaluation of fast ADCs for direct sampling RF field detection for the European XFEL and ILC, *LINAC*, Victoria, BC, Canada, 2008, p. 1030.

[11] T. Schilcher, Ph.D. thesis, Hamburg University, 1998.

[12] M. Omet *et al.*, Development and test of klystron linearization packages for FPGA-based low level RF control systems of ILC-like electron accelerators, 2014 19th IEEE-NPSS Real Time Conf. (RT), 260030, 2014 (IEEE, Piscataway, NJ, 2014), p. 1.

[13] *9-cell TESLA-Type SRF-Cavity*, 2016. http://tt.desy.de/desy_technologies/accelerators_magnets_und_cryogenic_technologies/weld_free_cavity/index_eng.html

[14] E. Vogel, *Phys. Rev. Spec. Top. Accel. Beams* **10** (2007) 052001.

[15] *Laplace Transform*, 2016. https://en.wikipedia.org/wiki/Laplace_transform

[16] L. Ljung, *System Identification, Theory for the User*, 2nd ed. (Prentice-Hall, Upper Saddle River, NJ,1999).

[17] S. Pfeiffer, Ph.D. thesis, Hamburg-Harburg University of Technology, 2014. https://doi.org/10.15480/882.1191

[18] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. (Wiley, Chichester, 2005).

[19] R.G. Lyons, *Understanding Digital Signal Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 2004).

[20] G. Stein, *IEEE Control Syst.*, **23** (2003) 12. http://dx.doi.org/10.1109/mcs.2003.1213600

[21] N. Amann *et al.*, *IEE Proc. Control Theory Appl.* **143** (1996) 217.

[22] C. Schmidt, Ph.D. thesis, Hamburg University of Technology, 2010. http://www-library.desy.de/cgi-bin/showprep.pl?desy-thesis-11-019

[23] A. Neumann, Ph.D. thesis, Humboldt University of Berlin, 2008. http://edoc.hu-berlin.de/docviews/abstract.php?id=29419

[24] A. Neumann *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **13** (2010) 082001. http://link.aps.org/doi/10.1103/PhysRevSTAB.13.082001

[25] F. Qiu *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **18** (2015) 092801. http://link.aps.org/doi/10.1103/PhysRevSTAB.18.092801

[26] F. Qiu, Disturbance observer-based control in low-level radio-frequency system in a compact energy recovery linac at KEK, in *LLRF Workshop 2015*, Shanghai, China, 2015.

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Coherence Properties of the Radiation from X-ray Free Electron Lasers

E.A. Schneidmiller and M.V. Yurkov
Deutsches Elektronen Synchrotron (DESY), Hamburg, Germany

Abstract
A description of the statistical and coherence properties of the radiation from X-ray free electron lasers is one of the most complicated topics in free electron laser physics. Results of studies in this field are distributed over many papers published during the past three decades. We have made an attempt to put together all knowledge on the subject available to date. We avoid complicated mathematical derivations, and put the main emphasis on physical features and final results. Free electron laser theory has reached mature status, making it possible to present final results in an elegant form. Application of similarity techniques to the results of numerical simulations allowed us to derive universal scaling relations for the main characteristics of an X-ray free electron laser operating in the saturation regime: efficiency, coherence time, degree of transverse coherence, and pointing stability of the radiation. Statistical and coherence properties of the higher harmonics of the radiation are highlighted as well.

Keywords
Free electron laser; X-ray laser; self-amplified spontaneous emission FEL; SASE FEL; statistics; fluctuations; coherence; temporal coherence; spatial coherence; harmonics.

1 Introduction

Single-pass free electron laser (FEL) amplifiers starting from shot noise in the electron beam have been intensively developed during the last decades. An origin for this development was an idea born in the early 1980s to extend the operating wavelength range of FELs to the vacuum ultraviolet and X-ray bands [1–5]. Following the terminology of quantum lasers (amplified spontaneous emission), the term 'self-amplified spontaneous emission (SASE)' in connection with an FEL amplifier, starting from the shot noise, has been introduced in Ref. [5]. We note that this essentially quantum terminology does not reflect the physical properties of the device. In fact, FELs belong to a separate class of vacuum tube devices, and their operation is completely described in terms of classical physics (see Ref. [6] for more details).

Significant amounts of effort have been invested in the development of high brightness injectors, beam formation systems, linear accelerators, and undulators. The result was a rapid extension of the wavelength range from infrared to hard X-rays within one decade [7–18]. The first dedicated user facility, FLASH at DESY in Hamburg, has been in operation since 2005 and provides a wavelength range from 3.1 nm to 80 nm [10–16]. The Linac Coherent Light Source (LCLS) at Stanford was the first FEL operating in the hard X-ray regime, in the 0.15–1.5 nm wavelength range [17]. SACLA extended the wavelength range down to 0.06 nm [18]. The PAL X-ray free electron laser (XFEL) has produced its first light and is being commissioned at the time of writing this review [19–21]. Two other dedicated facilities are under commissioning at the moment, the European XFEL and the SwissFEL [22–24]. The main mode of operation of these facilities is SASE. One more soft X-ray FEL user facility is FERMI FEL, which uses external seeding [25].

The high-gain FEL amplifier starting from shot noise in the electron beam is a simple system consisting of a relativistic electron beam and an undulator. The FEL collective instability in the electron

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.539

beam produces an exponential growth (along the undulator) of the modulation of the electron density on the scale of the undulator radiation wavelength. The initial seed for the amplification process is fluctuation of the electron beam current. Since shot noise in the electron beam is a stochastic process, the radiation produced by a SASE FEL also possesses stochastic features. Its properties are naturally described in terms of statistical optics using notions of probability density distribution functions of the fields and intensities, correlation functions, notions of coherence time, degree of coherence, etc.

Development of the theoretical description of the coherence properties of the radiation from SASE FELs has spanned more than twenty years (see [26–45]). This subject is rather complicated, and it is worth mentioning that theoretical predictions agree well with experimental results obtained at FLASH [10–13, 46–49]. Some averaged output characteristics of SASE FELs in the framework of the one-dimensional model have been obtained [26, 27]. An approach for time-dependent numerical simulations of SASE FELs has been developed [28, 29]. Realization of this approach allowed some statistical properties of the radiation from a SASE FEL operating in linear and non-linear regimes to be obtained [30, 31]. A comprehensive study of the statistical properties of the radiation from the SASE FEL in the framework of the same model is presented in Ref. [32]. It has been shown that a SASE FEL operating in the linear regime is a completely chaotic polarized radiation source, described with Gaussian statistics. Short-pulse effects (for pulse durations comparable with the coherence time) have been studied [29, 33, 34]. An important practical result was prediction of the significant suppression of the fluctuations of the radiation intensity after a narrow-band monochromator, for the case of SASE FEL operation in the saturation regime [33]. Statistical description of the chaotic evolution of the radiation from SASE FEL has been presented [35, 36].

The first analytical studies of the problem of transverse coherence relate to the late 1980s [37, 38]. Later on, more detailed studies have been performed [39]. The problem of start-up from the shot noise has been studied analytically and numerically for the linear stage of amplification, using an approach developed in Ref. [38]. It has been found that the process of formation of transverse coherence is more complicated than that given by a naive physical picture of transverse mode selection. Namely, in the case of perfect mode selection, the degree of transverse coherence is defined by the interdependence of the longitudinal and transverse coherence. Comprehensive studies of the evolution of transverse coherence in the linear and non-linear regime of SASE FEL operation have been performed [40–43]. It has been found that the coherence time and the degree of transverse coherence reach maximum values at the end of the linear regime. Maximum brilliance of the radiation is achieved at the very beginning of the non-linear regime, which is also referred as a saturation point [41, 42]. The output power of the SASE FEL increases continuously in the non-linear regime, while the brilliance decreases after passing the saturation point.

Another important subject related to mode degeneration is fundamental limitation of the pointing stability of SASE FEL radiation. The radiation from a SASE FEL always has a limited degree of transverse coherence. When the transverse size of the electron beam significantly exceeds the diffraction limit, the mode competition effect does not provide selection of the ground mode, and spatial coherence degrades, owing to the contribution of the higher-order transverse modes. Important consequences of this effect are fluctuations of the spot size and of the pointing stability of the photon beam [50–52]. These fluctuations are fundamental and originate from the shot noise in the electron beam. The pointing instability effect becomes more pronounced for shorter wavelengths. We analyse in detail the case of an optimized SASE FEL and derive universal dependencies applicable to all operating and planned X-ray FELs. It is shown that the hard X-ray FELs driven by low-energy beams may exhibit poor spatial coherence and bad pointing stability.

Radiation of the electron beam in the planar undulator contains a rich harmonic spectrum. This refers to both incoherent and coherent radiation. Over the past years, significant research efforts have been devoted to studying the process of generation of higher harmonics in high-gain FELs [53–66]. Such an interest has mainly been driven by practical needs for prediction of the properties of XFELs. A fraction of a higher-harmonic content is very important for users planning experiments at an XFEL facility. On the one hand, higher harmonics constitute rather harmful background for some classes of experiment. On the other hand, higher-harmonic radiation can significantly extend the operating band of the user facility. In both cases, it is highly desirable to know the properties of the higher-harmonic radiation. Analytical techniques have been used to predict properties of the higher harmonics for FEL amplifiers operating in the linear mode of operation [56, 57]. However, most of the radiation power is produced in the non-linear regime, and a set of assumptions needs to be accepted in order to estimate the saturation power of higher harmonics on the base of extrapolation of analytical results. As for statistical properties, these could not be extrapolated from linear theory at all. Many studies have been performed with numerical simulation codes. These studies developed in two directions. The first direction is investigations of higher-harmonic phenomena by means of steady-state codes [58–61]. Although the results of these studies are applicable to externally seeded FEL amplifiers, it is relevant to appreciate that they gave the first predictions for high radiation power in higher harmonics of SASE FELs. Another direction was an extraction of the time structure for the beam bunching from time-dependent simulation code with subsequent use of analytical formulae of linear theory [56]. Recent studies of the mechanism of non-linear harmonic generation in SASE FEL have been presented [64, 65].

This review is mainly based on our papers published during the last two decades. References to original papers of our colleagues working on the same problems are also made, and we believe that the reader can find further details therein. Analysis of higher harmonics of FEL radiation is restricted to the odd harmonics in the planar undulator. We consider only the non-linear harmonic generation mechanism. Owing to limitations of space, it was impossible to derive the results from first principles, and we decided to concentrate on a description of physical models and final results. The results themselves are strict results of FEL theory. The theory has now reached a mature status and allows rather complicated phenomena to be described in an elegant way. In addition, we widely use application of the similarity techniques to the results of numerical simulations. This allowed us to describe the saturation regime in detail, which is important from a practical point of view. Universal formulae that can be used in practical calculation are distributed across the text, which can be inconvenient for the reader. The review proceeds as follows. In Section 2, we present general statistical properties of SASE FEL radiation. We show that a SASE FEL operating in the exponential growth regime can be described by Gaussian statistics, and that properties of the radiation correspond to the properties of completely chaotic polarized light. In the framework of the one-dimensional model, in Section 3, we describe temporal properties of the radiation. Topics related to spatial coherence are described in Section 4, in the framework of three-dimensional FEL theory. Essential practical formulae are collected in Appendices A and B for the results of the one-dimensional and three-dimensional FEL theory, respectively. Finally, Appendix C describes statistical techniques for measurements of the main parameters of SASE FEL radiation. This is an extremely powerful technique allowing measurement of the FEL parameter ρ, coherence time, photon pulse duration, number of longitudinal and transverse modes, and degree of transverse coherence. It is based on fundamental principles, and measured values have strict physical meaning.

2 General properties of SASE FEL radiation

The amplification process in the SASE FEL starts from the shot noise in the electron beam; then it passes the stage of exponential amplification (high-gain linear stage) and finally enters saturation stage (see Fig. 1). At the initial stage of amplification, coherence properties are poor, and the radiation consists of a large number of transverse and longitudinal modes [6, 67–73]. In the exponential stage of amplification, transverse modes with higher gain dominate over modes with lower gain as the undulator length progresses. This feature is also known as the mode competition process. Longitudinal coherence also improves in the high-gain linear regime [32, 64, 71]. The mode selection process stops at the onset of the non-linear regime, and maximum values of the degree of transverse coherence and the coherence time are reached at this point. If one traces the evolution of the brilliance of the radiation along the undulator length, there is always a point, which we define as the saturation point, where the brilliance reaches a

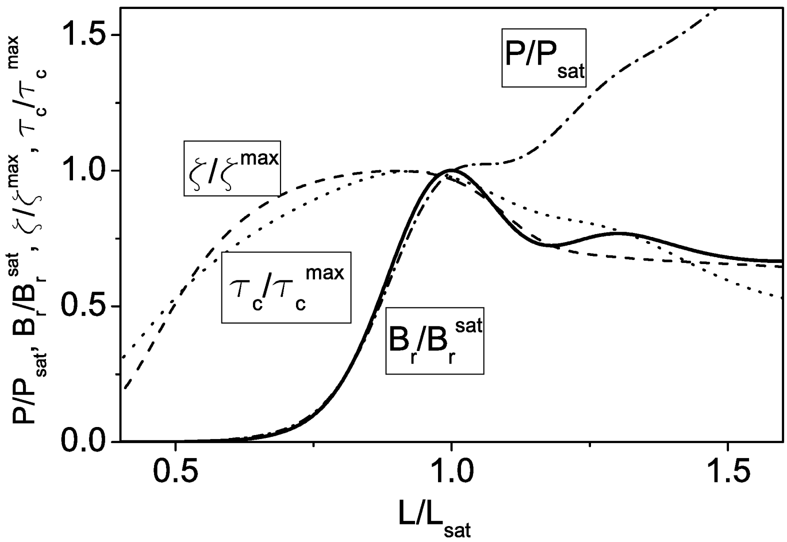


Fig. 1: Evolution of main characteristics of SASE FEL along the undulator: brilliance (solid line), radiation power (dash-dotted line), degree of transverse coherence (dashed line), and coherence time (dotted line). Brilliance and radiation power are normalized to saturation values. Coherence time and degree of transverse coherence are normalized to the maximum values. Undulator length is normalized to saturation length. The plot has been derived from the parameter set corresponding to $\hat{\epsilon} = 1$. Calculations made with simulation code FAST [77].

maximum value [41, 42]. The undulator length to saturation is in the range from about nine (hard X-ray SASE FELs) to eleven (visible-range SASE FELs) field gain lengths L_{g} [41, 42]. Figures 2 and 3 show the evolution of the temporal and spectral structure of the radiation pulse along the undulator: at $0.5L_{\mathrm{g}}$ (beginning of the undulator), $5L_{\mathrm{g}}$ (high-gain linear regime), and $10L_{\mathrm{g}}$ (saturation regime). Figure 4 shows snapshots of the intensity distributions across a slice of the photon pulse. We see that many transverse radiation modes are excited when the electron beam enters the undulator. The radiation field generated by a SASE FEL consists of wavepackets (spikes [29]) which originate from fluctuations of the electron beam density. The typical length of a spike is approximately the coherence length. The spectrum of the SASE FEL radiation also exhibits a spiky structure. The spectrum width is inversely proportional to the coherence time, and a typical width of a spike in a spectrum is inversely proportional to the pulse duration. The amplification process selects a narrow band of the radiation, the coherence time increases, and the spectrum shrinks. Transverse coherence is also improved, owing to the mode selection process (Eq. (25)).

Figure 5 shows probability distributions of the instantaneous power density $I \propto |\tilde{E}|^2$ (top) and the instantaneous radiation power $P \propto \int I(\vec{r}_\perp)\mathrm{d}\vec{r}_\perp$ (bottom). We see that transverse and longitudinal distributions of the radiation intensity exhibit rather chaotic behaviour. However, probability distributions of the instantaneous power density I and of the instantaneous radiation power P look more elegant and seem to be described by simple functions. The origin of this fundamental simplicity relates to the properties of the electron beam. The shot noise in the electron beam has a statistical nature that significantly influences the characteristics of the output radiation from a SASE FEL. Fluctuations of the electron beam current density serve as input signals in a SASE FEL. These fluctuations always exist in the electron beam, owing to the effect of shot noise. Initially, fluctuations are not correlated in space and time, but when the electron beam enters the undulator, beam modulation at frequencies close to the resonance frequency of the FEL amplifier initiates the process of the amplification of coherent radiation.

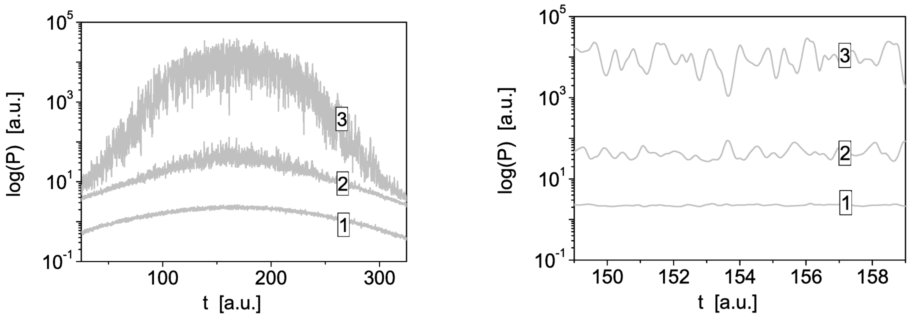


Fig. 2: Temporal structure of the radiation pulse at different undulator lengths. Indexes 1, 2, and 3 correspond to undulator lengths of $0.5L_{\mathrm{g}}$, $5L_{\mathrm{g}}$, and $10L_{\mathrm{g}}$, respectively. The plot in the right panel is an enlarged portion of the plot in the left panel. Calculations made using simulation code FAST [77].

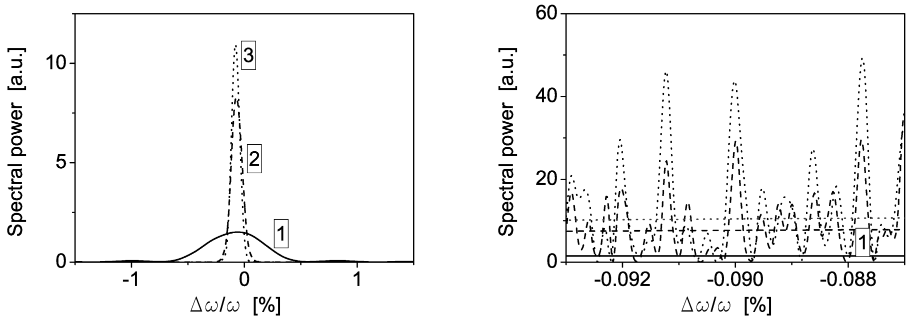


Fig. 3: Spectral structure of the radiation pulse at different undulator lengths. Solid, dashed, and dotted lines correspond to undulator lengths of $0.5L_{\mathrm{g}}$, $5L_{\mathrm{g}}$, and $10L_{\mathrm{g}}$, respectively. Left: envelope of the radiation spectrum. Right: enlarged portion of the radiation spectrum. Calculations made with simulation code FAST [77].

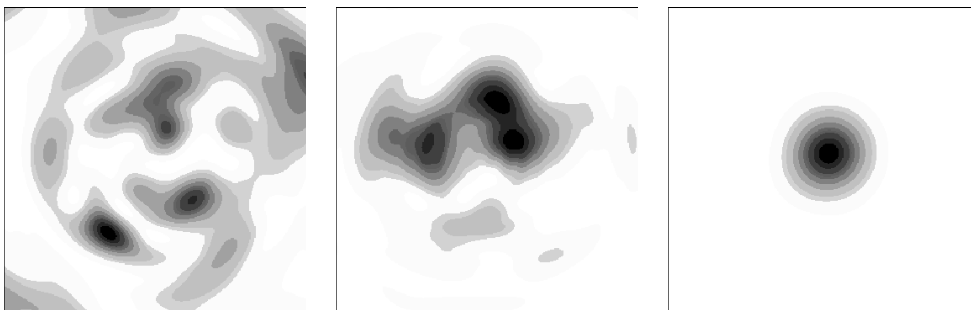


Fig. 4: Snapshots of the power density distribution in a slice at an undulator length of (left) $0.5L_{\mathrm{g}}$, (middle) $5L_{\mathrm{g}}$, and (right) $10L_{\mathrm{g}}$. Calculations made with simulation code FAST [77].

Let us consider a microscopic picture of the electron beam current at the entrance of the undulator. The electron beam current consists of moving electrons randomly arriving at the entrance of the undulator:

$$I(t) = (-e) \sum_{k=1}^{N} \delta(t - t_k) \ ,$$

where $\delta(\ldots)$ is the delta function, $(-e)$ is the charge of the electron, N is the number of electrons in a bunch and t_k is the random arrival time of the electron at the undulator entrance. The electron beam

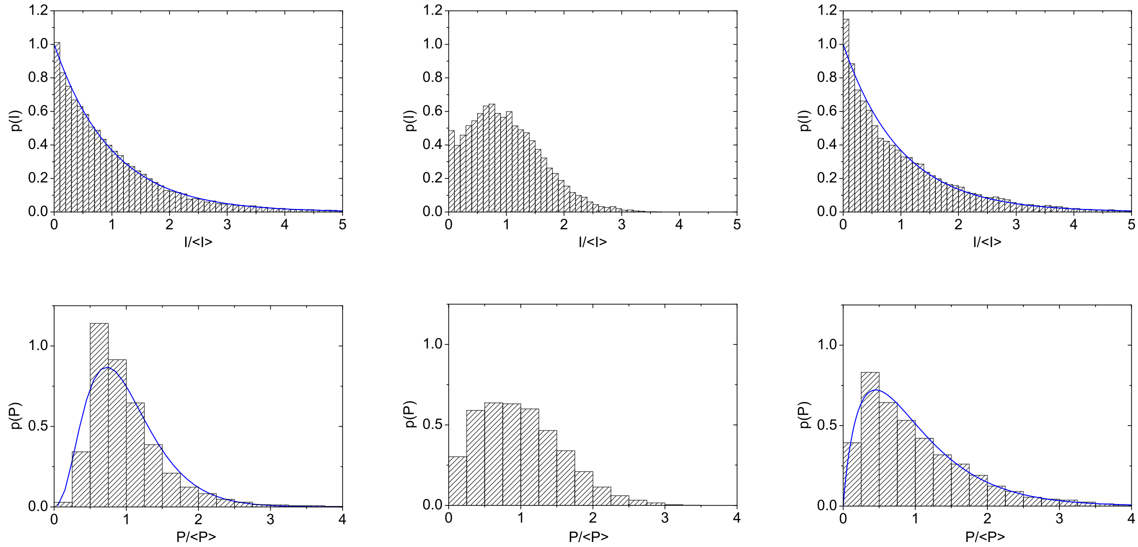


Fig. 5: Probability density distributions of (top) instantaneous power density $I = |\tilde{E}|^2$ and (bottom) instantaneous radiation power P from a SASE FEL at different stages of amplification: linear regime, saturation regime, and deep non-linear regime (undulator lengths, $5L_\mathrm{g}$, $10L_\mathrm{g}$, and $15L_\mathrm{g}$, respectively). Top: solid lines on the power density histograms represent negative exponential distribution (Eq. (2)). Bottom: solid lines on power histograms represent gamma distribution (Eq. (3)) with $M = 1/\sigma_\mathrm{P}^2$. $\hat{\epsilon} = 2$. Calculations made with simulation code FAST [77].

current $I(t)$ and its Fourier transform $\bar{I}(\omega)$ are:

$$I(t) = (-e)\sum_{k=1}^{N}\delta(t - t_k) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\bar{I}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega \ ,$$

$$\bar{I}(\omega) = \int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\omega t}I(t)\mathrm{d}t = (-e)\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}\omega t_k} \ . \tag{1}$$

It follows from Eq. (1) that the Fourier transformation of the input current, $\bar{I}(\omega)$, is the sum of large number of complex phasors with random phases $\phi_k = \omega t_k$. Thus, harmonics of the electron beam current can be described with Gaussian statistics.

The FEL process is simply an amplification of the initial shot noise in the narrow band near the resonance wavelength λ when both harmonics of the beam current and radiation are increasing. An FEL amplifier operating in the linear regime is simply a linear filter, and the Fourier harmonic of the radiation field is simply proportional to the Fourier harmonic of the electron beam current, $\bar{E}(\omega) = H_A(\omega - \omega_0)\bar{I}(\omega)$. Thus, the statistics of the radiation are Gaussian—the same as of the shot noise in the electron beam. This kind of radiation is usually referred to as completely chaotic polarized light, a well-known object in the field of statistical optics [74]. For instance, the higher-order correlation functions (time and spectral) are expressed via the first-order correlation function. The spectral density of the radiation energy and the first-order time correlation function form a Fourier transform pair (Wiener–Khinchin theorem). The real and imaginary parts of the slowly varying complex amplitudes of the electric field of the electromagnetic wave, $\tilde{E}$, have a Gaussian distribution. The instantaneous power density, $I = |\tilde{E}|^2$, fluctuates in accordance with the negative exponential distribution (see Fig. 5):

$$p(I) = \frac{1}{\langle I\rangle}\exp\left(-\frac{I}{\langle I\rangle}\right) \ . \tag{2}$$

Any integral of the power density, such as radiation power P (or radiation pulse energy E), fluctuates in accordance with the gamma distribution:

$$p(P) = \frac{M^M}{\Gamma(M)} \left(\frac{P}{\langle P \rangle} \right)^{M-1} \frac{1}{\langle P \rangle} \exp\left(-M \frac{P}{\langle P \rangle} \right) . \tag{3}$$

where $\Gamma(M)$ is the gamma function, with argument $M = 1/\sigma_P^2$, and $\sigma_P^2 = \langle (P - \langle P \rangle)^2 \rangle / \langle P \rangle^2$ is the relative dispersion of the radiation power. For completely chaotic polarized light, the parameter M has a clear physical interpretation—it is the number of modes [6]. Thus, the relative dispersion of the radiation power directly relates to the coherence properties of the SASE FEL operating in the linear regime.

When the amplification process enters a non-linear stage and reaches saturation, the statistics of the radiation significantly deviate from Gaussian. A particular signature of this change is illustrated in Fig. 5. We see that the probability distribution of the radiation intensity is not negative exponential, and that the probability distribution of the radiation power visibly deviates from a gamma distribution. As yet, there is no analytical description of the statistics in the saturation regime, and we refer the reader to the analysis of the results of numerical simulations [41]. A general feature of the saturation regime is that fluctuations in radiation intensity are significantly suppressed. When we trace the amplification process further in the non-linear regime, we find that fluctuations of the radiation intensity and radiation power increase, and that relevant probability distributions tend to those given by Eqs. (2) and (3). This behaviour hints that the properties of the radiation from a SASE FEL operating in the deep non-linear regime tend to be those of completely chaotic polarized light [32, 41].

Another practical problem refers to the probability distributions of the radiation intensity in the frequency domain, like that filtered by a monochromator. For SASE FEL radiation produced in the linear regime, the probability distribution radiation intensity is defined by Gaussian statistics, and it is negative exponential for a narrow-band monochromator. When the amplification process enters the saturation regime, this property still holds for the case of long electron pulses [6, 32] but is violated significantly for the case of short electron bunches, approximately equal to or less than the coherence length. In the latter case, fluctuations of the radiation intensity after the narrow-band monochromator are significantly suppressed, as has been predicted theoretically and measured experimentally at the FLASH FEL at DESY, operating in a femtosecond mode [33, 46].

All these considerations are related to the fundamental harmonic of the SASE FEL radiation. Radiation from a SASE FEL with a planar undulator has rich harmonics contents. Intensities of even harmonics are suppressed [75], but odd harmonics provide visible contribution to the total radiation power [53–65]. Comprehensive studies of the statistical properties of the odd harmonics have been conducted [64]. It has been found that the statistics of the high-harmonic radiation from the SASE FEL changes significantly with respect to the fundamental harmonic (with respect to Gaussian statistics). For the fundamental harmonic, the probability density function of the intensity is the negative exponential distribution: $p(W) = \langle W \rangle^{-1} \exp(-W/\langle W \rangle)$. The mechanism of higher-harmonic generation is equivalent to the transformation of the intensity W as $z = (W)^h$, where h is the harmonic number. It has been shown [64] that the probability distribution for the intensity of the hth harmonic is given by:

$$p(z) = \frac{z}{h\langle W \rangle} z^{(1-h)/h} \exp(-z^{1/h}/\langle W \rangle) . \tag{4}$$

The expression for the mean value is $\langle z \rangle = h! \langle W \rangle^h$. Thus, the hth-harmonic radiation for the SASE FEL has an intensity level roughly $h!$ times larger than the corresponding steady-state case, but with more shot-to-shot fluctuations than the fundamental [56]. The non-trivial behaviour of the intensity of the high harmonics reflects the complicated non-linear transformation of the fundamental harmonic statistics. In this case, Gaussian statistics are no longer valid. Practically, this behaviour occurs only at the very end of a high-gain exponential regime when coherent radiation intensity exceeds incoherent radiation intensity. When amplification enters the non-linear stage, probability distributions change dramatically on the scale of the gain length, and in the saturation regime (and further downstream the undulator) the probability distributions of the radiation intensity of higher harmonics are already close to the negative exponential distribution [64].

2.1 Definitions of the statistical properties of radiation

In the proceeding sections we present a systematic description of the main statistical properties of SASE FEL radiation. We describe the radiation in terms of statistical optics [74]. Longitudinal and transverse coherence are described in terms of correlation functions. The first-order time correlation function, $g_1(t,t')$, is defined as:

$$g_1(\vec{r},t-t') = \frac{\langle \tilde{E}(\vec{r},t)\tilde{E}^*(\vec{r},t')\rangle}{\left[\langle \mid \tilde{E}(\vec{r},t)\mid^2\rangle\langle \mid \tilde{E}(\vec{r},t')\mid^2\rangle\right]^{1/2}} \ . \tag{5}$$

For a stationary random process, the time correlation functions are dependent on only one variable, $\tau = t - t'$. The coherence time is defined as [6, 76]:

$$\tau_{\mathrm{c}} = \int\limits_{-\infty}^{\infty} |g_1(\tau)|^2 \mathrm{d}\tau \ . \tag{6}$$

The transverse coherence properties of the radiation are described in terms of the transverse correlation functions. The first-order transverse correlation function is defined as:

$$\gamma_1(\vec{r}_\perp,\vec{r\prime}_\perp,z,t) = \frac{\langle \tilde{E}(\vec{r}_\perp,z,t)\tilde{E}^*(\vec{r\prime}_\perp,z,t)\rangle}{\left[\langle|\tilde{E}(\vec{r}_\perp,z,t)|^2\rangle\langle|\tilde{E}(\vec{r\prime}_\perp,z,t)|^2\rangle\right]^{1/2}} \ ,$$

where $\tilde{E}$ is the slowly varying amplitude of the amplified wave, $E = \tilde{E}(\vec{r}_\perp,z,t)\mathrm{e}^{\mathrm{i}\omega_0(z/c-t)} + \mathrm{C.C.}$ We consider the model of a stationary random process, meaning that γ_1 does not depend on time. Following Ref. [41], we define the degree of transverse coherence as:

$$\zeta = \frac{\int|\gamma_1(\vec{r}_\perp,\vec{r\prime}_\perp)|^2 I(\vec{r}_\perp)I(\vec{r\prime}_\perp)\mathrm{d}\vec{r}_\perp\mathrm{d}\vec{r\prime}_\perp}{\left[\int I(\vec{r}_\perp)\mathrm{d}\vec{r}_\perp\right]^2} \ , \tag{7}$$

where $I \propto |\tilde{E}|^2$ is the radiation intensity.

An important figure of merit of the radiation source is the degeneracy parameter δ, the number of photons per mode (coherent state). Note that when $\delta \gg 1$, classical statistics are applicable, while a quantum description of the field is necessary as soon as δ is comparable to (or less than) one. Using the definitions of the coherence time (Eq. (6)) and of the degree of transverse coherence (Eq. (7)), we define the degeneracy parameter as

$$\delta = \dot{N}_{\mathrm{ph}}\tau_{\mathrm{c}}\zeta \ , \tag{8}$$

where $\dot{N}_{\mathrm{ph}}$ is the photon flux. The peak brilliance of the radiation from an undulator is defined as a transversely coherent spectral flux:

$$B_{\mathrm{r}} = \frac{\omega\mathrm{d}\dot{N}_{\mathrm{ph}}}{\mathrm{d}\omega}\frac{\zeta}{(\lambda/2)^2} = \frac{4\sqrt{2}c\delta}{\lambda^3} \ . \tag{9}$$

When deriving the right-hand term of the equation, we used the fact that the spectrum shape of SASE FEL radiation in a high-gain linear regime and near saturation is close to Gaussian [6]. In this case, the r.m.s. spectrum bandwidth σ_ω and coherence time obey the equation $\tau_{\mathrm{c}} = \sqrt{\pi}/\sigma_\omega$.

3 Temporal coherence of SASE FEL radiation

In this section, we present a comprehensive study of the phenomena related to the temporal coherence of SASE FEL radiation. The study is performed in the framework of the one-dimensional model with time-dependent simulation code FAST [6, 77]. We restrict our study to odd harmonics produced in the SASE FEL. We omit from consideration an effect of self-consistent amplification of the higher harmonics [56, 66]. In other words, we solve only the electrodynamic problem, assuming that particle motion is governed by the fundamental harmonic. The latter approximation is valid when the power in higher harmonics is much less than in the fundamental. We apply similarity techniques to the results of numerical simulations and derive universal relations describing general properties of the odd harmonics in the SASE FEL: power, statistical, and spectral properties. The results are illustrated for the first, third, and fifth harmonics having practical importance for X-ray FELs.

We consider planar undulator with the magnetic field:

$$H_z(z) = H_\mathrm{w} \cos(2\pi z/\lambda_\mathrm{w}) \ ,$$

where λ_w is undulator period, and H_w is the peak magnetic field. In the SASE FEL, the radiation is produced by the electron beam during a single pass of the undulator. The amplification process starts from shot noise in the electron beam. During the amplification process, a powerful, coherent radiation is produced, having narrow-band near-resonance wavelength:

$$\lambda_0 = \frac{\lambda_\mathrm{w}}{2\gamma^2}(1+K^2) \ , \tag{10}$$

where $K = e\lambda_\mathrm{w}H_\mathrm{w}/(2\sqrt{2}\pi mc^2)$ is the r.m.s. undulator parameter, γ is relativistic factor, and $(-e)$ and m are the charge and mass of an electron, respectively.

In the framework of the one-dimensional model, we consider amplification of the plane electromagnetic wave by the electron beam in the undulator. When space charge and energy spread effects can be neglected, operation of an FEL amplifier is described in terms of the gain parameter Γ and efficiency parameter (or FEL parameter) ρ (see, e.g. [4, 6]):

$$\rho = \frac{\lambda_\mathrm{w}}{4\pi}\left[\frac{4\pi^2 j_0 K^2 A_\mathrm{JJ}^2}{I_\mathrm{A}\lambda_\mathrm{w}\gamma^3}\right]^{1/3} \ , \qquad \Gamma = \frac{4\pi\rho}{\lambda_\mathrm{w}} \ . \tag{11}$$

Here, j_0 is the beam current density, $I_\mathrm{A} = mc^3/e \simeq 17$ kA, and $\omega = 2\pi c/\lambda$ is frequency of electromagnetic wave. The coupling factor K_h is given by

$$K_h = K(-1)^{(h-1)/2}\left[J_{(h-1)/2}(Q) - J_{(h+1)/2}(Q)\right] \ , \tag{12}$$

where $Q = K^2/[2(1+K^2)]$. The FEL amplifier is a resonance device with an amplification bandwidth of about $\Delta\omega/\omega_0 \simeq 2\rho$ around the resonance frequency $\omega_0 = 2\pi c/\lambda_0$. In the linear stage of amplification, the radiation power W increases exponentially along the undulator length, $W \propto \exp[2z/L_\mathrm{g}]$, and the field gain length is about $L_\mathrm{g} \simeq 2/(\Gamma\sqrt{3})$. Saturation of the FEL amplifier occurs when relative energy loss by the electrons at one field gain length is approximately equal to the saturation parameter ρ.

A complete description of the start-up from shot noise in the FEL amplifier can be made only with time-dependent simulations of the FEL process. We do not present here general technical details of the time-dependent simulations; they are described in detail elsewhere [6, 77]. Details related to the particle loading tool can be found in Ref. [78]. We note merely that, within the accepted approximation (the particle's dynamics are governed by the fundamental harmonic), we can simply calculate the odd harmonics from the particle distribution, and the amplitude of the electric field scales as

$$E(z,t) \propto K_h \int\limits_0^z a_h(z', t - z'/c)\mathrm{d}z' \ , \tag{13}$$

where a_h is the hth harmonic of the beam bunching. Thus, we find that the coupling factor K_h, and the time-dependent integral of the beam bunching are factorized. This allows us to extract the universal ratio of the power of higher harmonics to the power of the fundamental harmonic.

3.1 Statistical properties of the odd harmonics of the radiation from SASE FEL

In this section, we present the results of numerical studies of the operation of the SASE FEL in the linear and non-linear regimes. In the framework of the accepted model, the input parameter of the system is the number of cooperating electrons $N_\mathrm{c} = I/(e\rho\omega_0)$, where I is the beam current. Most of the statistical characteristics of the SASE FEL process are functions of N_c, only at fixed z coordinate [6,32]. A typical range of the values of N_c is 10^6–10^9 for the SASE FELs of wavelength range from X rays to the infrared. The numerical results presented in this section, are calculated for the value $N_\mathrm{c} = 3\times 10^7$, which is typical for a vacuum ultraviolet FEL. It is worth mentioning that the dependence of the output parameters of the SASE FEL on the value of N_c is rather weak, in fact logarithmic. Therefore, the obtained results are pretty general and can be used for the estimation of the parameters of actual devices with sufficient accuracy.

3.1.1 Temporal characteristics

Figure 6 presents a typical time structure of the first and the third harmonic of the radiation from a SASE FEL at different undulator lengths $\hat{z} = \Gamma z = 10$–13. The normalized power of the hth harmonic is defined as $\hat{\eta}_h = \langle W_h \rangle \times (K_1/K_h)^2/(\rho W_\mathrm{b})$. Here, $\langle W_h \rangle$ is averaged radiation power in the hth harmonic, and $W_\mathrm{b} = \gamma mc^2 I/e$ is the electron beam power. The longitudinal coordinate along the pulse is $\hat{s} = \rho\omega_0(z/\bar{v}_z - t)$, and $\bar{v}_z = v - cK^2/(2\gamma^2)$ is the velocity of the electron along the z axis, averaged over the undulator period. The head of the pulse is located in the positive direction of $\hat{s}$. The plot for the averaged power of the first harmonic is shown in Fig. 7 with a solid line. It can be seen that saturation is achieved at the undulator $\hat{z} = 13$. The saturation length is described well in terms of the number of cooperating electrons N_c [6,32]:

$$\hat{z}_\mathrm{sat} \simeq 3 + \frac{1}{\sqrt{3}} \ln N_\mathrm{c} \ . \qquad (14)$$

The normalized efficiency at saturation, $\hat{\eta}_\mathrm{sat} = \langle W_\mathrm{sat} \rangle/(\rho W_\mathrm{b}) \simeq 1.08$, is almost independent of the value of N_c. The dashed and dotted lines in the figure show a normalized power ratio, $\hat{\eta}_h/\hat{\eta}_1 = (\langle W_h \rangle/\langle W_1 \rangle) \times (K_1/K_h)^2$, for the third and fifth harmonics. One can notice that the power of the higher harmonics is greater than the shot noise level only at the end of the linear regime. This becomes clear if one takes into account the fact that the shot noise level of the beam bunching is about $1/\sqrt{N_\mathrm{c}}$. We consider an example typical for a vacuum ultraviolet FEL with $N_\mathrm{c} = 3 \times 10^7$, which corresponds to the shot noise beam bunching $a \simeq 2 \times 10^{-4}$. When the FEL amplifier operates in the linear regime, odd harmonics increase as a_1^h, and we expect from this simple physical estimate that coherent contribution into higher harmonics can exceed the shot noise level only for the values of the beam bunching at the fundamental harmonic $a_1 \gtrsim 0.1$, i.e. at the end of the linear regime. Note that the shot noise level increases when approaching the X-ray region.

The plots presented in Fig. 6 allow the evolution of the third harmonic power to be traced from $\hat{z} = 10$ (when it just starts to exceed the shot noise level) up to saturation point $\hat{z} = 13$. At $\hat{z} = 10$–11, the SASE FEL operates in the high-gain linear regime, and the beam bunching in the fundamental harmonic is small, $|a_1| \ll 1$. In this case, one can expect the well-known mechanism of the higher-harmonic generation, i.e. $a_h \propto a_1^h$, and the spikes of the third harmonic radiation become rather pronounced. However, the noisy nature of the SASE FEL makes a big difference to the behaviour of the growth rates with respect to predictions given in the framework of steady-state simulations [58, 59]. Analysing the plot for the power growth rate (see Fig. 8) we can state that, in practical situations, the prediction of the steady-state theory (the growth rate of higher harmonics is proportional to the harmonic number) is valid only for the third harmonic, and then only on a short piece of undulator close to saturation, of about one

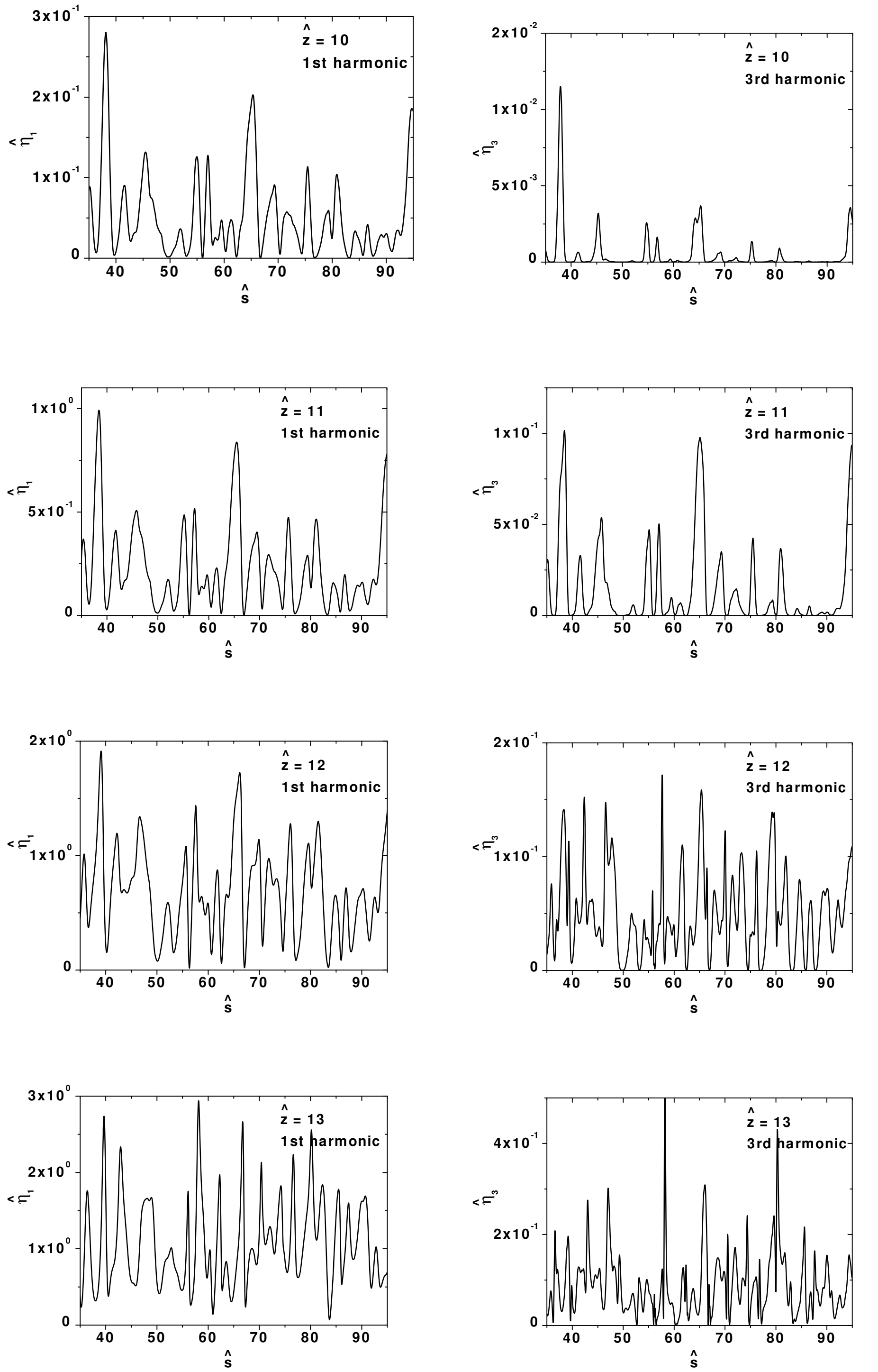


Fig. 6: Normalized power in the radiation pulse versus $\hat{s} = \rho\omega_0(z/\bar{v}_z - t)$ at different lengths of the FEL amplifier $\hat{z} = 10$–13. Left and right columns correspond to the fundamental and third harmonic, respectively. Calculations made with simulation code FAST [77].

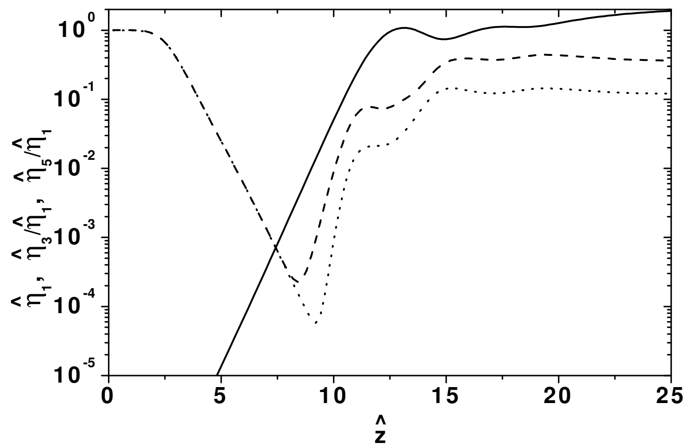


Fig. 7: Normalized averaged power of a fundamental harmonic of SASE FEL, $\hat{\eta}_1 = P_1/(\rho P_{\mathrm{beam}})$, as a function of a normalized undulator length (solid line). Dashed and dotted lines show normalized power ratio, $\hat{\eta}_h/\hat{\eta}_1 = (W_h/W_1) \times (K_1/K_h)^2$ for the third and fifth harmonics, respectively. Calculations made with simulation code FAST [77].

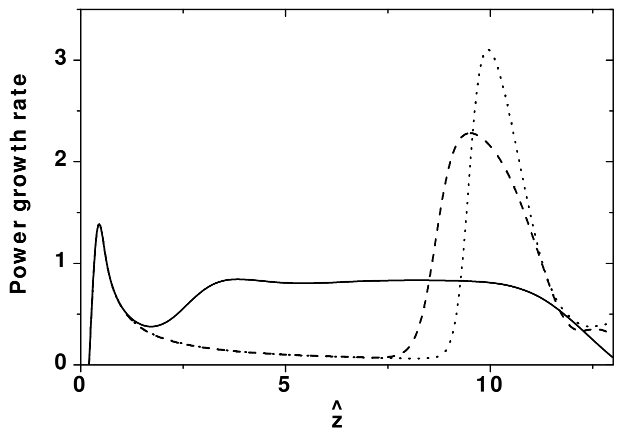


Fig. 8: Normalized power growth rate for the first, third, and fifth harmonic (solid, dashed, and dotted line, respectively). Calculations made with simulation code FAST [77].

gain length. Also, a prediction for the relation between averaged values of the beam bunching at the third harmonic, $< |a_3|^2 >= 6 < |a_1|^2 >^3$, holds only approximately, and is strongly violated for higher harmonics, because of the strong contribution of the shot noise. This feature of the SASE FEL has been highlighted qualitatively in early papers [56] with the analysis of simulation results obtained with code GINGER [55]. Here we simply present a more quantitative study.

The plots in Fig. 7 present a general result for a ratio of the power in the higher harmonics with respect to the fundamental one. For the saturation, we find a universal dependency:

$$\frac{\langle W_3 \rangle}{\langle W_1 \rangle}|_{\mathrm{sat}} = 0.094 \times \frac{K_3^2}{K_1^2} \, , \qquad \frac{\langle W_5 \rangle}{\langle W_1 \rangle}|_{\mathrm{sat}} = 0.03 \times \frac{K_5^2}{K_1^2} \, . \tag{15}$$

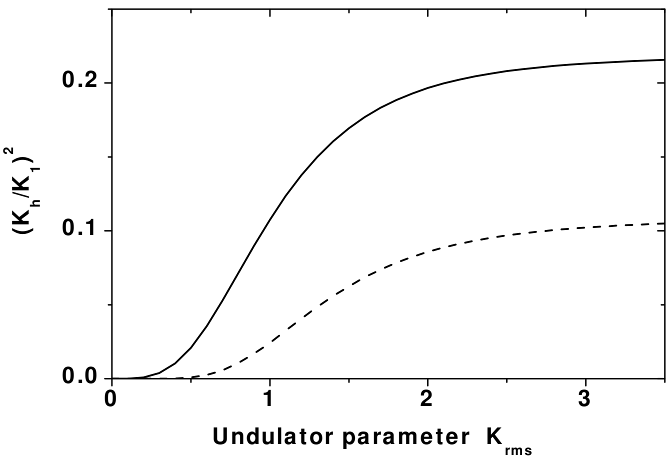


Fig. 9: Ratio of coupling factors, $(K_h/K_1)^2$, for the third (solid line) and the fifth (dashed line) harmonics with respect to the fundamental harmonic versus the r.m.s. value of undulator parameter K_{rms}. Calculations made with simulation code FAST [77].

Universal functions for the ratio $(K_h/K_1)^2$ are plotted in Fig. 9. Asymptotic values for large values of the undulator parameter are $(K_3/K_1)^2 \simeq 0.22$ and $(K_5/K_1)^2 \simeq 0.11$. Thus, we can state that the contribution of the third harmonic into the total radiation power of SASE FEL at saturation could not exceed a level of 2%. Thus, its influence on the beam dynamics should be small. This result justifies a basic assumption used for derivation of a universal relation, Eq. (15). A contribution of the fifth harmonic to the total power at saturation could not exceed the value of 0.3%.

Another important topic is an impact of the electron beam quality on the non-linear harmonic generation process. In the framework of the one-dimensional theory, this effect is described by the energy spread parameter $\hat{\Lambda}^2_{\mathrm{T}}$ [6]:

$$\hat{\Lambda}^2_{\mathrm{T}} = \frac{\langle(\Delta E)^2\rangle}{\rho^2 E_0^2} ,$$

where $\langle(\Delta E)^2\rangle$ is the r.m.s. energy spread and $E_0 = \gamma mc^2$ is the nominal energy of the electrons. Thus, the result given by Eq. (15) is generalized to the case of finite energy spread with the plot presented in Fig. 10. We see that the energy spread in the electron beam suppresses the power of the higher harmonics. Within the practical range of $\hat{\Lambda}^2_{\mathrm{T}}$, this suppression can be about a factor of three for the third harmonic, and about an order of magnitude for the fifth harmonic. For practical estimates, one should use an effective value of the energy spread, describing the contribution of the energy spread and the emittance to the longitudinal velocity spread [6]:

$$\frac{\langle(\Delta E)^2\rangle_{\mathrm{eff}}}{E_0^2} = \frac{\langle(\Delta E)^2\rangle}{E_0^2} + \frac{2\gamma_z^4\epsilon^2}{\beta^2} ,$$

where γ_z is the longitudinal relativistic factor ($\gamma_z^2 = \gamma^2/(1+K^2)$), ϵ is the beam emittance, and β is the focusing beta function. The plot in Fig. 10 covers the practical range of parameters for X-ray FELs. The saturation length at $\hat{\Lambda}^2_{\mathrm{T}} = 0.5$ is increased by a factor of 1.5 with respect to the 'cold' beam case $\hat{\Lambda}^2_{\mathrm{T}} = 0$.

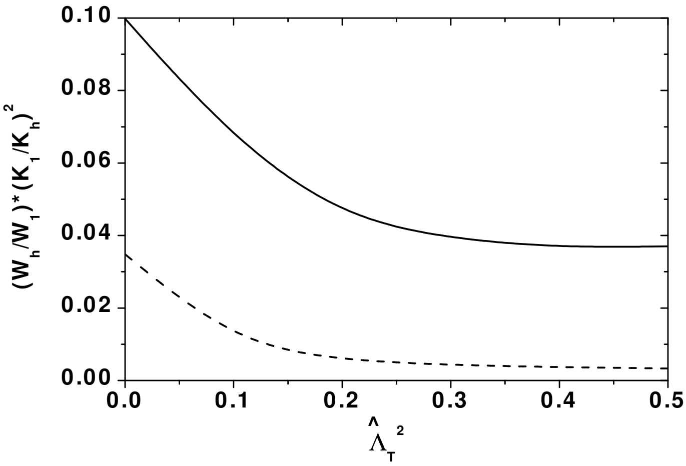


Fig. 10: Normalized power ratio at saturation, $(W_h/W_1) \times (K_1/K_h)^2$, for the third (solid line) and fifth (dashed line) harmonic as a function of energy spread parameter $\hat{\Lambda}_{\mathrm{T}}^2$. SASE FEL operates at saturation. Calculations made with simulation code FAST [77].

3.1.2 Probability distributions

The next step of our study is the behaviour of the probability distribution of the instantaneous power. In Fig. 11 we show the normalized r.m.s. deviation of the instantaneous radiation power, $\sigma_{\mathrm{w}} = \langle (W - \langle W \rangle)^2 \rangle^{1/2} / \langle W \rangle$, as a function of the undulator length. We see that, at the initial stage of SASE FEL operation, the r.m.s. deviation of the instantaneous power is equal to one for all harmonics. As we already discussed in Section 2, this is a consequence of the start-up from the shot noise in the electron beam. The statistical properties of the undulator radiation and of the radiation from SASE FEL operating in the linear regime are governed by Gaussian statistics [6, 32]. An important feature of the Gaussian statistics is that the normalized r.m.s. deviation of the instantaneous radiation power is equal to the unity. For the fundamental harmonic, the statistics of the radiation are non-Gaussian when the amplification process enters the non-linear mode [6, 32]. For the higher harmonics, non-Gaussian statistics take place when the non-linear harmonic generation starts to dominate over incoherent radiation (at $\hat{z} \gtrsim 8$ in the present numerical example). Analytical theory of non-linear harmonic generation [56] predicts a value of $\sigma_{\mathrm{w}} \simeq 4$ for the third harmonic. Analysis of the relevant curve in Fig. 11 shows that this prediction holds approximately in a short piece of the undulator length only. As we explained, this is because non-linear harmonic generation starts to dominate over incoherent radiation only at the values of the beam bunching at the fundamental harmonic $a_1 \sim 0.1$. However, at such a value of the beam bunching, the modulation of the beam density already deviates from a sinusoidal shape, owing to non-linear effects.

Probability density distributions for the instantaneous power of the fundamental and the third harmonic are presented in Fig. 12. The SASE radiation is a stochastic object and, at a given time, it is impossible to predict the amount of energy that flows to a detector. The initial modulation of the electron beam is defined by the shot noise and has a white spectrum. The high-gain FEL amplifier cuts and amplifies only a narrow frequency band of the initial spectrum $\Delta\omega/\omega \ll 1$. In the time domain, the temporal structure of the fundamental harmonic radiation is chaotic with many random spikes, with a typical duration given by the inverse width of the spectrum envelope. Even without performing numerical simulations, we can describe some general properties of the fundamental harmonic of the radiation from the SASE FEL operating in the linear regime. Indeed, in this case we deal with Gaussian statistics. As a result, the probability distribution of the instantaneous radiation intensity W should be the negative exponential probability density distribution of Eq. (2) [6, 32]. One should remember that the notion of

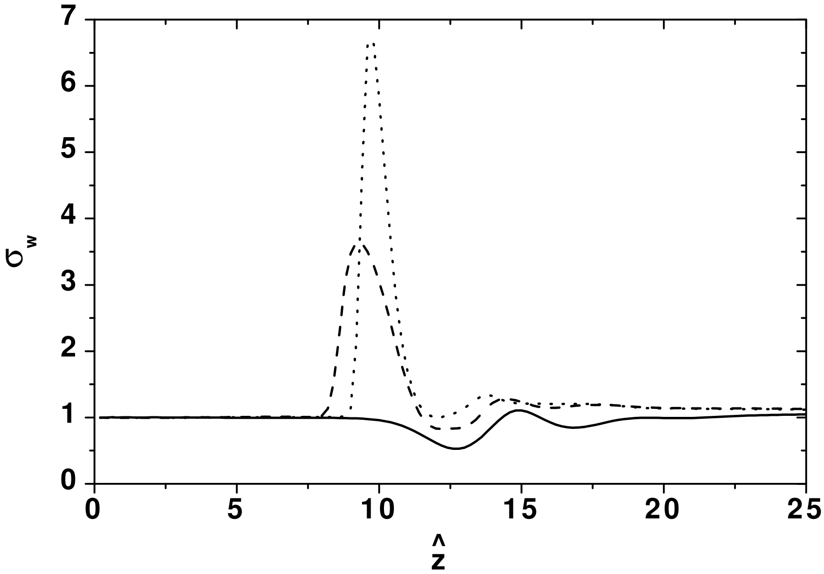


Fig. 11: Normalized r.m.s. deviation of fluctuations of instantaneous radiation power as a function of normalized undulator length. Solid, dashed, and dotted lines correspond to fundamental, third, and fifth harmonics, respectively. Calculations made with simulation code FAST [77].

the instantaneous intensity refers to a certain moment in time, and that the analysis must be performed over an ensemble of pulses. Also, the energy in the radiation pulse E_{rad} should fluctuate in accordance with the gamma distribution of Eq. (3) [6, 32]. These properties are well known in statistical optics as properties of completely chaotic polarized radiation [74].

As discussed in Section 2, the statistics of the high-harmonic radiation from the SASE FEL change significantly with respect to the fundamental harmonic (e.g., with respect to Gaussian statistics), The probability density function of the instantaneous intensity of higher harmonics SASE radiation (Eq. (4)) is a non-linear transformation of Eq. (2). Using this distribution, we obtain the expression for the mean value: $\langle z \rangle = h! \langle W \rangle^h$. Thus, the hth-harmonic radiation for the SASE FEL has an intensity level roughly $h!$ times larger than the corresponding steady-state case, but with more shot-to-shot fluctuations than the fundamental [56]. The non-trivial behaviour of the intensity of the high harmonics reflects the complicated non-linear transformation of the fundamental harmonic statistics. One can see that Gaussian statistics are no longer valid. The upper plots in Fig. 12 give an illustration to these consideration. Although in our practical example we do not have a pure linear amplification regime, the probability density functions for the instantaneous power follow the prediction of Eq. (4) rather well.

Analysis of the probability distributions in Fig. 12 shows that, in the non-linear regime, near the saturation point, the distributions change significantly with respect to the linear regime for both the fundamental and the third harmonic. An important message is that at the saturation point the third harmonic radiation exhibits much more noisy behaviour (nearly negative exponential) while stabilization of the fluctuations of the fundamental harmonics takes place.

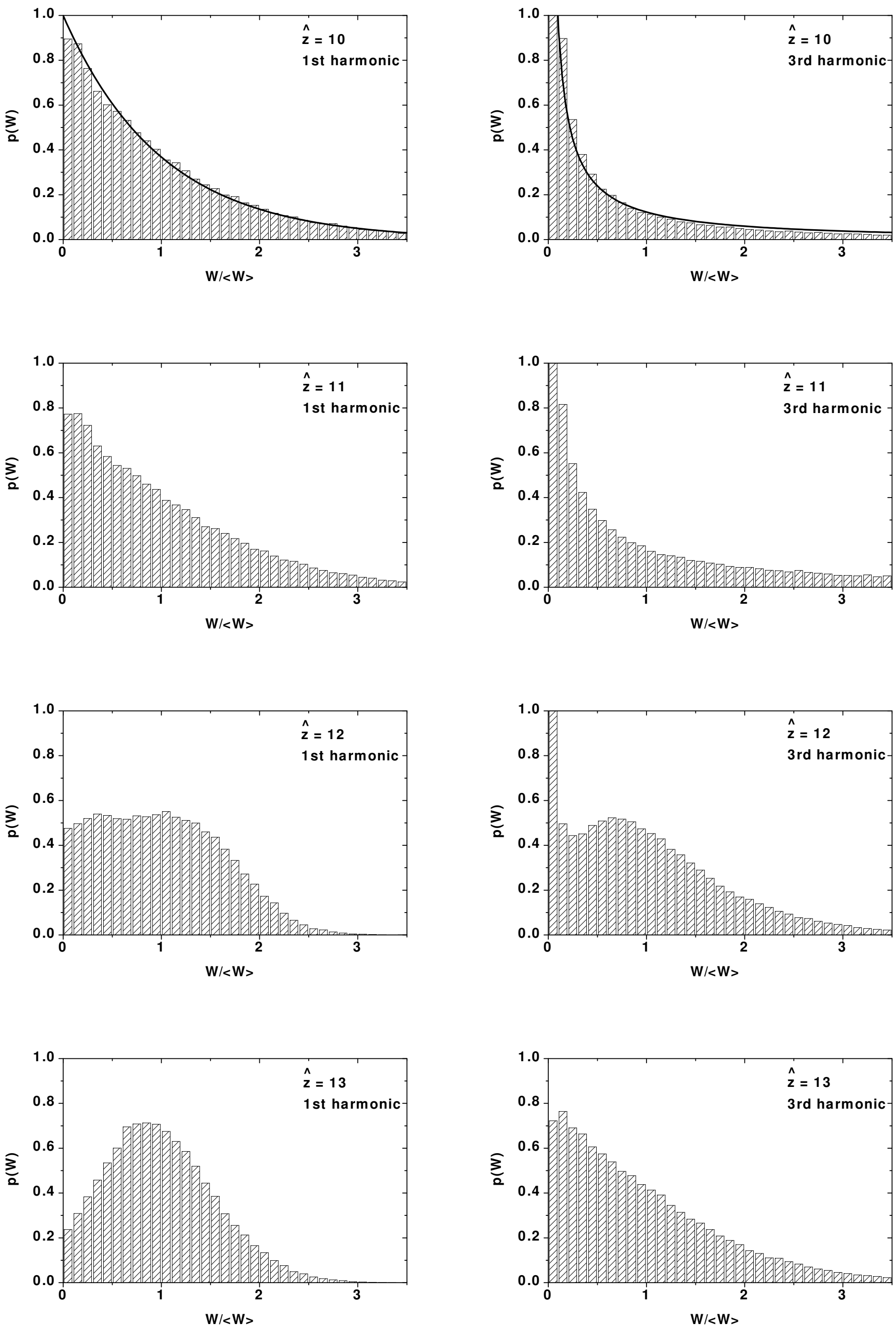


Fig. 12: Probability distribution of instantaneous radiation power at different lengths of the FEL amplifier $\hat{z} = 10$–13. Left and right columns correspond to fundamental and 3rd harmonic, respectively. Solid line shows probability density function (Eq. (4)). Calculations made with simulation code FAST [77].

3.1.3 Correlation functions

The first- and second-order time correlation functions are defined as:

$$g_1(t-t') = \frac{\langle \tilde{E}(t)\tilde{E}^*(t')\rangle}{\left[\langle|\tilde{E}(t)|^2\rangle\langle|\tilde{E}(t')|^2\rangle\right]^{1/2}}\ , \qquad g_2(t-t') = \frac{\langle|\tilde{E}(t)|^2|\tilde{E}(t')|^2\rangle}{\langle|\tilde{E}(t)|^2\rangle\langle|\tilde{E}(t')|^2\rangle}\ . \tag{16}$$

In Fig. 13 we show the evolution of the time correlation functions of the first and second order. At each normalized position along the undulator, $\hat{z}$, they are plotted against the normalized variable $\hat{\tau} = \rho\omega_0(t-t')$. The upper plot in Fig. 13 corresponds to the linear stage of SASE FEL operation. In the case of the fundamental harmonic, we deal with a Gaussian random process and the relation between the correlation functions holds for $g_2(t-t') = 1 + |g_1(t-t')|^2$. This feature does not hold for higher harmonics. The non-trivial behaviour of the correlation functions reflects the complicated non-linear evolution of the SASE FEL process. The second-order correlation function of the zero argument, $g_2(0)$, takes values smaller or larger than two, but always larger than unity. Note that there is a simple relation between $g_2(0)$ and the normalized r.m.s. power deviation: $g_2(0) = 1 + \sigma_{\mathrm{w}}^2$ (see Fig. 11). It is a well-known result of statistical optics that the cases of $g_2(0) = 1$ and $g_2(0) = 2$ correspond to stabilized single-mode laser radiation and to completely chaotic radiation from a thermal source, respectively. The values of $g_2(0)$ between 1 and 2 belong to some intermediate situation. In classical optics, a radiation source with $g_2(0) < 1$ cannot exist but the case of $g_2(0) > 2$ is possible. As one can see from Fig. 13, the latter phenomenon (known as superbunching) occurs for higher harmonics of SASE FEL, or for the fundamental when the SASE FEL operating in the non-linear regime.

Figure 14 shows the dependence on the undulator length of the normalized coherence time $\hat{\tau}_{\mathrm{c}} = \rho\omega_0\tau_{\mathrm{c}}$, where τ_{c} is given by Eq. (6). For the fundamental harmonic, the coherence time achieves its maximal value near the saturation point and then decreases drastically. The maximum value of $\hat{\tau}_{\mathrm{c}}$ depends on the saturation length and, therefore, on the value of the parameter N_{c}. With logarithmic accuracy, we have the following expression for the coherence time of the fundamental harmonic:

$$(\hat{\tau}_{\mathrm{c}})_{\max} \simeq \sqrt{\frac{\pi \ln N_{\mathrm{c}}}{18}}\ . \tag{17}$$

Longitudinal coherence for higher harmonics evolves in three different stages. Initially (up to $\hat{z} = 7$–8, see Fig. 14) coherence time increases linearly, as it should, for the spontaneous emission of radiation from the undulator. When the process of non-linear harmonic generation starts to dominate over the spontaneous emission, the coherence time drops sharply. At positions around $\hat{z} = 10$–11, we obtain some plateau where the ratio of the coherence time of the hth harmonic to that of the first harmonic scales as $1/\sqrt{h}$. At these distances, the SASE FEL still operates in the exponential regime when the amplitude of the beam bunching is visibly less than unity, and the intensity of the hth harmonic scales as I_1^h. Such a mechanism of non-linear harmonic generation leads to scaling of the coherence time as $1/\sqrt{h}$. To explain this, we refer to Fig. 6, which presents the temporal structure of the radiation pulse. The radiation pulse consists of a number of spikes (wavepackets). For the sake of simplicity, let us approximate a wavepacket with a Gaussian, $G_1(\hat{s})$, with an r.m.s. width σ_1. Non-linear transformation of the intensity for the hth harmonic gives us envelope $G_h \propto G_1^h$. Therefore, the relevant spike for the hth harmonic is $\sigma_h = \sigma_1/\sqrt{h}$. In other words, sharpening of the peaks in intensity distribution leads to suppression of coherence times for higher harmonics. When the amplification process enters the non-linear stage ($\hat{z} \gtrsim 11$), the relative sharpening of the intensity peaks of higher harmonics becomes stronger, and coherence time starts decrease again. In fact, as one can find from Fig. 13, the coherence time at the saturation point ($\hat{z} = 13$) for higher harmonics approximately decreases inversely proportional to the harmonic number h.

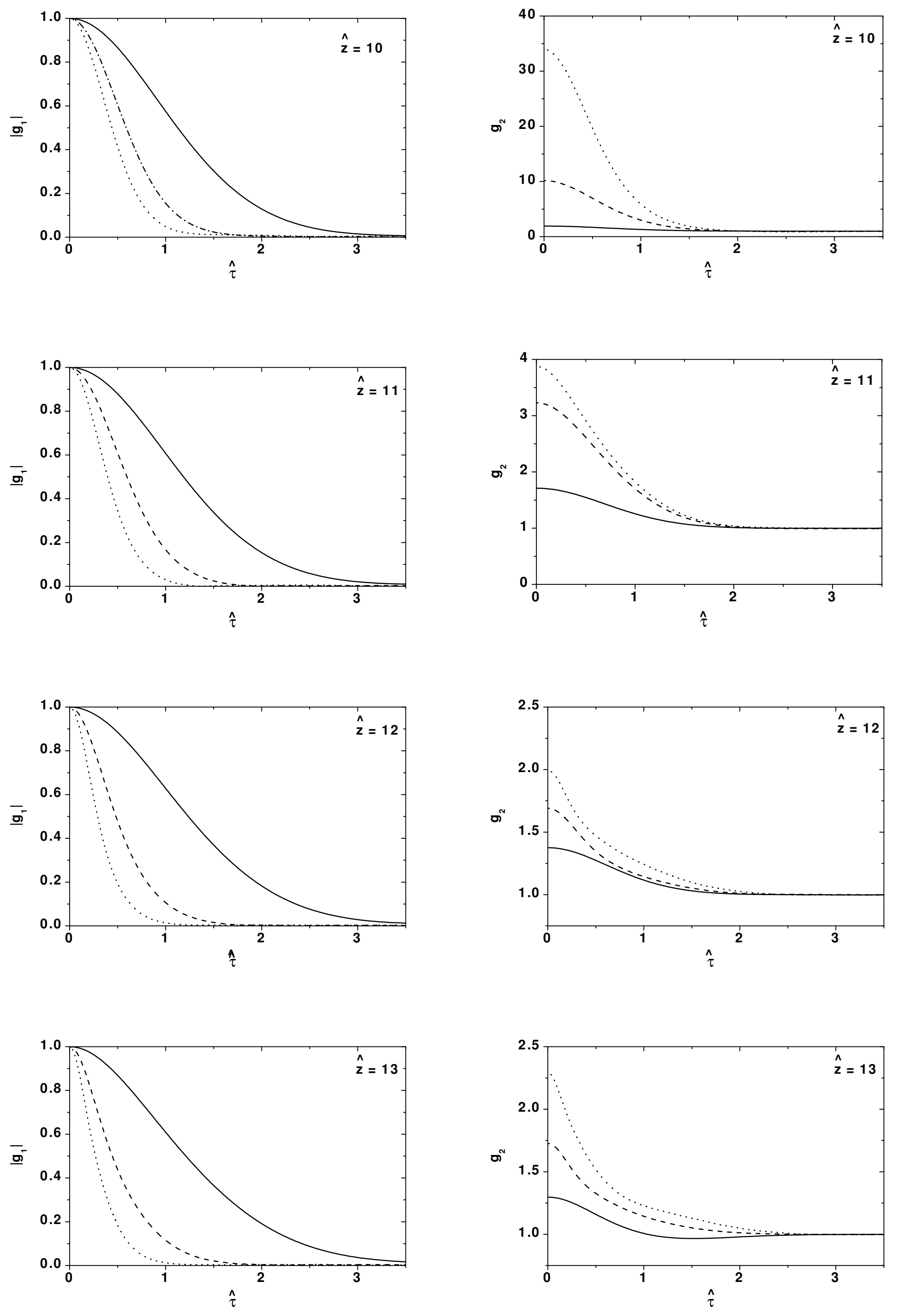


Fig. 13: First-order (left column) and second-order (right column) correlation function at different lengths of the FEL amplifier $\hat{z} = 10$–13. Solid, dashed, and dotted lines correspond to fundamental, third, and fifth harmonics, respectively. Calculations made with simulation code FAST [77].

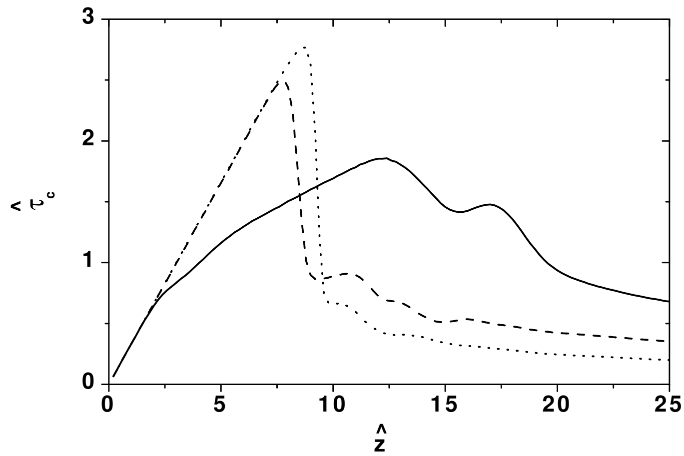


Fig. 14: Normalized coherence time of SASE FEL as a function of normalized undulator length. Solid, dashed, and dotted lines correspond to fundamental, third, and fifth harmonics, respectively

When comparing radiation spectra, it is convenient to use the normalized spectral density, $H(\hat{C})$, defined as

$$\int_{-\infty}^{\infty} \mathrm{d}\hat{C} H(\hat{C}) = 1 \ .$$

Here $\hat{C} = [2\pi/\lambda_\mathrm{w} - \omega(1 + K^2)/(2c\gamma^2)]/\Gamma$ is the detuning parameter. The frequency deviation, $\Delta\omega$, from the nominal value of ω_h can be recalculated as $\Delta\omega = -2\rho\omega_h\hat{C}$. Since we consider the model of a long rectangular bunch, the function $H(\hat{C})$ can be treated as the normalized spectral density of both the radiation energy and the power.

The spectral density of the radiation energy and the first-order time correlation function form a Fourier transform pair [74]:

$$G(\Delta\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\tau g_1(\tau) \exp(-\mathrm{i}\Delta\omega\tau) \ . \tag{18}$$

This is the so-called Wiener-Khinchin theorem.

The temporal structures of the radiation pulses (see Fig. 6) are used to calculate the first-order time correlation function (see Fig. 13). Then the radiation spectra are reconstructed by Fourier transformation of the first-order time correlation function. Figure 15 shows the evolution of the radiation spectra of the SASE FEL radiation from the end of the linear regime to saturation. Note that the spectrum width of the higher harmonics from SASE FEL differs significantly from that of incoherent radiation. For the case of incoherent radiation, the relative spectrum width $\Delta\omega/\omega_h$ scales inversely proportional to the harmonic number h [79]. One can see that the situation changes dramatically for the case when the non-linear harmonic generation process starts to dominate. At saturation, we find that the relative spectrum bandwidth is nearly the same for all odd harmonics.

3.2 Short-pulse effects in SASE FEL

Up to now we have studied the properties of the SASE FEL radiation in the framework of a stationary process, i.e. we considered the model of a long electron bunch with a rectangular profile. In this section,

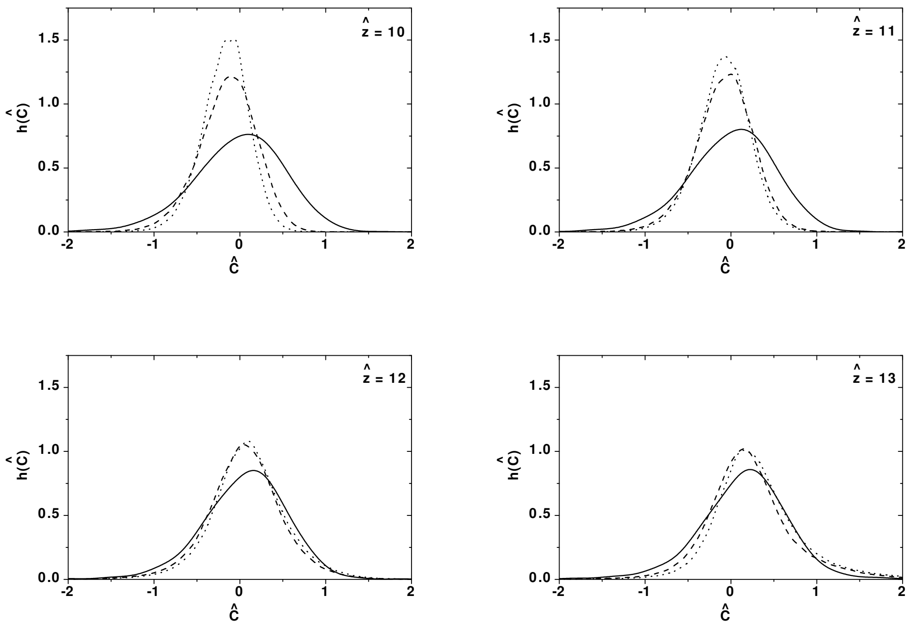


Fig. 15: Normalized spectrum at different lengths of the undulator: $\hat{z} = 10$–13. Solid, dashed, and dotted lines correspond to fundamental, third, and fifth harmonics, respectively. Calculations made with simulation code FAST [77].

we analyse the statistical properties of the radiation from a SASE FEL driven by short electron bunches. In the linear regime, the radiation from a SASE FEL is a Gaussian random process. When approaching saturation point, the statistical properties of the radiation change drastically on a scale of one field gain length. We devote particular attention to the analysis of fluctuations of the total energy in the radiation pulse and after a narrow-band monochromator. We show that slippage effects result in a set of novel features of SASE FEL operating in the non-linear regime. In particular, for very short pulses, we observe the effect of stabilization of fluctuations of energy in the radiation pulse and fluctuations of the energy after the narrow-band monochromator. The suppression factor scales as a square root of the pulse length.

To be specific, we consider an electron beam with a Gaussian axial profile of the current density:

$$S(\hat{s}) = \frac{j(\hat{s})}{j_{\max}} = \exp\left(-\frac{\hat{s}^2}{2\hat{\sigma}_{\mathrm{b}}^2}\right) \ ,$$

where $\hat{\sigma}_{\mathrm{b}} = \rho\omega_0\sigma_{\mathrm{b}}/c$ and σ_{b} is the r.m.s. bunch length. Here and in the following, the normalization is performed with respect to the maximum current density, $j_{\max}$. The r.m.s. bunch length is assumed to be large, $\omega_0\sigma_{\mathrm{b}}/c \gg 1$, or, in normalized form: $\hat{\sigma}_{\mathrm{b}} \gg \rho$. Under this assumption, we can neglect the contribution of the coherent seed to the input signal of the FEL amplifier starting from the shot noise. Since ρ is always much less than unity, we can investigate short-pulse effects, when the bunch is comparable to (or even much shorter than) the typical slippage distance $c/(\rho\omega_0)$.

Even without simulations, we can predict the general properties of the radiation from a SASE FEL operating in the linear regime. Indeed, shot noise in the electron beam is a Gaussian random process [32]. The FEL amplifier, operating in the linear regime, can be considered as a linear filter that does not change statistics. As a result, the radiation is also a Gaussian random process. In this case, the probability distribution of the instantaneous radiation power should be the negative exponential distribution (the notion of instantaneous power refers to a certain moment of time and a certain z coordinate, and that the analysis must be performed over an ensemble of pulses). Also, the finite-time integrals of the instantaneous power and the integrated spectral density (measured after the monochromator) should fluctuate in accordance with the gamma distribution. Nevertheless, a reasonable question arises as to what are the features of the radiation from SASE FEL operating in the non-linear mode and, in particular, at saturation and post-saturation regime.

In the case of a SASE FEL driven by a short electron bunch, we deal with non-stationary random process, and temporal coherence can be described in terms of an effective correlation function [6]:

$$g_1^{(\mathrm{eff})}(\tau) = \frac{\int\limits_{-\infty}^{\infty} \langle \tilde{E}(\bar{t}+\tau/2)\tilde{E}^*(\bar{t}-\tau/2)\rangle \mathrm{d}\bar{t}}{\int\limits_{-\infty}^{\infty} \langle |\tilde{E}(\bar{t})|^2\rangle \mathrm{d}\bar{t}} , \qquad (19)$$

where $\bar{t} = (t+t')/2$ and $\tau = t - t'$.

The normalized envelope of the radiation power spectrum and the effective correlation function are connected by a Fourier transform [6]:

$$\frac{\langle |\bar{E}(\Delta\omega)|^2\rangle}{\int\limits_{-\infty}^{\infty} \langle |\bar{E}(\Delta\omega)|^2\rangle \mathrm{d}(\Delta\omega)} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{d}\tau g_1^{(\mathrm{eff})}(\tau) \mathrm{e}^{\mathrm{i}\Delta\omega\tau} . \qquad (20)$$

In other words, the correlation function $g_1^{(\mathrm{eff})}(\tau)$ effectively describes the case of a long rectangular bunch producing the same spectrum as that of a bunch with a gradient profile.

It is natural to define the coherence time for the non-stationary process as

$$\tau_{\mathrm{c}} = \int\limits_{-\infty}^{\infty} \mathrm{d}\tau |g_1^{(\mathrm{eff})}(\tau)|^2 . \qquad (21)$$

To obtain the output characteristics of the radiation from SASE FEL, one should perform a large number of simulation runs with a time-dependent simulation code. The result of each run contains parameters of the output radiation (field and phase) stored in the boxes over the full length of the radiation pulse. At the next stage of the numerical experiment, the data arrays should be handled to extract information on statistical properties of the radiation. Probability distribution functions of the instantaneous radiation power, of the finite-time integrals of the instantaneous power, and of the radiation energy after the monochromator installed at the exit of the SASE FEL are calculated by plotting histograms of large amounts of statistical data.

3.2.1 Average energy and fluctuations

The time structure of a sample radiation pulse and the radiation power averaged over an ensemble are presented in Fig. 16 for different bunch lengths. It can be seen that, for values of $\hat{\sigma}_{\mathrm{b}}$ of about unity, the radiation pulse is visibly shifted from the centre of the electron bunch, owing to the slippage effect. In Fig. 17, the number of modes M is plotted against the longitudinal coordinate for different bunch lengths. The SASE FEL operates in the linear regime, and the number of modes is $M = 1/\sigma_{\mathcal{E}}^2$, where $\sigma_{\mathcal{E}}^2 = \langle(\mathcal{E} - \langle\mathcal{E}\rangle)^2\rangle/\langle\mathcal{E}\rangle^2$ is the relative energy dispersion in the radiation pulse.

Figure 18 shows the evolution of the averaged efficiency along the undulator length. The averaged efficiency is defined as $\langle\hat{\eta}\rangle = \langle E_{\mathrm{rad}}\rangle/(\rho\gamma m_{\mathrm{e}}c^2 N)$, where E_{rad} is energy in the radiation pulse, and N is the number of electrons in the bunch. The dashed line in Fig. 18 represents the averaged efficiency for the case of a long electron bunch with a rectangular profile. The amplification process involves three stages: start-up from shot noise, a stage of exponential gain, and a non-linear stage. Let us define the saturation

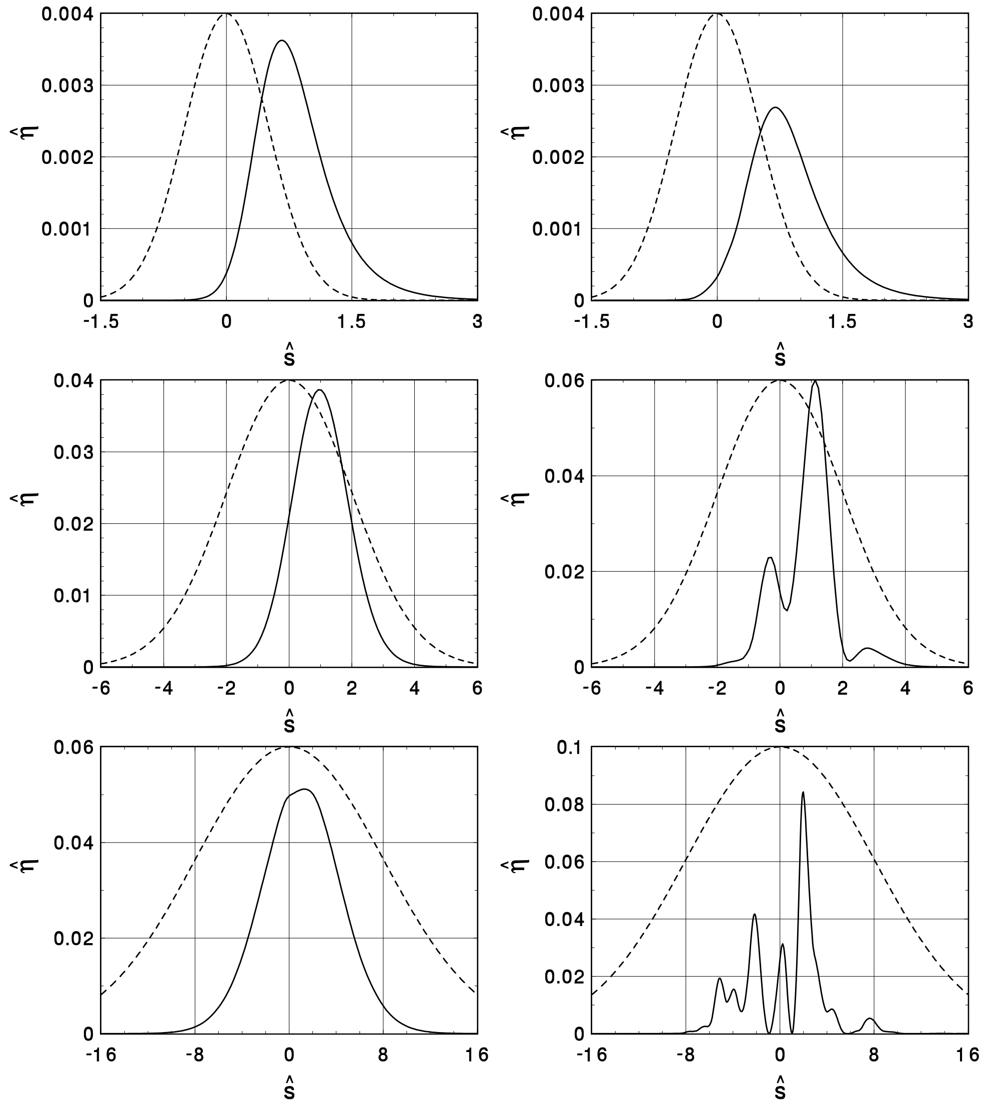


Fig. 16: Averaged (left column) and typical single-shot (right column) axial distribution over $\hat{s} = \rho\omega_0(z/\bar{v}_z - t))$ of normalized radiation power from SASE FEL for different r.m.s. bunch lengths of $\hat{\sigma}_\mathrm{b} = 0.5$, 2, and 8 (upper, middle, and lower plots, respectively). Dashed lines represent axial profile of the beam current. The normalized length of the undulator is $\hat{z} = 10$. Calculations made with simulation code FAST [77].

point as the first maximum of the gain curve. At $\hat{\sigma}_\mathrm{b} \lesssim 2$, the saturation length increases as $\hat{z}_\mathrm{sat} \propto 1/\sqrt{\hat{\sigma}_\mathrm{b}}$, and the averaged saturation efficiency decreases as $\langle\hat{\eta}\rangle \propto \sqrt{\hat{\sigma}_\mathrm{b}}$. These features of short bunch effects were addressed in an earlier paper [29]. Figure 19 shows the dependence of the saturation efficiency on the bunch length. It can be seen that the saturation efficiency quickly approaches an asymptotical value. A comparison with the case of a long electron pulse with a rectangular profile (the dashed line in the plot) shows that only 60% of electrons produce radiation. This is a consequence of the gradient profile of the electron bunch.

Figure 20 (left plot) shows the evolution along the undulator of the radiation pulse energy, fluctuations of the radiation pulse energy, and the r.m.s. photon pulse length. A maximum of the fluctuations of the radiation pulse energy and a minimum radiation pulse duration are obtained at the end of the exponential gain regime. Normalized values of these parameters (E/E_sat, $\sigma_\mathrm{E}/\sigma_\mathrm{E}^\mathrm{max}$, and $\sigma_\mathrm{ph}/\sigma_\mathrm{ph}^\mathrm{min}$) exhibit nearly universal dependencies for the r.m.s. electron pulse duration $\rho\omega\sigma_\mathrm{z} \gtrsim 1$, as shown in the right plot in Fig. 20. This allows us to derive the universal dependency of the r.m.s. electron pulse length

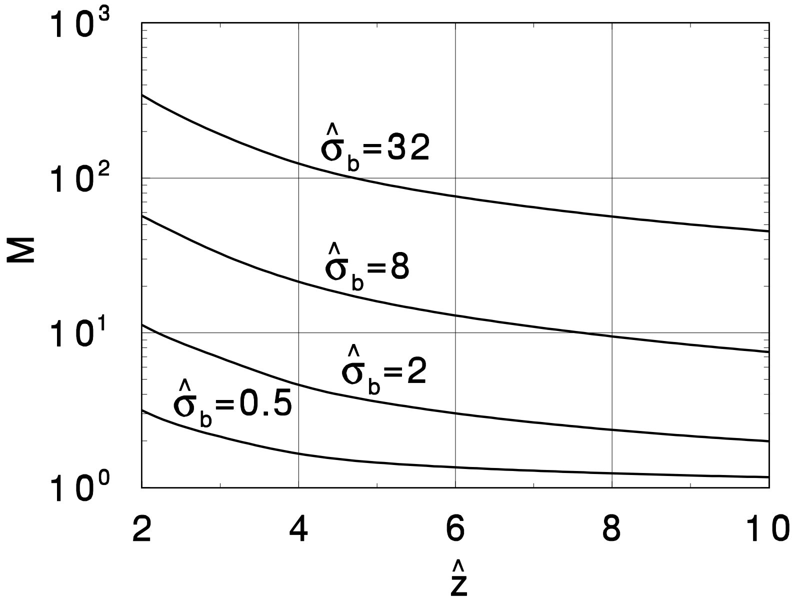


Fig. 17: Number of modes in the radiation pulse for Gaussian electron bunches of different lengths versus undulator length: SASE FEL operates in linear regime. Calculations were made with simulation code FAST [77].

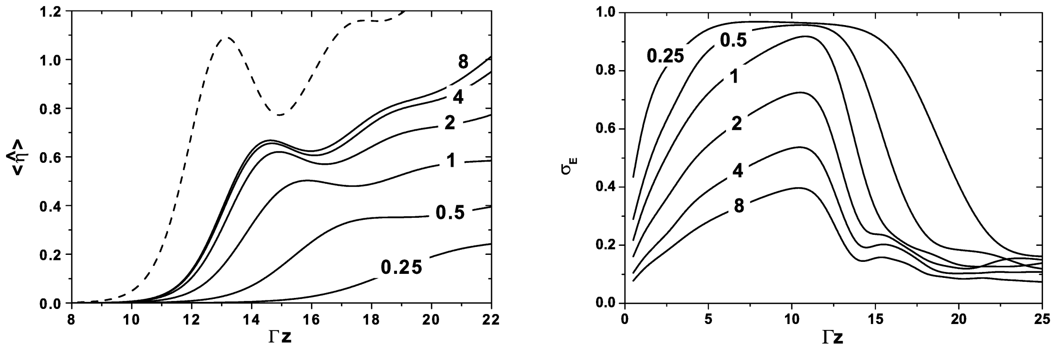


Fig. 18: Left: Averaged efficiency of SASE FEL versus undulator length $\hat{z} = \Gamma z$ for different lengths of electron bunch $\hat{\sigma}_{\rm b} = 0.25$-8. Right: normalized r.m.s. deviation of energy in radiation pulse; $N_{\rm c} = 10^8$. Dashed line represents case of long electron pulse with rectangular profile. Calculations made with simulation code FAST [77].

and the minimum FWHM radiation pulse length $\tau_{\rm ph}^{\rm min} = \sqrt{2\pi}\sigma_{\rm ph}^{\rm min}$ at the end of the linear regime as a function of the number of modes in the radiation pulse (see Fig. 21). For $M \gtrsim 2$ we have, with reasonable practical accuracy:

$$\sigma_z \simeq \tau_{\rm ph}^{\rm min} \simeq \frac{M\lambda}{5\rho} \simeq \frac{M\lambda L_{\rm sat}}{5c\lambda_{\rm w}} \, . \tag{22}$$

The minimum radiation pulse duration expressed in terms of the coherence time is $\tau_{\rm ph}^{\rm min} \simeq 0.7 \times M \times \tau_{\rm c}$.

Lengthening of the radiation pulse occurs when the amplification process enters the saturation regime. This happens because of two effects. The first effect is lasing to saturation of the tails of the electron bunch, and the second effect is pulse lengthening due to slippage effects (just simply one radiation wavelength per one undulator period). The effect of lasing tails gives the same relative radiation pulse lengthening as is illustrated in the bottom plot in Fig. 20. At the saturation point, pulse lengthening is about a factor of 1.4 with respect to the minimum pulse for the linear regime given by Eq. (22), and is increased up to a factor of two in the deep non-linear regime. It can also be seen that the slippage effect is more pronounced for relative lengthening of short pulses.

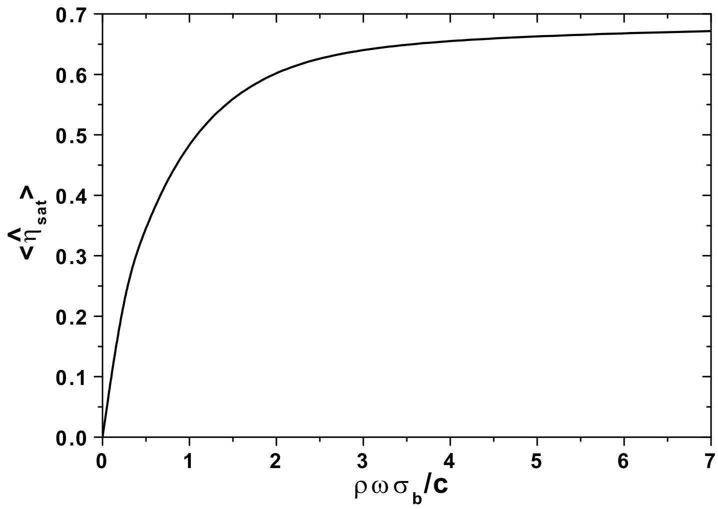


Fig. 19: Averaged efficiency of SASE FEL at saturation versus length of electron bunch. Calculations made with simulation code FAST [77].

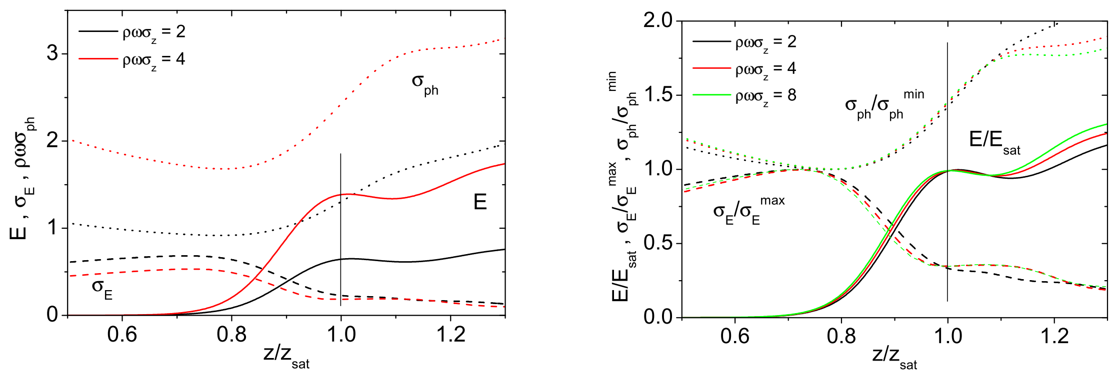


Fig. 20: Evolution of the energy in the radiation pulse E (solid line), fluctuations of the radiation energy σ_{E} (dashed line), and r.m.s. radiation pulse duration $\rho\omega\sigma_{\mathrm{ph}}$ (dotted line) versus undulator length. Black, red, and green lines correspond to electron r.m.s. pulse durations $\rho\omega\sigma_{\mathrm{z}}$ of 2, 4, and 8, respectively. Values on the right plot are normalized as E/E_{sat}, $\sigma_{\mathrm{E}}/\sigma_{\mathrm{E}}^{\mathrm{max}}$, and $\sigma_{\mathrm{ph}}/\sigma_{\mathrm{ph}}^{\mathrm{min}}$.

3.2.2 Correlation functions and coherence time

The effective correlation function is calculated as follows. We perform a large number of simulation runs, calculate the spectrum of each pulse, and calculate the envelope of the normalized averaged spectrum. Equation (20) is used to calculate the first-order correlation function. The plot of the effective correlation function $g_1(\tau)$ at saturation is shown in Fig. 22. The circles in this plot represent the correlation function for the case of a long electron pulse with a rectangular profile [32]. Figure 23 shows plots for the coherence time $\hat{\tau}_{\mathrm{c}}$ (Eq. (21)) for different electron pulse lengths. Analysis of these results allows us to state that for the bunch length $\hat{\sigma}_{\mathrm{b}} \gtrsim 4$ we have good agreement between the case of finite pulse duration and asymptotic results for a long pulse with a rectangular profile. Conversely, at short bunch length, $\hat{\sigma}_{\mathrm{b}} \lesssim 2$, the behaviour of the coherence time differs visibly from an asymptotic one. This is a clear indication of a different physical process.

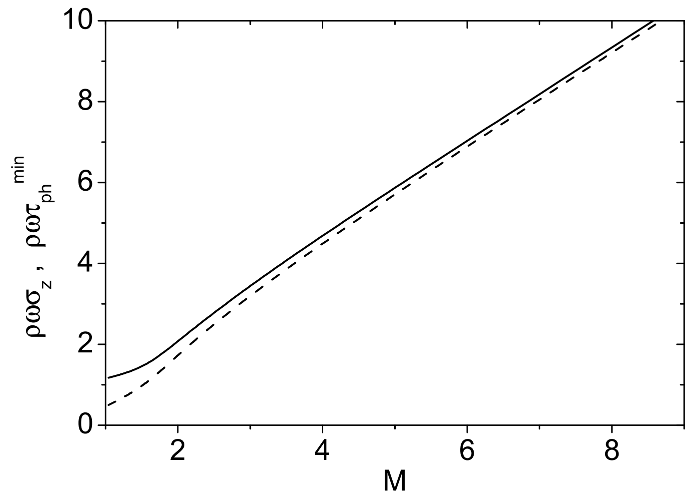


Fig. 21: The r.m.s. electron pulse duration σ_{z} (dashed curve) and minimum FWHM photon pulse duration $\tau_{\mathrm{ph}}^{\mathrm{min}}$ at the end of the linear regime (solid curve) versus the number of modes in the radiation pulse M.

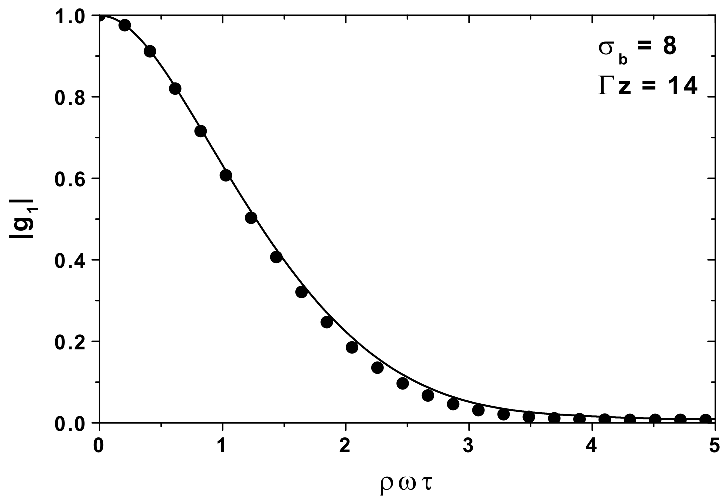


Fig. 22: Module of first-order time correlation function, where $\hat{\sigma}_{\mathrm{b}} = 8$ and $N_{\mathrm{c}} = 10^8$; SASE FEL operates at saturation, $\hat{z} = 14$. Circles correspond to the case of a long electron bunch with a rectangular profile. Calculations made with simulation code FAST [77].

3.2.3 Fluctuations of energy in the radiation pulse

The right plot in Fig. 18 shows the normalized r.m.s. deviation of energy in the radiation pulse. There is a straightforward explanation for the behaviour of the fluctuations in the linear regime: the number of longitudinal modes decreases with the pulse length and the undulator length. The radiation of the SASE FEL operating in the linear regime is a Gaussian random process, so the probability distribution of the energy in the radiation pulse is a gamma distribution. The situation changes drastically when the amplification process enters a non-linear stage. It is seen that deviation of energy drops quickly on a scale of a field gain length. For a long pulse, fluctuations are suppressed as $1/\sqrt{\hat{\sigma}_{\mathrm{b}}}$, which is a consequence of increasing the number of statistically independent spikes in the radiation pulse [29]. The physical

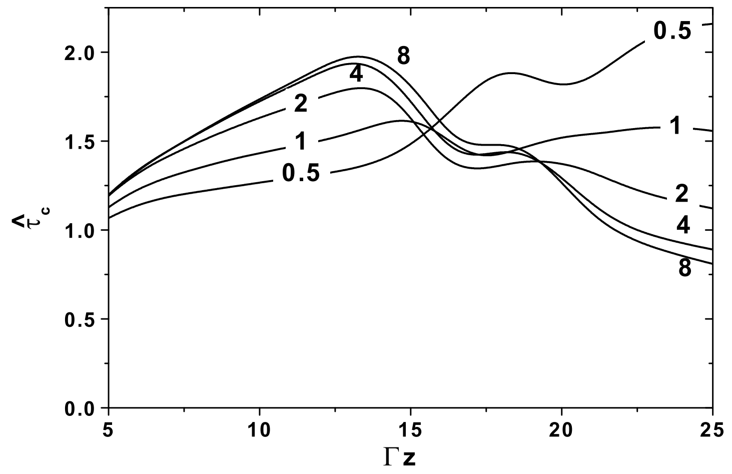


Fig. 23: Normalized coherence time $\hat{\tau}_c$ versus undulator length. Electron bunch length varies within the limits $\hat{\sigma}_b = 0.5$–8; $N_c = 10^8$. Calculations made with simulation code FAST [77].

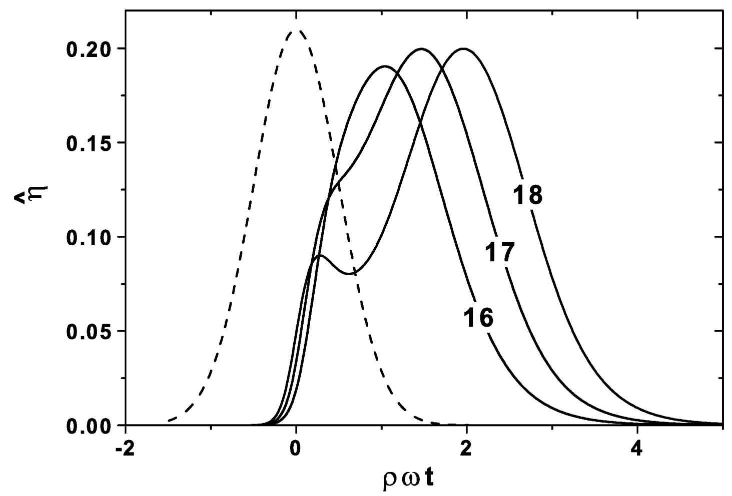


Fig. 24: Averaged normalized radiation power versus time; SASE FEL operates in the non-linear regime. Undulator lengths are $\hat{z} = 16$, 17, and 18. Dashed line represents electron bunch profile ($\hat{\sigma}_b = 0.5$). Calculations made with simulation code FAST [77].

picture becomes quite different for the bunch length $\hat{\sigma}_b \lesssim 2$: the deviation of energy in the radiation pulse starts to decrease with the bunch length as $\sqrt{\hat{\sigma}_b}$. The nature of this phenomenon can be understood by analysing the structure of the radiation pulse (see Fig. 24). At the end of the linear mode of operation, a SASE FEL driven by a short electron bunch produces radiation pulses of nearly the same shape, but with amplitudes fluctuating by almost negative exponential distribution. When the amplification process enters the non-linear stage, the radiation power is saturated, and pulses sleep forward. A further increase in the total energy occurs, due to the radiation of bunched electron beam. Since maximal bunching of the electron beam is limited to unity, this additional radiation is well stabilized, leading to overall stability of the total energy in the radiation pulse.

Simulations show that the statistics of the radiation also change drastically near the saturation point, on a scale of one field gain length. Figure 25 shows the evolution of the probability density

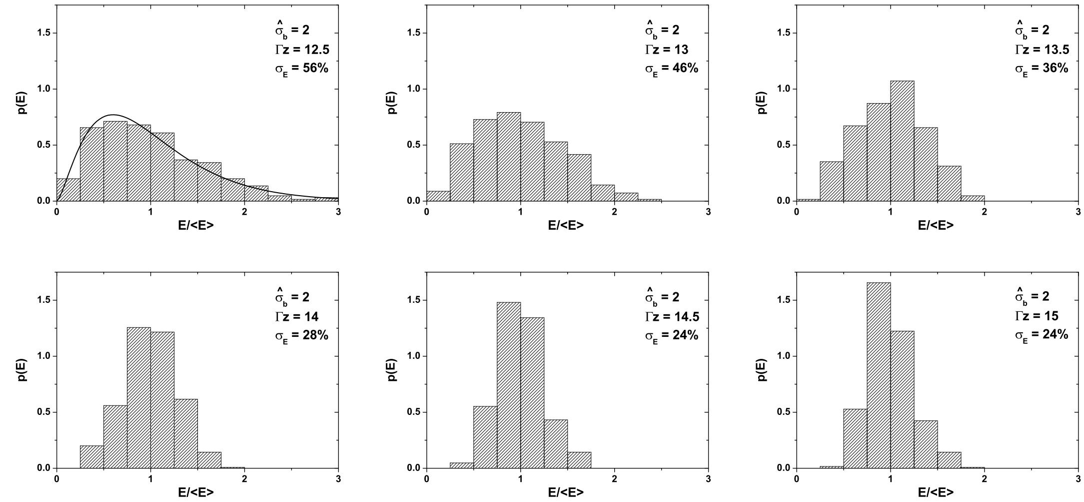


Fig. 25: Probability density distribution of energy in radiation pulse for different undulator lengths; electron pulse length, $\hat{\sigma}_{\mathrm{b}} = 2$. Solid curve represents gamma distribution. Calculations made with simulation code FAST [77].

distribution of energy in the radiation pulse for different lengths of the undulator. The electron pulse length is $\hat{\sigma}_{\mathrm{b}} = 2$. The solid curve represents a gamma distribution. Such behaviour is also typical for shorter bunch lengths.

3.2.4 Fluctuations of energy after narrow-band monochromator

Let us study the statistics of SASE FEL radiation filtered through a narrow-band monochromator. Plots of the normalized r.m.s. deviation of energy after a narrow-band monochromator versus undulator length are presented in Fig. 26. In the linear stage of SASE FEL operation, the value of normalized energy deviation is equal to unity, and energy fluctuates in accordance with a negative exponential distribution. This is a consequence of the fact that, in this case, radiation is a Gaussian random process. However, in the non-linear mode of operation, we obtain a significant decrease in fluctuations when the pulse length decreases (see Fig. 27). This effect has a simple physical explanation. When a SASE FEL driven by a short bunch operates in the linear regime, radiation pulses have a similar shape each time, but the amplitude fluctuates by an almost negative exponential distribution (see left plot in Fig. 16). When the amplification process enters the non-linear stage, amplitudes of different pulses are equalized, owing to saturation effects, while keeping the close shape (see right plot in Fig. 28). The spectrum of the radiation pulse is given by the Fourier transform of the radiation field, and at saturation we obtain a nearly similar spectrum envelope for different pulses (see right plot in Fig. 29). As a result, we can expect that fluctuations of the radiation energy after a narrow-band monochromator should follow fluctuations of the total energy in the radiation pulse. Figure 30 confirms this simple physical consideration. At saturation, fluctuations of the energy after the narrow-band monochromator decrease as $\sigma_{\mathrm{M}} \simeq \sigma_{\mathrm{E}} \propto \sqrt{\hat{\sigma}_{\mathrm{b}}}$.

Study of the amplification process in the SASE FEL driven by an electron bunch of finite pulse duration allows us to make the following conclusions. For the bunch length $\hat{\sigma}_{\mathrm{b}} \gtrsim 4$, asymptotical results [32] for a long rectangular bunch are applicable. At $\hat{\sigma}_{\mathrm{b}} \lesssim 2$, the SASE FEL exhibits quite different behaviour, caused by the strong influence of slippage effects. In addition to a reduction in the FEL gain and efficiency [29], short-pulse effects strongly influence the statistical properties of the radiation in the non-linear regime. In particular, for very short pulses, we found the effect of stabilization of fluctuations of energy in the radiation pulse and fluctuations of the energy after the narrow-band monochromator. The suppression factor scales as $\sqrt{\hat{\sigma}_{\mathrm{b}}}$ with electron bunch length. This effect has been measured experimentally at the TTF FEL [46].

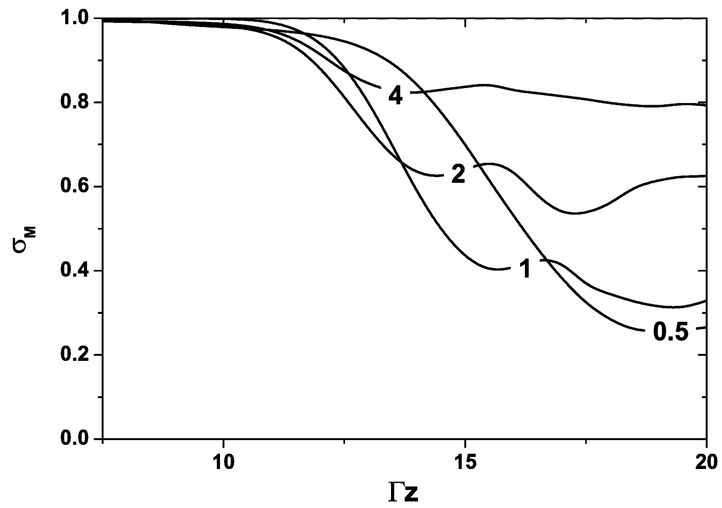


Fig. 26: Normalized r.m.s. deviation of energy after narrow-band monochromator versus undulator length. Electron bunch length changes within the limits $\hat{\sigma}_{\mathrm{b}} = 0.5$–$4$. $N_{\mathrm{c}} = 10^8$. Calculations made with simulation code FAST [77].

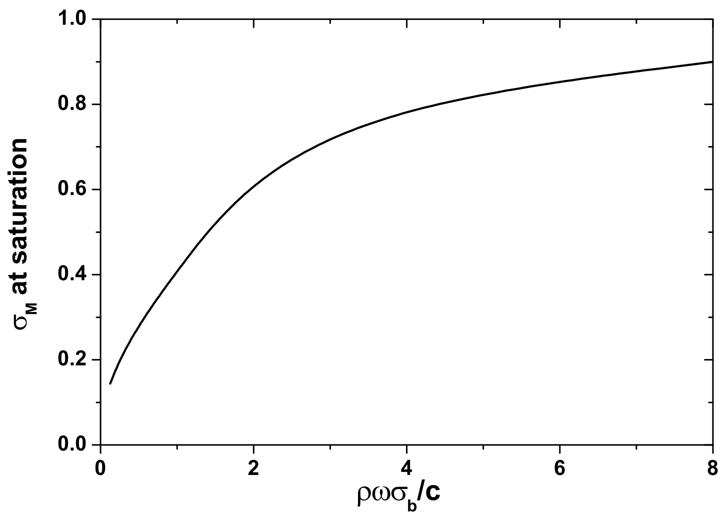


Fig. 27: Normalized r.m.s. deviation of energy after narrow-band monochromator versus length of electron bunch; SASE FEL operates at saturation. Calculations made with simulation code FAST [77].

4 Transverse coherence

At the initial stage of amplification, the spatial coherence is poor, and the radiation consists of a large number of transverse modes. Longitudinal coherence is poor as well. In the exponential stage of amplification, transverse modes with higher gain dominate over modes with lower gain as the undulator length progresses. This feature is also known as the mode competition process. Longitudinal coherence is also improving in the high-gain linear regime. The mode selection process stops at the onset of the non-linear regime, and maximum values of the degree of the transverse coherence and of the coherence time are reached at this point. The undulator length to saturation is in the range of about nine (hard X-ray SASE FELs) to eleven (visible-range SASE FELs) field gain lengths [41]. The situation with

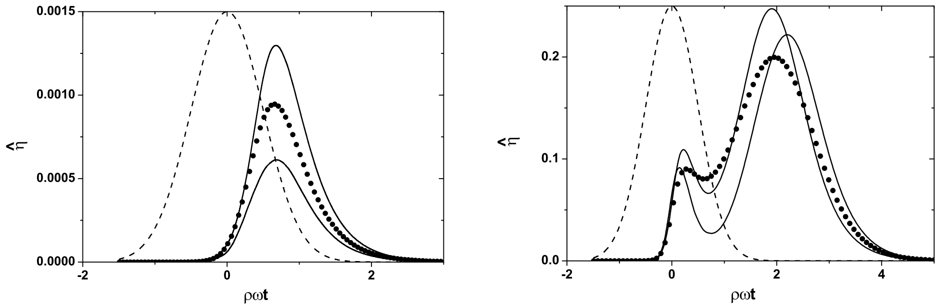


Fig. 28: Normalized power of SASE FEL radiation in (left) the linear regime and (right) at saturation. Solid curves are single pulses, and circles represent averaging over many pulses; dashed curve is electron bunch profile, $\hat{\sigma}_{\rm b} = 0.5$. Calculations made with simulation code FAST [77].

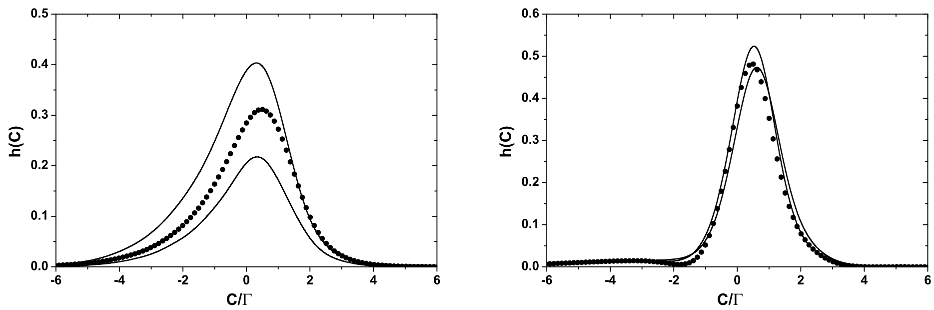


Fig. 29: Spectrum of SASE FEL radiation in (left) the linear regime and (right) at saturation. Solid curves are single pulses (see Fig. 16); circles represent averaging over many pulses. Electron bunch length, $\hat{\sigma}_{\rm b} = 0.5$. Calculations made with simulation code FAST [77].

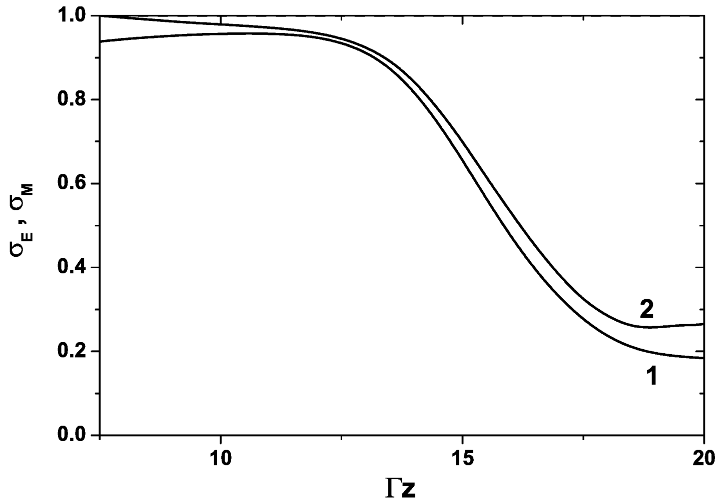


Fig. 30: Normalized r.m.s. deviation of (1) total energy in the radiation pulse and (2) energy after narrow-band monochromator versus undulator length. Length of electron bunch, $\hat{\sigma}_{\rm b} = 0.5$. Calculations made with simulation code FAST [77].

transverse coherence is favourable when the relative separation of the field gain between the fundamental and higher modes exceeds 25–30%. In this case, the maximum degree of transverse coherence can exceed 90% [39, 41]. Further development of the amplification process in the non-linear stage leads to visible degradation of the coherence properties.

Separation of the gain of the FEL radiation modes mainly depends on the value of the diffraction parameter. An increase in the value of the diffraction parameter results in less relative separation of the gain of the modes. In this case, we deal with the mode degeneration [6, 68]. Since the number of gain lengths to saturation is limited, the contribution of the higher spatial modes to the total power increases with the value of the diffraction parameter, and the transverse coherence degrades. The range of large diffraction parameter values is typical for SASE FELs operating in the hard X-ray wavelength range. It is also worth noticing that a spread of longitudinal velocities (due to energy spread and emittance) helps to suppress high-order modes, thus improving transverse coherence properties. This consideration suggests that a tight focusing of the electron beam in the undulator can be important for reaching a good coherence, owing to a reduction of the diffraction parameter and an increase of the velocity spread.

In this section, we present a thorough analysis of the coherence properties of the radiation from a SASE FEL. The analysis is performed in the framework of three-dimensional theory. We find that there is such a parameter range where the degree of transverse coherence of the radiation from SASE FEL is visibly less than unity. We also show that the pointing stability of the SASE FEL beam suffers from insufficient mode selection of higher spatial radiation modes, which happens for large values of the diffraction parameter.

4.1 FEL radiation modes

We consider an axisymmetrical model of the electron beam. It is assumed that the transverse distribution function of the electron beam is Gaussian, so that the r.m.s. transverse size of the matched beam is $\sigma = \sqrt{\epsilon\beta}$, where ϵ is the r.m.s. beam emittance and β is the beta function. In the framework of the three-dimensional theory, the operation of a short-wavelength FEL amplifier is described by the following parameters: the diffraction parameter B, the energy spread parameter $\hat{\Lambda}^2_{\mathrm{T}}$, the betatron motion parameter $\hat{k}_\beta$, and the detuning parameter $\hat{C}$ [6, 70]:

$$B = 2\bar{\Gamma}\sigma^2\omega/c\,, \qquad \hat{C} = C/\bar{\Gamma}\,,$$
$$\hat{k}_\beta = 1/(\beta\bar{\Gamma})\,, \qquad \hat{\Lambda}^2_{\mathrm{T}} = (\sigma_{\mathrm{E}}/E)^2/\bar{\rho}^2\,, \tag{23}$$

The gain parameter $\bar{\Gamma}$ and efficiency parameter $\bar{\rho}$ are given by:

$$\bar{\Gamma} = \left[\frac{I}{I_{\mathrm{A}}}\frac{8\pi^2 K^2 A^2_{\mathrm{JJ}}}{\lambda\lambda_{\mathrm{w}}\gamma^3}\right]^{1/2}\,, \qquad \bar{\rho} = \frac{\lambda_{\mathrm{w}}\bar{\Gamma}}{4\pi}\,. \tag{24}$$

Here, $E = \gamma mc^2$ is the energy of electron, γ is a relativistic factor, and $C = 2\pi/\lambda_{\mathrm{w}} - \omega/(2c\gamma_z^2)$ is the detuning of the electron with nominal energy $\mathcal{E}_0$. Note that the efficiency parameter $\bar{\rho}$ entering the equations of the three-dimensional theory relates to the one-dimensional parameter ρ as $\rho = \bar{\rho}/B^{1/3}$ [4, 6]. The following notation is used here: I is the beam current, $\omega = 2\pi c/\lambda$ is the frequency of the electromagnetic wave, λ_{w} is undulator period, K is the r.m.s. undulator parameter, $\gamma_z^{-2} = \gamma^{-2} + \theta_{\mathrm{s}}^2$, $I_{\mathrm{A}} = mc^3/e = 17$ kA is the Alfven current, $A_{\mathrm{JJ}} = 1$ for a helical undulator and $A_{\mathrm{JJ}} = J_0(K^2/2(1+K^2)) - J_1(K^2/2(1+K^2))$ for a planar undulator. J_0 and J_1 are the Bessel functions of the first kind. The energy spread is assumed to be Gaussian with r.m.s. deviation σ_{E}.

The amplification process in a SASE FEL starts from the shot noise in the electron beam. At the initial stage of amplification, the coherence properties are poor, and the radiation consists of a large number of transverse and longitudinal modes [6, 39, 67–73]:

$$\tilde{E} = \sum_{m,n}\int \mathrm{d}\omega A_{mn}(\omega, z)\Phi_{mn}(r,\omega)\exp[\Lambda_{mn}(\omega)z + \mathrm{i}m\phi + \mathrm{i}\omega(z/c - t)]\,. \tag{25}$$

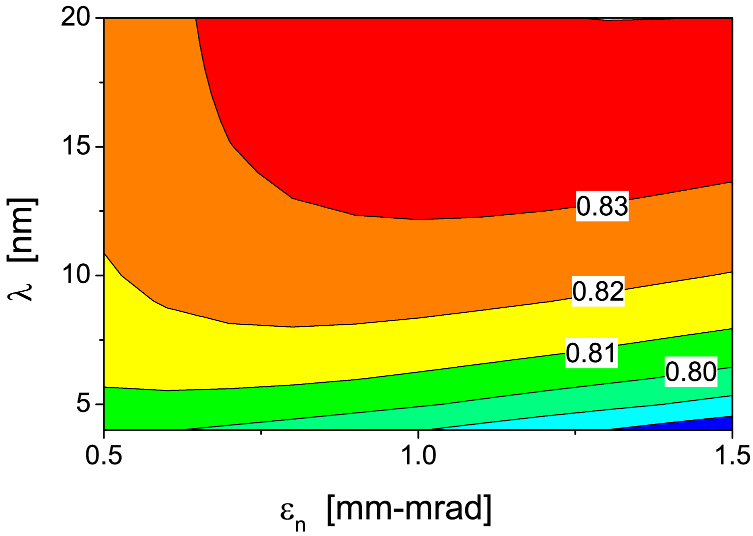


Fig. 31: Contour plot of ratio of maximum field gain of TEM_{10} to field gain of the ground TEM_{00} mode versus radiation wavelength and emittance: beam current, 1.5 kA; beta function, 10 m. Calculations made with simulation code FAST [77].

Each mode is characterized by the eigenvalue $\Lambda_{mn}(\omega)$ and the field distribution eigenfunction $\Phi_{mn}(r,\omega)$. The real part of the eigenvalue $\mathrm{Re}(\Lambda_{mn}(\omega))$ is referred to as the field gain. The field gain length is $L_g = 1/\mathrm{Re}(\Lambda_{mn}(\omega))$. Eigenvalues and eigenfunctions are the solutions of the eigenvalue equation [69, 70]. Each eigenvalue has a maximum at a certain frequency (or, at a certain detuning), so that the detuning for each mode is chosen automatically in the case of a SASE FEL (in contrast with seeded FELs, where the detuning can be set to any value). Thus, we will in fact deal with the three dimensionless parameters: B, $\hat{k}_\beta$, and $\hat{\Lambda}^2_{\mathrm{T}}$.

Let us look closer at the properties of the radiation modes. The gains for several modes are depicted in Fig. 31 as functions of the diffraction parameter. The values for the gain correspond to the maximum of the scan over the detuning parameter $\hat{C}$. The curve for the TEM_{00} mode shows the values of normalized gain $\mathrm{Re}(\Lambda_{00}/\bar{\Gamma})$. Curves for the higher spatial modes show the ratio of the gain of the mode to the gain of the fundamental mode, $\mathrm{Re}(\Lambda_{mn}/\Lambda_{00})$. Sorting of the modes by the gain results in the following ranking: TEM_{00}, TEM_{10}, TEM_{01}, TEM_{20}, TEM_{11}, TEM_{02}. The gain of the fundamental TEM_{00} mode is always above the gain of higher-order spatial modes. The difference in the gain between the fundamental TEM_{00} mode and higher-order spatial modes is pronounced for small values of the diffraction parameter $B \lesssim 1$. The gain of higher-order spatial modes approaches, asymptotically, the gain of the fundamental mode for large values of the diffraction parameter. In other words, the effect of mode degeneration takes place. Its origin can be understood through a qualitative analysis of the eigenfunctions (distribution of the radiation field in the near zone). Figure 32 shows eigenfunctions of the FEL radiation modes for two values of the diffraction parameter, $B = 1$ and $B = 10$. We observe that, for small values of the diffraction parameter, the field of the higher spatial modes spans far away from the core of the electron beam, while the fundamental TEM_{00} mode is more confined. This feature provides a higher coupling factor of the radiation with the electron beam and higher gain. For large values of the diffraction parameter, all radiation modes shrink to the beam axis, which results in equalization of coupling factors and of the gain. Asymptotically, the eigenvalues of all modes tends to the one-dimensional asymptote as [43]:

$$\Lambda_{mn}/\bar{\Gamma} \simeq \frac{\sqrt{3}+\mathrm{i}}{2B^{1/3}} - \frac{(1+\mathrm{i}\sqrt{3})(1+n+2m)}{3\sqrt{2}B^{2/3}} \,. \tag{26}$$

For a SASE FEL, the undulator length to saturation is in the range from about nine (hard X-ray range) to eleven (visible range) field gain lengths [41, 42, 65]. The mode selection process stops

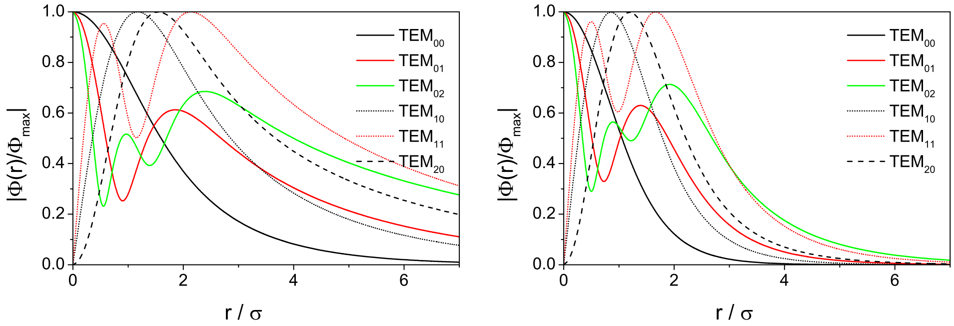


Fig. 32: Amplitude of the eigenfunctions of the FEL radiation modes, $|\Phi_{mn}(r)|/|\Phi_{\max}|$. Left: diffraction parameter $B = 1$. Right: $B = 10$. Detuning corresponds to the maximum of the gain. Energy spread parameter, $\hat{\Lambda}_{\mathrm{T}}^2 \to 0$; betatron motion parameter,s $\hat{k}_\beta \to 0$. Colour codes refer to the radial index of the mode: 0, black; 1, red; 2, green. Line type codes refer to the azimuthal index of the mode: 0, solid line; 1, dotted line; 2, dashed line. Calculations made with simulation code FAST [77].

at the onset of the non-linear regime, about two field gain lengths before saturation. Let us make a simple estimate of the value of the diffraction parameter $B = 10$ and the cold electron beam, $\hat{\Lambda}_{\mathrm{T}}^2 \to 0$, and $\hat{k}_\beta \to 0$. We get from Fig. 31 that the ratio of the gain $\mathrm{Re}(\Lambda_{10}/\Lambda_{00})$ is equal to 0.87. With an assumption of similar values of coupling factors, we find that the ratio of the field amplitudes at the onset of the non-linear regime is about of factor of three only. An estimate of the contribution of the higher spatial modes to the total power is about 10%. Another numerical example for $B = 1$ gives the ratio $\mathrm{Re}(\Lambda_{10}/\Lambda_{00}) = 0.73$, and the ratio of field amplitudes exceeds a factor of 10. Thus, an excellent transverse coherence of the radiation is not expected for a SASE FEL with a diffraction parameter $B \gtrsim 10$ and a small velocity spread in the electron beam.

A longitudinal velocity spread due to the energy spread and emittance serves as a tool for selective suppression of the gain of the higher spatial modes [6,68]. Figures 33 and 34 show the dependence of the gain of TEM_{00} and TEM_{10} modes on the betatron motion parameter and the energy spread parameter. We see that with the fixed value of the diffraction parameter, the mode degeneration effect can be relaxed at the price of gain reduction.

The betatron motion can influence the gain of different modes (and, therefore, transverse coherence properties) via two different mechanisms. First, the particles move across the beam, thus transferring the information between different points in the beam cross-section. Second, as already mentioned, there is a spread of longitudinal velocities that has a similar effect as the energy spread (and is usually more important than the first one). One can introduce a combination of parameters B and $\hat{k}_\beta$ that can, to some extent, be similar to the energy spread parameter:

$$(\hat{\Lambda}_{\mathrm{T}}^2)_{\mathrm{eff}} = B^2 \hat{k}_\beta^4 \tag{27}$$

Finally, let us note that the situation with transverse coherence is favourable when relative separation of the gain between the fundamental and higher spatial modes is more than 25–30%. In this case, the degree of transverse coherence can reach values above 90% at the end of the high-gain linear regime [39, 43]. Further development of the amplification process in the non-linear stage leads to a significant degradation of the spatial and temporal coherence [41, 42, 65].

4.2 Coherence properties of the radiation from the FLASH free electron laser

In the current experimental situation, many parameters of the electron beam at FLASH depend on practical tuning of the machine. Analysis of measurements and numerical simulations shows that, depending on the tuning of the machine, emittance may change from about 1 to about 1.5 mm-mrad. Tuning at

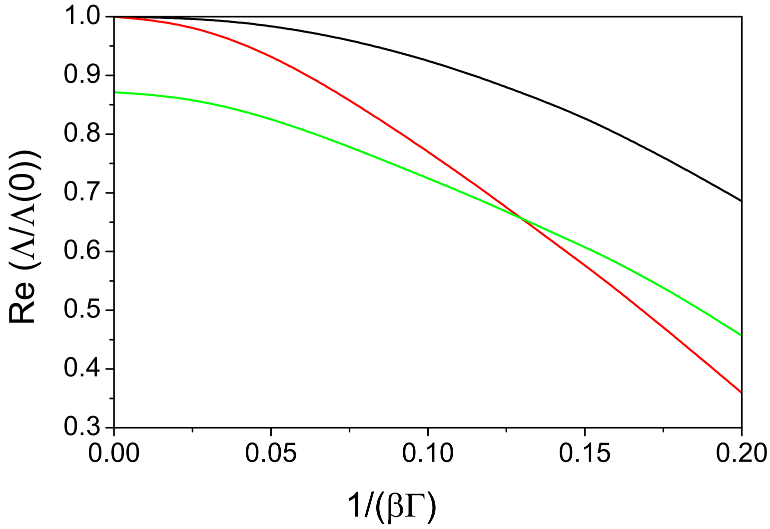


Fig. 33: Dependence of gain of (black) TEM_{00} and (red) TEM_{10} modes on betatron motion parameter $\hat{k}_\beta = 1/(\beta\bar{\Gamma})$. Values normalized to those at $\hat{k}_\beta \to 0$. Green curve shows ratio of gain of TEM_{10} mode to gain of TEM_{00} mode. Diffraction parameter, $B = 10$; energy spread parameter, $\hat{\Lambda}_\mathrm{T}^2 \to 0$. Calculations made with simulation code FAST [77].

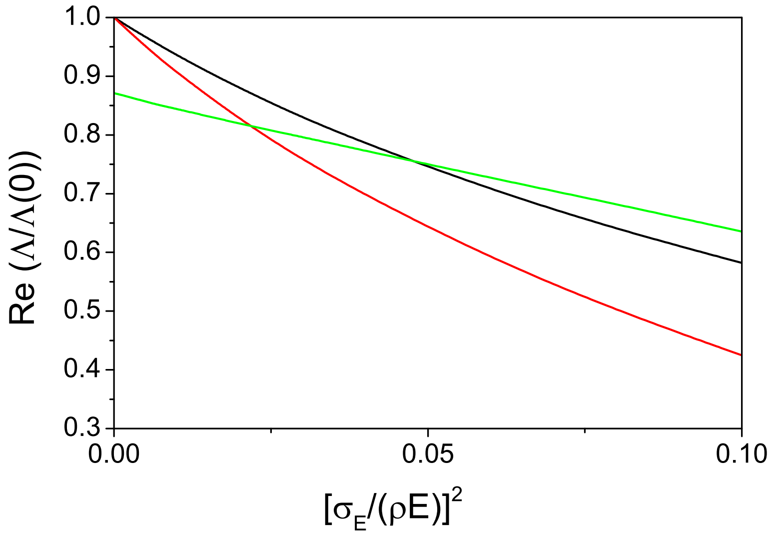


Fig. 34: Dependence of gain of (black) TEM_{00} and (red) TEM_{10} modes on energy spread parameter $\hat{\Lambda}_\mathrm{T}^2$. Values normalized to those at $\hat{\Lambda}_\mathrm{T}^2 \to 0$. Green curve shows ratio of gain of TEM_{10} mode to gain of TEM_{00} mode. Diffraction parameter, $B = 10$; betatron oscillation parameter, $\hat{k}_\beta \to 0$. Calculations made with simulation code FAST [77].

small charges may enable smaller values of the emittance, down to 0.5 mm-mrad, to be reached. The peak current may change in the range from 1 kA to 2 kA, depending on the tuning of the beam formation system. An estimate of the local energy spread is σ_{E} [MeV] $\simeq 0.1 \times I$ [kA]. The average beta function in the undulator is about 10 m.

Let us choose a reference working point with a radiation wavelength 8 nm, r.m.s. normalized emittance 1 mm-mrad, and beam current 1.5 kA. The parameters of the problem for this reference point are: diffraction parameter, $B = 17.2$; energy spread parameter, $\hat{\Lambda}^2_{\mathrm{T}} = 1.7 \times 10^{-3}$; betatron motion parameter, $\hat{k}_\beta = 5.3 \times 10^{-2}$. Then the reduced parameters at other working points can be easily recalculated using the scaling:

$$\hat{k}_\beta \propto \frac{1}{\beta I^{1/2} \lambda^{1/4}} , \qquad \hat{\Lambda}^2_{\mathrm{T}} \propto I \lambda^{1/2} , \qquad B \propto \frac{\epsilon_n \beta I^{1/2}}{\lambda^{1/4}} .$$

The effective contribution of the emittance to the longitudinal velocity spread (Eq. (27)) scales as

$$(\hat{\Lambda}^2_{\mathrm{T}})_{\mathrm{eff}} \propto \frac{\epsilon_n^2}{\beta^2 I \lambda^{3/2}}$$

and equals 2.3×10^{-3} at the considered working point.

Analysing these simple dependencies in terms of their effect on mode separation, we can state that:

- Dependencies on the wavelength are relatively weak (except for $(\hat{\Lambda}^2_{\mathrm{T}})_{\mathrm{eff}}$), i.e. one should not expect a significantly better transverse coherence at longer wavelengths. Moreover, mode separation can even be somewhat improved at shorter wavelengths, owing to a significant increase in $(\hat{\Lambda}^2_{\mathrm{T}})_{\mathrm{eff}}$.
- Reduction of the peak current (by going to a weaker bunch compression) would lead to an improvement of mode separation (even though the energy spread parameter would smaller). Obviously, the peak power at the saturation would be reduced.
- Dependence on the normalized emittance is expected to be weak because of the two competing effects. Mode separation due to a change of the diffraction parameter can be to a large extent compensated by a change of the longitudinal velocity spread. As we will see next, this does indeed indeed in the considered parameter range.

A contour plot for the value of the diffraction parameter B for the value of beta function of 10 m and the value of beam current 1.5 kA isprese0nted in Fig. 35. We see that the value of the diffraction parameter is $B \gtrsim 10$ in the whole parameter space. Figure 31 shows the ratio of the field gain $\mathrm{Re}(\Lambda_{10}(\omega))$ to the value of the field gain $\mathrm{Re}(\Lambda_{00}(\omega))$ of the fundamental mode. We see that this ratio is above 0.8 in the whole range of parameters, and we can expect significant contribution of the first azimuthal mode to the total radiation power. We can also notice relatively weak dependencies on the emittance and on the wavelength.

We illustrate the general characteristics of FLASH with a specific numerical example for FLASH operating at the wavelength of 8 nm, peak current 1.5 kA, and r.m.s. normalized emittance 1 mm-mrad.

4.2.1 Radiation power

Figure 36 shows the evolution along the undulator of the radiation power in the fundamental harmonic. Higher values of the peak current and smaller emittances would enable higher radiation powers to be achieved. When the amplification process enters the non-linear stage, a process of non-linear harmonic generation takes place [53–61, 64]. Contour plots in Fig. 37 show the relevant contribution to the total power of the third and the fifth harmonic at the saturation point. A general observation is that the relative contribution of the higher harmonic is higher for smaller values of the emittance. With the value of the normalized emittance of 1 mm-mrad, partial contributions for the third and the fifth harmonic are

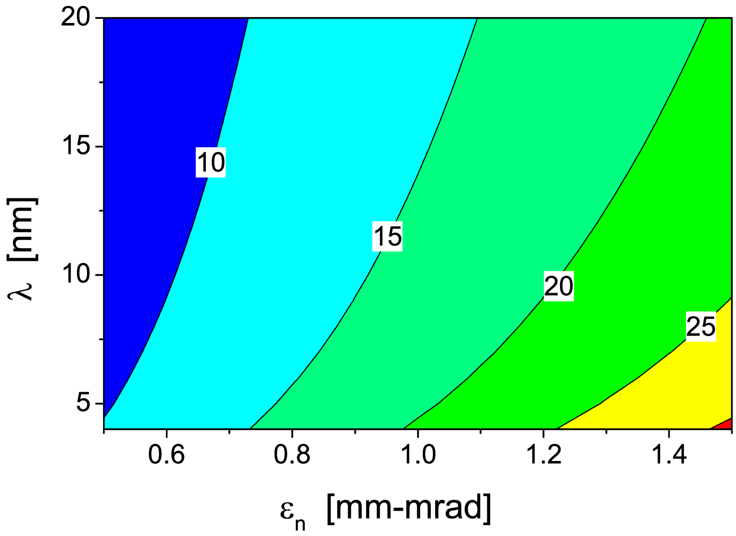


Fig. 35: Contour plot for diffraction parameter B versus normalized emittance and radiation wavelength. Beam current, 1.5 kA; beta function, 10 m. Calculations made with simulation code FAST [77].

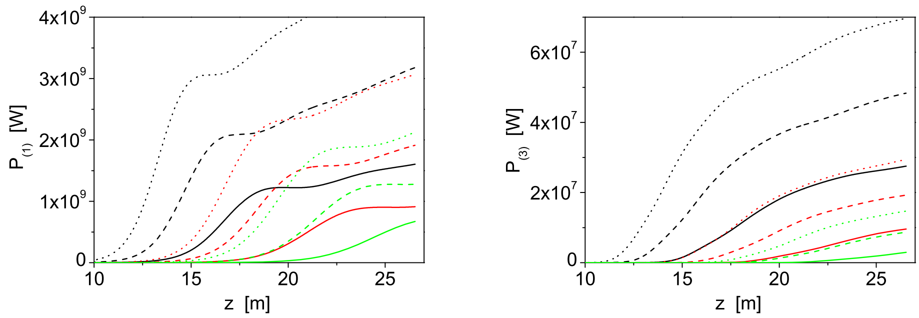


Fig. 36: Evolution of radiation power along undulator for (left) fundamental and (right) third harmonics. Colour codes (black, red, and green) refer to different emittances, $\epsilon_n = 0.5$, 1, and 1.5 mm-mrad. Line styles (solid, dash, and dot) refer to different values of peak current, 1 kA, 1.5 kA, and 2 kA. Radiation wavelength, 8 nm; beta function, 10 m. Calculations made with simulation code FAST [77].

0.7×10^{-2} and 2×10^{-4}, respectively. Note that this result is fairly close to that described by universal scaling law of Eq. (15), with an appropriate correction for longitudinal velocity spread derived in the previous section.

4.2.2 Temporal coherence

Plots in Fig. 38 show the coherence time for the whole parameter range for the fundamental and the third harmonic. In the high-gain linear regime, the coherence time increases as a square root of undulator length. It reaches a maximum value just before saturation point, and then decreases. The value of the coherence time at the saturation is close to that derived in the previous section in the framework of the one-dimensional theory (Eq. (17)), in terms of the FEL parameter ρ [4] and the number of cooperating electrons $N_{\rm c} = I/(e\rho_{\rm 1D}\omega)$ [6]. The coherence time for the higher harmonics at the saturation point and in the post-saturation amplification stage can be derived using the scaling derived in the previous section—it scales inversely proportional to the harmonic number, while the relative spectrum bandwidth remains constant with the harmonic number.

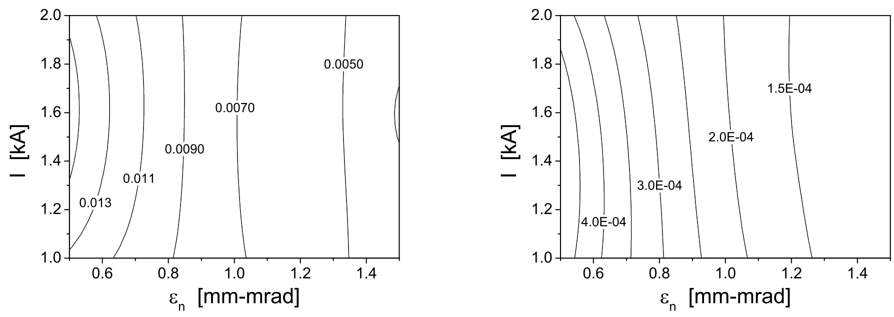


Fig. 37: Partial contribution of (left) third and (right) fifth harmonics to the total power versus peak beam current and emittance. Left and right parts of the plots are the saturation point and undulator end, respectively. Radiation wavelength, 8 nm; beta function, 10 m; beam current, 1.5 kA; r.m.s. normalized emittance, 1 mm-mrad. Calculations made with simulation code FAST [77].

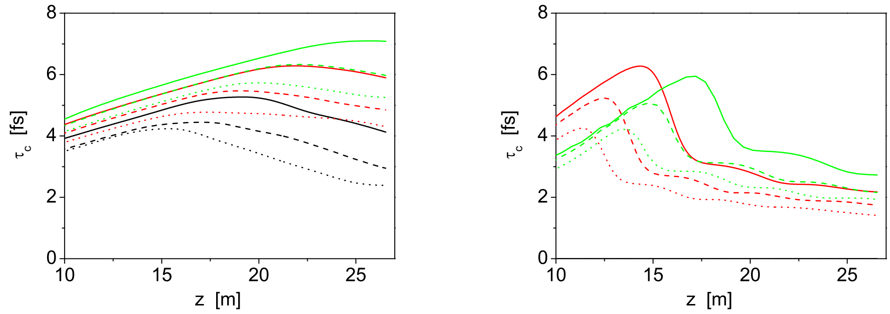


Fig. 38: Evolution along undulator of coherence time of radiation at (left) fundamental and (right) third harmonics. Colour codes (black, red, and green) refer to different emittance,$\epsilon_n = 0.5$, 1, and 1.5 mm-mrad. Line styles (solid, dash, and dot) refer to different values of peak current, 1 kA, 1.5 kA, and 2 kA. Radiation wavelength, 8 nm; beta function, 10 m. Calculations made with simulation code FAST [77].

4.2.3 Spatial coherence

Figure 39 presents an overview of the degree of transverse coherence in the considered parameter space. In our studies of coherent properties of FELs [41], we have found that, for an optimized SASE FEL, the degree of transverse coherence can be as high as 0.96. One can see from Fig. 39 that, in the considered cases, the degree of transverse coherence for the first harmonic is visibly less.

We should distinguish two effects limiting the degree of transverse coherence at FLASH. The first one is called mode degeneration and was intensively discussed in this paper. This physical phenomenon takes place at large values of the diffraction parameter [6]. Figure 40 shows the contribution of higher azimuthal modes to the total power for the specific example of emittance 1 mm-mrad and peak current 1.5 kA. The contribution of the first azimuthal modes decreases in the high-gain linear regime, but only to 12%, and then starts to increase in the non-linear regime, reaching 16% at the undulator end.

The second effect is connected with a finite longitudinal coherence; it was discovered in Ref. [39] and discussed in Refs. [41, 42]. The essence of the effect is a superposition of mutually incoherent fields produced by different longitudinally uncorrelated parts of the electron bunch. In the exponential gain regime, this effect is relatively weak, but it prevents a SASE FEL from reaching full transverse coherence, even when only one transverse eigenmode survives [39]. In the deep non-linear regime beyond FEL saturation, this effect can be strong and can lead to a significant degradation of the degree of transverse coherence [41, 42]. In particular, as one can see from Fig. 39, this effect limits the degree of transverse coherence to about 50% when FLASH operates in the deep non-linear regime.

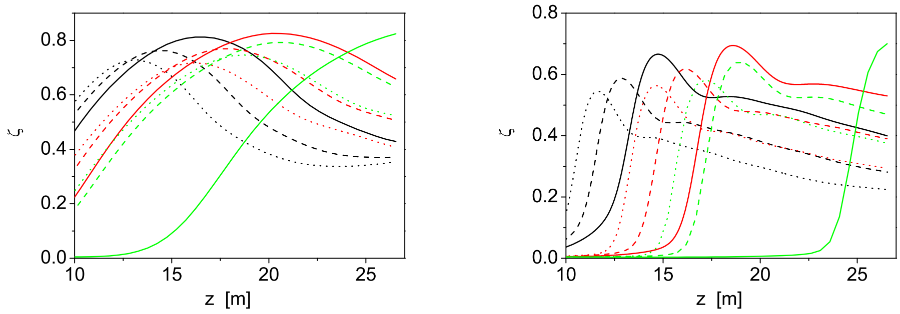


Fig. 39: Evolution along undulator of degree of transverse coherence of radiation. Left: fundamental frequency (8 nm). Right: third harmonic (2.66 nm). Colour codes (black, red, and green) refer to different emittances, $\epsilon_n = 0.5$, 1, and 1.5 mm-mrad. Line styles (solid, dash, and dot) refer to different values of peak current, 1 kA, 1.5 kA, and 2 kA. Radiation wavelength, 8 nm; beta function, 10 m. Calculations made with simulation code FAST [77].

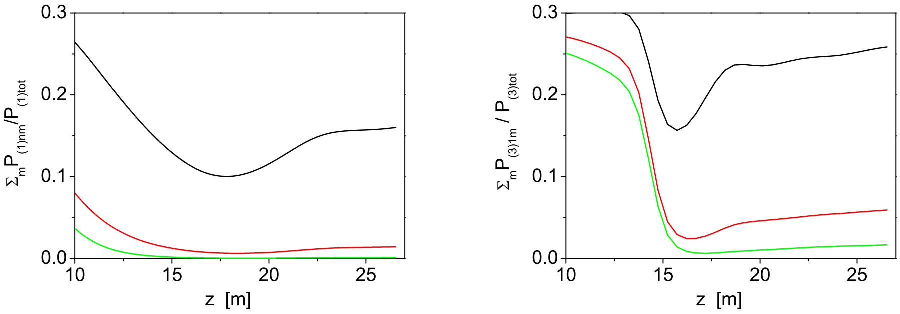


Fig. 40: Partial contribution of the higher azimuthal modes for (left) fundamental and (right) third harmonic. Black, red, and green curves refer to modes with $n = \pm 1$, $n = \pm 2$, and $n = \pm 3$, respectively. Radiation wavelength, 8 nm; beta function, 10 m; beam current, 1.5 kA; r.m.s. normalized emittance, 1 mm-mrad. Calculations made with simulation code FAST [77].

Higher harmonics are derived from the non-linear process governed by the fundamental harmonic. As a result, the coherence properties of the harmonics follow the same tendencies as the fundamental, but with visibly lower degree of transverse coherence [65].

Note that an easier way to improve the transverse coherence dramatically would be to decrease the beam current such that saturation is achieved at the very end of the undulator. This would eliminate not only the degradation in the deep non-linear regime, but would also improve the mode selection process because the diffraction parameter is then reduced while the velocity spread due to emittance is increased. According to our expectations, the degree of transverse coherence might reach around 90% in this regime. Such a regime was realized at FLASH on a user‘s demand, but it is not typical for the machine operation because the peak power is low due to a low peak current.

One can also suppress unwanted effects in the deep non-linear regime by kicking the electron beam at the saturation point (or close to it) when the peak current is high. Then one can still have a high power and an improved (about 70–80%) degree of transverse coherence. Further improvement could be achieved by reducing the beta function (thus improving the mode selection, as discussed).

4.2.4 Pointing stability and mode degeneration

Mode degeneration has a significant impact on the pointing stability of a SASE FEL. Let us illustrate this effect with a specific example for FLASH, operating with average energy in the radiation pulse of 60 µJ. The left plot in Fig. 41 shows evolution along the undulator of the radiation energy in azimuthally

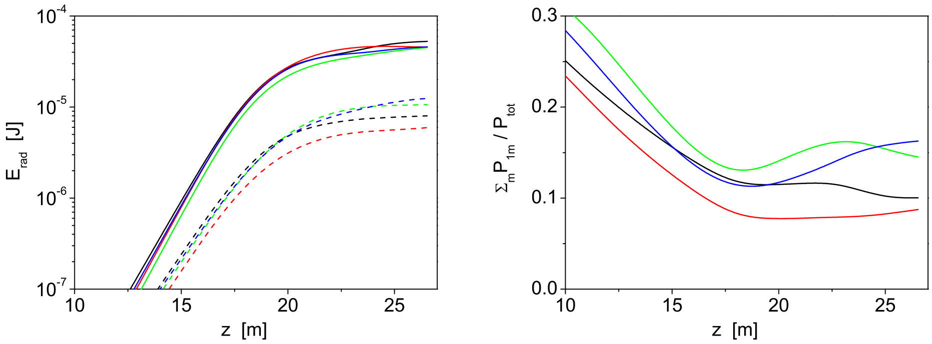


Fig. 41: Left: evolution of energy in radiation pulse versus undulator length. Colour codes (black to blue) correspond to different shots. Line styles correspond to total energy in the azimuthally symmetrical $\sum \mathrm{TEM}_{0m}$ modes (solid lines), and in the first azimuthal modes $\sum \mathrm{TEM}_{1m}$ (dashed lines). Right: partial contribution of the first azimuthal modes to the total radiation power, $\sum P_{1m}/P_{\mathrm{tot}}$. Radiation wavelength, 8 nm; beta function, 10 m; beam current, 1.5 kA; r.m.s. normalized emittance, 1 mm-mrad. Calculations made with simulation code FAST [77].

symmetrical modes and of the energy in the modes with azimuthal index $n = \pm 1$. The right plot in this figure shows relative contribution to the total radiation energy of the modes with azimuthal index $n = \pm 1$. Four consecutive shots are shown here. Temporal profiles of the radiation pulses are presented in Fig. 42. The intensity distributions in the far zone for these four shots are shown in four rows in Fig. 43. Four profiles on the left-hand side of each row show intensity distributions in the single slices for the times 40 fs, 50 fs, 60 fs, and 70 fs. The right column presents intensity profiles averaged over full shots. We see that the transverse intensity patterns in the slices have a rather complicated shape, owing to interference of the fields of statistically independent modes with different azimuthal indexes. The shape of the intensity distributions changes on a scale of coherence length. Averaging the slice distributions over the radiation pulse results in a smoother distribution. However, it can clearly be seen that the spot shape of a short radiation pulse changes from pulse to pulse. The centre of gravity of the radiation pulse visibly jumps from shot to shot. The position of the pulse also jumps from shot to shot; this is frequently referred to as bad pointing stability. Note that the effect illustrated here is a fundamental one, which takes place as a result of the mode degeneration when the contribution of the higher azimuthal modes to the total power is pronounced (10–15% in our case). Only in the case of a long radiation pulse, or after averaging over many pulses, do we come asymptotically to an azimuthally symmetrical radiation distribution.

4.3 Optimized SASE FEL

The best properties of the output radiation from SASE FEL correspond to the case when the fundamental TEM_{00} mode dominates over higher-order spatial modes. Thus, the standard procedure for optimization of the SASE FEL is optimization for the maximum gain of the fundamental mode. For given parameters of the electron beam and the undulator there always exists an optimum focusing beta function β_{opt}, which provides a minimum gain length L_{g} of the fundamental mode [66, 80]:

$$
\begin{aligned}
L_{\mathrm{g}0} &= 1.67 \left(\frac{I_{\mathrm{A}}}{I}\right)^{1/2} \frac{(\epsilon_n \lambda_{\mathrm{w}})^{5/6}}{\lambda^{2/3}} \frac{(1+K^2)^{1/3}}{K A_{\mathrm{JJ}}} (1+\delta)\,, \\
\beta_{\mathrm{opt}} &\simeq 11.2 \left(\frac{I_{\mathrm{A}}}{I}\right)^{1/2} \frac{\epsilon_n^{3/2} \lambda_{\mathrm{w}}^{1/2}}{\lambda K A_{\mathrm{JJ}}} (1+8\delta)^{-1/3}\,, \\
\delta &= 131\, \frac{I_{\mathrm{A}}}{I} \frac{\epsilon_n^{5/4}}{\lambda^{1/8} \lambda_{\mathrm{w}}^{9/8}} \frac{\sigma_\gamma^2}{(K A_{\mathrm{JJ}})^2 (1+K^2)^{1/8}}\,.
\end{aligned} \tag{28}
$$

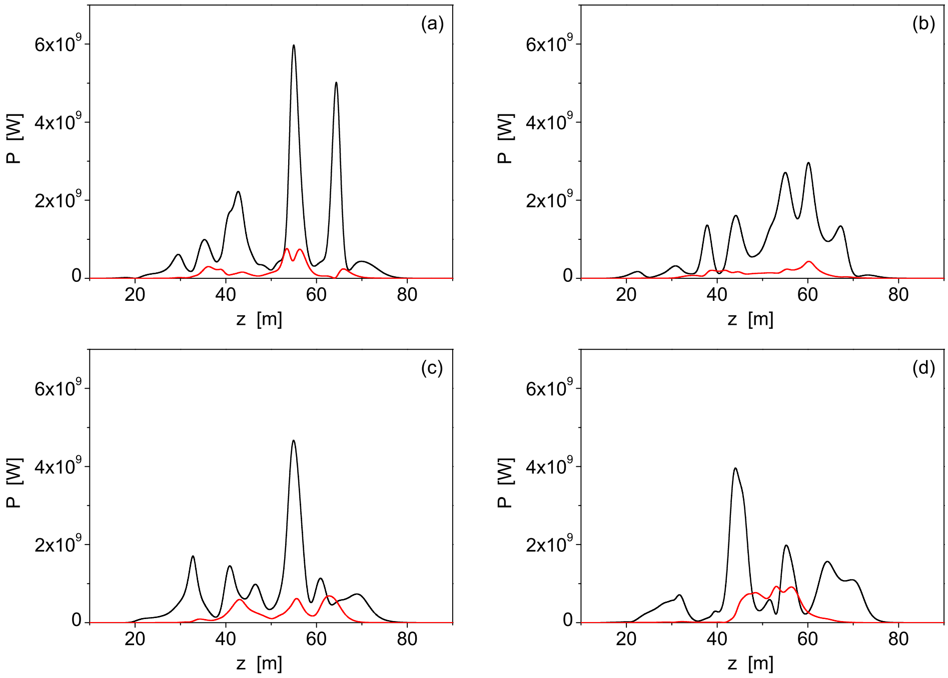


Fig. 42: Temporal structure of four radiation pulses. Black: power of azimuthally symmetrical modes. Red: power of first azimuthal modes. Radiation wavelength, 8 nm; beta function, 10 m; beam current, 1.5 kA; r.m.s. normalized emittance, 1 mm-mrad; undulator length, 27 m. Calculations made with simulation code FAST [77].

A realization of the conditions of Eq. (28) is referred to as an optimized FEL amplifier. Sometimes, for technical reasons, the focusing beta function might be $\beta > \beta_{\rm opt}$. In such a case, the gain length can be approximated as follows:

$$L_{\rm g}(\beta) \simeq L_{\rm g}(\beta_{\rm opt}) \left[1 + \frac{(\beta - \beta_{\rm opt})^2(1+8\delta)}{4\beta_{\rm opt}^2}\right]^{1/6} , \qquad \text{for } \beta > \beta_{\rm opt} \tag{29}$$

Here $\epsilon_n = \gamma\epsilon$ is the r.m.s. normalized emittance, $\sigma_\gamma = \sigma_{\mathcal{E}}/mc^2$ is the r.m.s. energy spread (in units of the rest energy), and factor $A_{\rm JJ}$ is the usual coupling factor defined in the previous sections. Equation (28) provides an accuracy better than 5% in the range of parameters

$$1 < \frac{2\pi\epsilon}{\lambda} < 5 , \qquad \delta < 2.5 \left\{1 - \exp\left[-\frac{1}{2}\left(\frac{2\pi\epsilon}{\lambda}\right)^2\right]\right\} \tag{30}$$

The saturation length of the optimized SASE FEL is given by [42]:

$$L_{\rm sat} \simeq 0.6\, L_{\rm g} \ln\left(\frac{N_\lambda L_{\rm g}}{\lambda_{\rm w}}\right) . \tag{31}$$

Here, $N_\lambda = I\lambda/c$ is the number of electrons per wavelength. When operating vacuum ultraviolet and X-ray SASE FELs, one typically has $L_{\rm sat} \simeq (10 \pm 1) \times\ L_{\rm g}$.

For small energy spread, $\delta \ll 1$, the physical parameters describing operation of the optimized FEL, the diffraction parameter B, and the parameter of betatron oscillations $\hat{k}_\beta$, are functions of the only parameter $\hat{\epsilon} = 2\pi\epsilon/\lambda$ [6,41]:

$$B = 2\bar{\Gamma}\sigma^2\omega/c \simeq 12.5 \times \hat{\epsilon}^{5/2} ,$$
$$\hat{k}_\beta = 1/(\beta\bar{\Gamma}) \simeq 0.158 \times \hat{\epsilon}^{-3/2} . \tag{32}$$

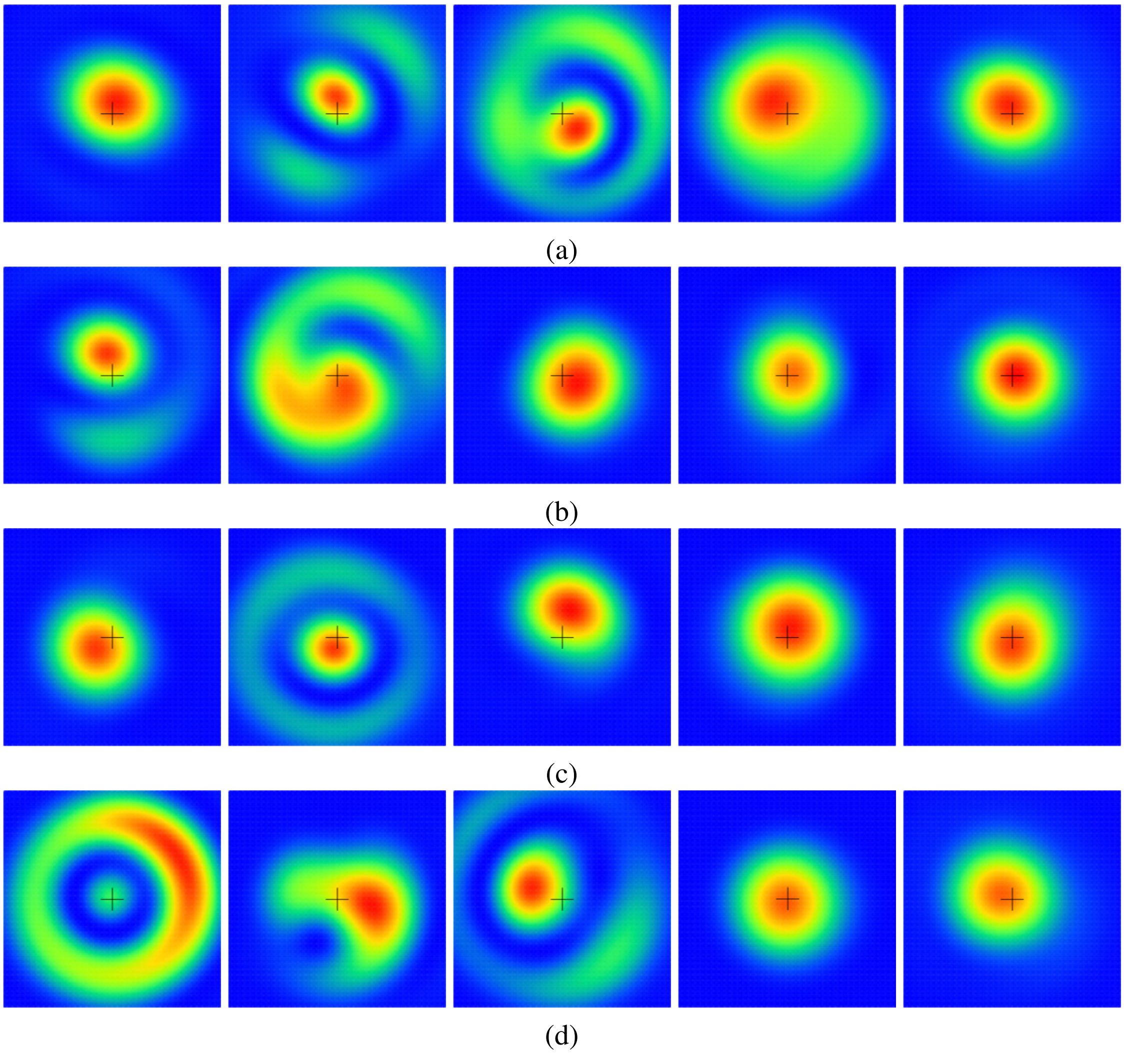


Fig. 43: Profiles of radiation intensity in the far zone. Rows (a–d) correspond to specific shots with temporal structures presented in Fig. 42 (plots a–d). Profiles on the right-hand side show average intensity over full pulse. Profiles 1 to 4 from the left-hand side show intensity distributions of selected slices corresponding to 40 fs, 50 fs, 60 fs, and 70 fs, respectively. Crosses denote the geometrical centre of the radiation intensity averaged over many shots. Radiation wavelength, 8 nm; beta function, 10 m; beam current, 1.5 kA; r.m.s. normalized emittance, 1 mm-mrad; undulator length, 27 m. Calculations made with simulation code FAST [77].

The diffraction parameter B directly relates to diffraction effects and the formation of transverse coherence. If diffraction expansion of the radiation on a scale of the field gain length is comparable to the transverse size of the electron beam, we can expect a high degree of transverse coherence. For this range of parameters, the value of the diffraction parameter is small. If diffraction expansion of the radiation is small (which happens at large values of the diffraction parameter), we can expect significant degradation in the degree of transverse coherence. This effect occurs simply because different parts of the beam produce radiation nearly independently. In terms of the radiation expansion in the eigenmodes of Eq. (25), this range of parameters corresponds to the degeneration of modes [6]. The diffraction parameter for an optimized XFEL exhibits strong dependence on the parameter $\hat{\epsilon}$ (see Eq. (32)), and we can expect the degree of transverse coherence to drop rapidly with the increase of the parameter $\hat{\epsilon}$.

4.3.1 Characteristics of the radiation from optimized SASE FEL operating in the saturation regime

Figure 1 shows the evolution of the main characteristics of a SASE FEL along the undulator. If one traces the evolution of the brilliance (degeneracy parameter) of the radiation along the undulator length, there is always a point (defined as the saturation point [41]) where the brilliance reaches a maximum value. The best properties of the radiation in terms of transverse and longitudinal coherence are achieved just before the saturation point, and these values then degrade significantly, despite the radiation power continuing to increase with undulator length.

Application of similarity techniques allows us to derive universal parametric dependencies of the output characteristics of the radiation at the saturation point. As mentioned in Section 2, within accepted approximations (optimized SASE FEL and negligibly small energy spread in the electron beam), normalized output characteristics of a SASE FEL at the saturation point are functions of only two parameters: $\hat{\epsilon} = 2\pi\epsilon/\lambda$ and the number of electrons in the volume of coherence $N_{\mathrm{c}} = IN_{\mathrm{g}}\lambda/c$, where $N_{\mathrm{g}} = L_{\mathrm{g}}/\lambda_{\mathrm{w}}$ is the number of undulator periods per gain length. Characteristics of practical interest are: saturation length L_{sat}, saturation efficiency $\eta_{\mathrm{sat}} = P_{\mathrm{sat}}/P_{\mathrm{b}}$ (ratio of the radiation power to the electron beam power $P_{\mathrm{b}} = \gamma mc^2 I/e$), coherence time τ_{c}, degree of transverse coherence ζ, degeneracy parameter δ, and brilliance B_{r}. Applications of similarity techniques to the results of numerical simulations of a SASE FEL [41] gives us the following result:

$$
\begin{aligned}
\hat{L}_{\mathrm{sat}} &= \bar{\Gamma} L_{\mathrm{sat}} \simeq 2.5 \times \hat{\epsilon}^{5/6} \times \ln N_{\mathrm{c}} \ , \\
\hat{\eta} &= P/(\bar{\rho} P_{\mathrm{b}}) \simeq 0.17/\hat{\epsilon} \ , \\
\hat{\tau_{\mathrm{c}}} &= \bar{\rho}\omega\tau_{\mathrm{c}} \simeq 1.16 \times \sqrt{\ln N_{\mathrm{c}}} \times \hat{\epsilon}^{5/6} \ , \\
\sigma_{\omega} &= \sqrt{\pi}/\tau_{\mathrm{c}} \ .
\end{aligned}
\qquad (33)
$$

These expressions provide reasonable practical accuracy for $\hat{\epsilon} \gtrsim 0.5$. With logarithmic accuracy in terms of N_{c}, characteristics of the SASE FEL expressed in a normalized form are only functions of the parameter $\hat{\epsilon}$. The saturation length, FEL efficiency, and coherence time exhibit monotonous behaviour in the parameter space of modern XFELs ($\hat{\epsilon} \simeq 0.5, \ldots, 5$). The situation with the degree of transverse coherence at saturation is more complicated, as we can see from Fig. 44—degradation of spatial coherence at small emittances seems to be counter-intuitive. To understand the origin of this phenomenon, we trace in Fig. 45 the degree of transverse coherence ζ versus the reduced propagation coordinate z/z_{sat} for different values of the parameter $\hat{\epsilon}$. The typical behaviour of the degree of transverse coherence is that it increases in the exponential stage of amplification, reaches a maximum value at the saturation point, and degrades in the post-saturation regime. The increase of the degree of transverse coherence in the exponential stage of amplification happens because of two physical effects. The first effect is the mode selection process, which is reflected in Eq. (25). As we have shown, smaller values of the diffraction parameter provide better mode selection. Simple calculations of the gain (see Fig. 46) shows that, for values of diffraction parameter less than or approximately equal to one (diffraction-limited beam), one should expect almost 100% contribution of the fundamental mode to the total power. However, this is not the case. The maximum degree of transverse coherence is degraded for smaller emittances, as we can see from Fig. 45. The physical effect responsible for such a degradation is the interdependence of the poor longitudinal coherence and transverse coherence [39]. We pay attention to the feature that, owing to the start-up from shot noise, every radiation mode entering Eq. (25) is excited within a finite spectral bandwidth. This means that, in the high-gain linear regime, the radiation of the SASE FEL is formed by many fundamental TEM_{00} modes with different frequencies. The transverse distribution of the radiation field of the mode is also different for different frequencies. A smaller value of the diffraction parameter (i.e. smaller value of $\hat{\epsilon}$) corresponds to larger deviation of the radiation mode from the plane wave. This effect explains the degradation of the transverse coherence at small values of $\hat{\epsilon}$. The degree of transverse coherence asymptotically approaches unity as $(1-\zeta) \propto 1/z \propto 1/\ln N_{\mathrm{c}}$ at small values of the emittance. The maximum value of the degree of transverse coherence is about 0.96 and is achieved at $\hat{\epsilon} \simeq 1$.

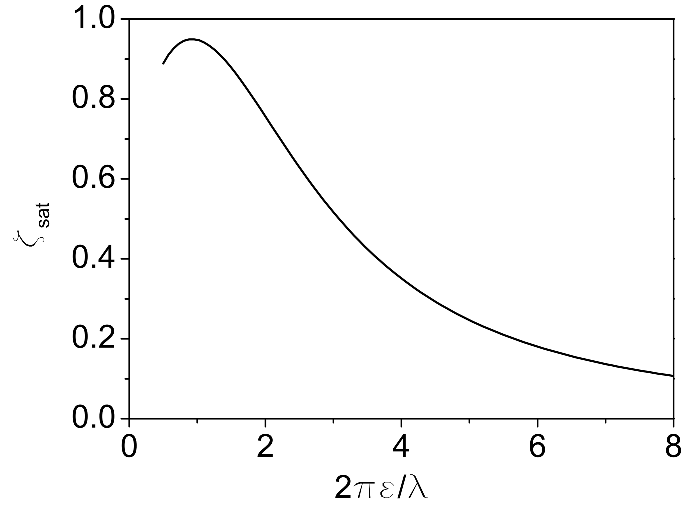


Fig. 44: Degree of transverse coherence ζ_{sat} in saturation point versus $\hat{\epsilon}$. Number of electrons in coherence volume, $N_{\mathrm{c}} = 4 \times 10^6$. Calculations made with simulation code FAST [77].

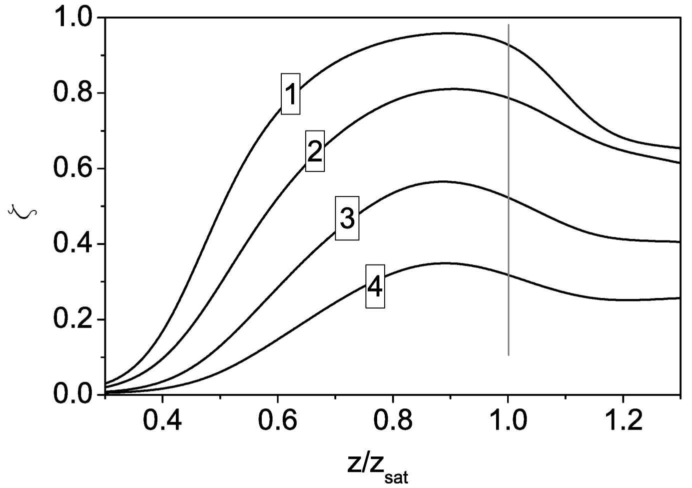


Fig. 45: Evolution of degree of transverse coherence along undulator length for $\hat{\epsilon} = 0.5$, 1, 2, 3, and 4. Undulator length normalized to saturation length. Calculations made with simulation code FAST [77].

When the parameter $\hat{\epsilon}$ is large, the diffraction parameter is also large, leading to degeneration of the radiation modes. The amplification process in the SASE FEL passes a limited number of field gain lengths and, starting from some value of $\hat{\epsilon}$ the linear stage of amplification, becomes too short to provide the mode selection process of Eq. (25). When the amplification process enters the non-linear stage, the mode content of the radiation becomes even richer, owing to independent growth of the radiation modes in the non-linear medium (see Fig. 47). Thus, at large values of $\hat{\epsilon}$, the degree of transverse coherence is limited by poor mode selection. Analytical estimates show that, in the limit of large emittance, $\hat{\epsilon} \gg 1$, the degree of transverse coherence scales as $1/\hat{\epsilon}^2$. To avoid complications, we present here just a fit for

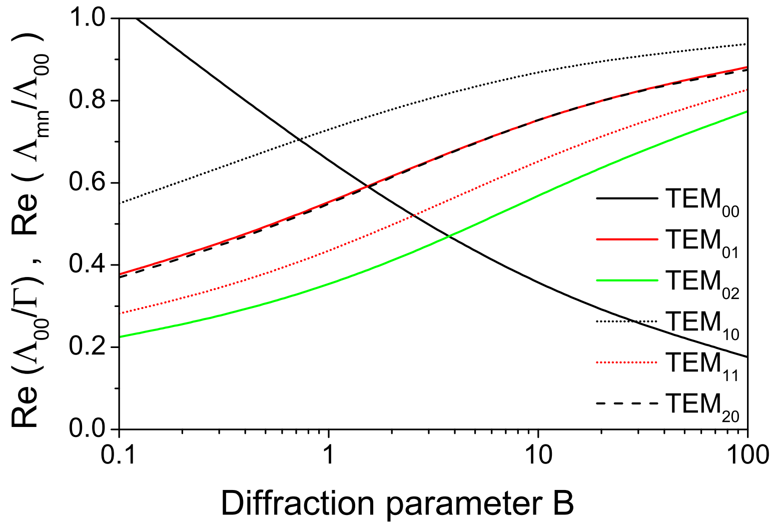


Fig. 46: Ratio of maximum gain of higher modes to maximum gain of fundamental mode $\mathrm{Re}(\Lambda_{mn})/\,\mathrm{Re}(\Lambda_{00})$ versus diffraction parameter B. Energy spread parameter, $\hat{\Lambda}_{\mathrm{T}}^2 \to 0$, betatron motion parameter, $\hat{k}_\beta \to 0$. Colour codes refer to radial index of the mode: 0, black; 1, red; 2, green. Line type codes refer to azimuthal index of the mode: 0, solid line, 1, dotted line; 2, dashed line. Black solid line shows gain of the fundamental mode $\mathrm{Re}(\Lambda_{00})/\bar{\Gamma}$. Calculations made with simulation code FAST [77].

the degree of transverse coherence for the number of electrons in the coherence volume $N_{\mathrm{c}} = 4 \times 10^6$:

$$\zeta_{\mathrm{sat}} \simeq \frac{1.1\hat{\epsilon}^{1/4}}{1 + 0.15\hat{\epsilon}^{9/4}} \,. \tag{34}$$

Recalculation from reduced to dimensional parameters is straightforward. For instance, the saturation length is $L_{\mathrm{sat}} \simeq 0.6 \times L_{\mathrm{g}} \times \ln N_{\mathrm{c}}$. Using Eq. (33), we can calculate the normalized degeneracy parameter $\hat{\delta} = \hat{\eta}\zeta\hat{\tau}_{\mathrm{c}}$ and then the brilliance (Eq. (9)):

$$B_r \left[\frac{\text{photons}}{\text{sec mrad}^2\ \text{mm}^2\ 0.1\%\ \text{bandw.)}}\right] \simeq 4.5 \times 10^{31} \times \frac{I[\text{kA}] \times E[\text{GeV}]}{\lambda[\,]} \times \hat{\delta} \,. \tag{35}$$

4.3.2 Coherence properties of the higher odd harmonics

We start the analysis with a specific numerical example corresponding to $\hat{\epsilon} = 0.5$. This operating point corresponds to the maximum degree of transverse coherence that can be achieved in SASE FEL [41,42, 44]. Figure 48 shows a slice of the temporal structure of the radiation pulse from a SASE FEL operating in the saturation regime. Already, this specific example provides a lot of physical information. We note that the spikes of all harmonics are well aligned in space, illustrating an effect of non-linear harmonic generation: higher harmonics radiate only by those parts of the electron bunch that have been effectively modulated by the fundamental harmonic. We also notice that the typical scale of the radiation intensities of the third (fifth) harmonic is in the range of a few per cent (per mille) of the fundamental. Even a brief look at the spike widths in Fig. 48 gives us an idea that the coherence time of the third harmonic is significantly less than that of the fundamental harmonic. The spikes of the fifth harmonic are shorter than those of the third harmonic; thus, the coherence time of the fifth harmonic should be even less.

The plots in Fig. 49 show the evolution along the undulator of the radiation power and brilliance. The longitudinal coordinate is normalized to the saturation length of the fundamental harmonic. The brilliance and power of the harmonics are normalized to the values corresponding to the saturation point of the fundamental harmonic. We see that the radiation powers of all harmonics continue to increase

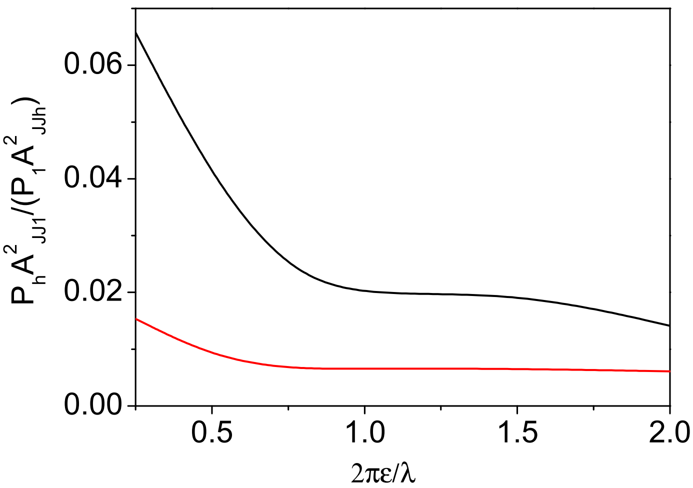


Fig. 47: Optimized XFEL. Ratio of powers in the (black) third and (red) fifth harmonics to the power of the fundamental harmonic versus $\hat{\epsilon} = 2\pi\epsilon/\lambda$; SASE FEL operates at saturation point. Calculations made with simulation code FAST [77].

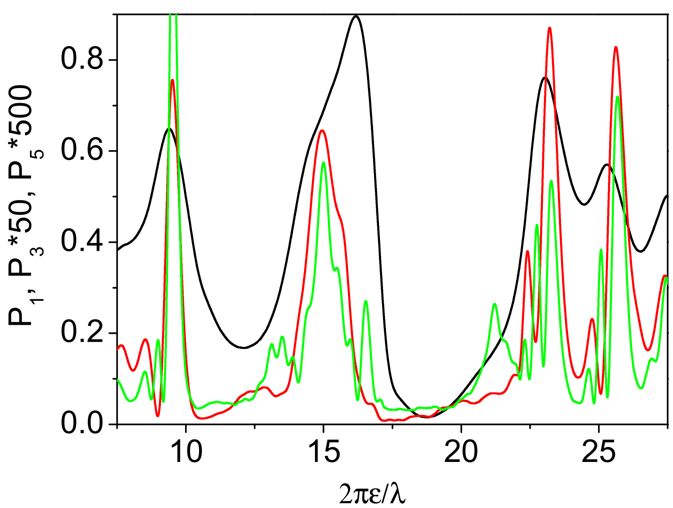
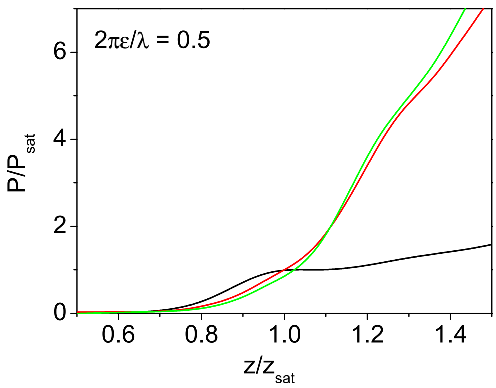


Fig. 48: Optimized XFEL. Temporal structure of radiation pulse in saturation point for $\hat{\epsilon} = 0.5$. Black, red, and green lines refer to first, third, and fifth harmonics, respectively. Calculations made with simulation code FAST [77].



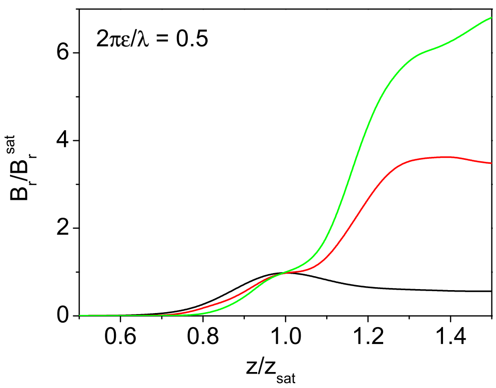


Fig. 49: Left: optimized XFEL: FEL power versus undulator length. Right: brilliance versus undulator length. All values normalized to values corresponding to the values at the saturation point of the first harmonic. Black, red, and green lines refer to first, third, and fifth harmonic, respectively. Calculations made with simulation code FAST [77].

after the saturation point of the fundamental harmonic. The increase in power of the third and the fifth harmonic is visibly faster than that of the fundamental. An important feature is that the brilliances of the higher harmonics also continue to increase after the saturation point. The maximum brilliance of the higher harmonics is reached in the deep non-linear regime, mainly as a result of faster growth of the harmonic radiation power with respect to the fundamental. This means that in a parameter range of $\hat{\epsilon} \approx 0.5$, the electron beam after saturation remains a relatively good amplification medium for higher harmonics. The contribution of higher harmonics to the total radiation power depends strongly on how long the amplification process develops after the saturation point.

The plots in Fig. 50 show the evolution along the undulator of the coherence time and degree of transverse coherence. We multiplied the coherence time by the harmonic number h to bring all curves into scale. We find an important feature; that coherence time in the saturation regime scales inversely proportional to harmonic number. Moreover, the relative spectrum bandwidth $\Delta\omega_h/\omega_h$ remains constant for all harmonics. This finding confirms the result obtained earlier in the framework of the one-dimensional model [64]. Note that recent measurements of the harmonic properties at FLASH and LCLS [13, 81] are in good qualitative agreement with the results reported here.

Figure 49 shows the evolution of the degree of transverse coherence along the undulator. Note that we illustrate the parameter space providing the maximum degree of transverse coherence in the fundamental harmonic (about 95%) for an optimized X-ray FEL. An important observation is that the degree of transverse coherence for higher harmonics is visibly less. There is nothing unusual in this result. Qualitatively, it can be explained by the general feature of frequency multiplication schemes, which also amplify noise progressively with harmonic number [82]. The fundamental harmonic already contains visible noise content, resulting in a reduced degree of transverse coherence, and we can readily expect further reduction for higher harmonics. An example of similar physical behaviour is the degradation of longitudinal coherence in the high-gain harmonic generation scheme [83].

As already mentioned, the characteristics of the optimized FEL in the saturation point depend only on the parameter, $\hat{\epsilon}$. Figure 51 shows the dependence of the degree of transverse coherence for the first, third, and fifth harmonic on the value of the emittance parameter. We see that the maximum values of the degree of transverse coherence correspond to values of $\hat{\epsilon} \approx 0.5$. While the coherence properties of the fundamental harmonic do not change too much when $\hat{\epsilon}$ increases to 2, we obtain a significant degradation for the third harmonic.

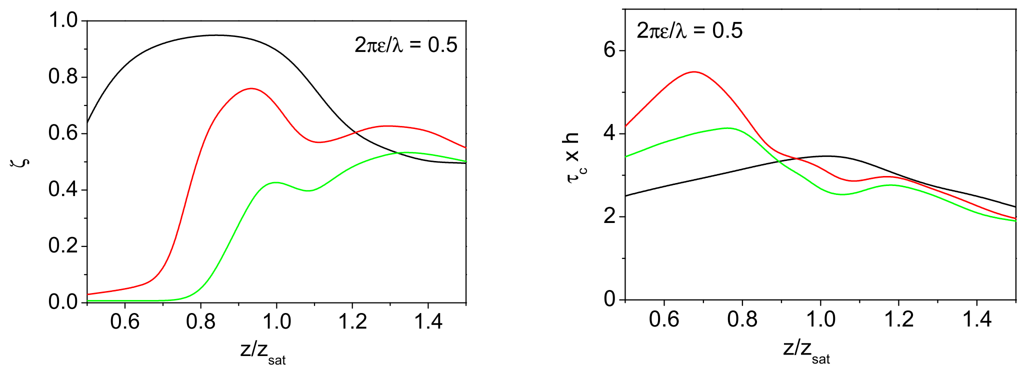


Fig. 50: Optimized XFEL. Left: degree of transverse coherence, ζ. Right: normalized coherence time,$\hat{\tau}_{\mathrm{c}}$, versus undulator length for $\hat{\epsilon} = 0.5$. Black, red, and green lines refer to first, third, and fifth harmonic, respectively. Coherence time is multiplied by corresponding harmonic number h. Calculations made with simulation code FAST [77].

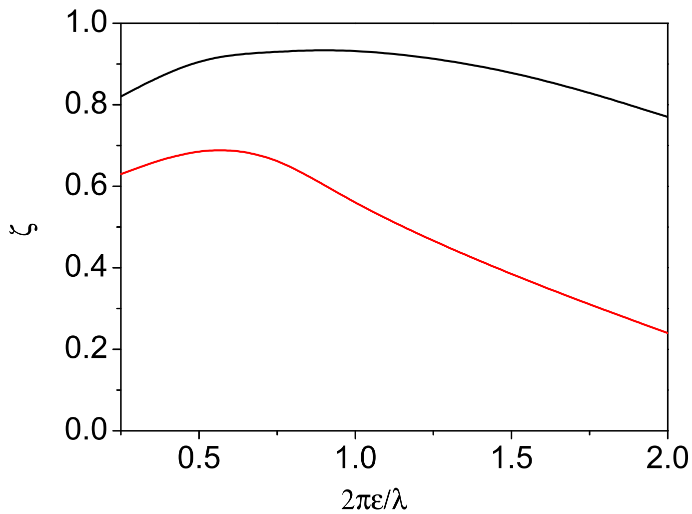


Fig. 51: Optimized XFEL: degree of transverse coherence ζ_{sat} in saturation versus parameter $\hat{\epsilon} = 2\pi\epsilon/\lambda$. Black and red lines refer to first and third harmonics, respectively. Calculations made with simulation code FAST [77].

When we analyse expressions for the radiation power, we find that the dependencies for the ratios of the power of higher harmonics to the fundamental become universal functions of emittance parameter when we factorize them with factor $A^2_{\mathrm{JJh}}/A^2_{\mathrm{JJ1}}$. Relevant plots are presented in Fig. 47. For large values of the undulator parameter K, asymptotic values of $A^2_{\mathrm{JJh}}/A^2_{\mathrm{JJ1}}$ are equal to 0.22 and 0.11 for the third, and fifth harmonics, respectively. In the range of emittance parameter from 0.25 to 2, contributions to the total power of the third (fifth) harmonic are between 0.3% and 1.4% (0.07% and 0.16%). Note that the contribution of higher harmonics to the total power increases in the deep non-linear regime, and may constitute a substantial amount (see Fig. 49).

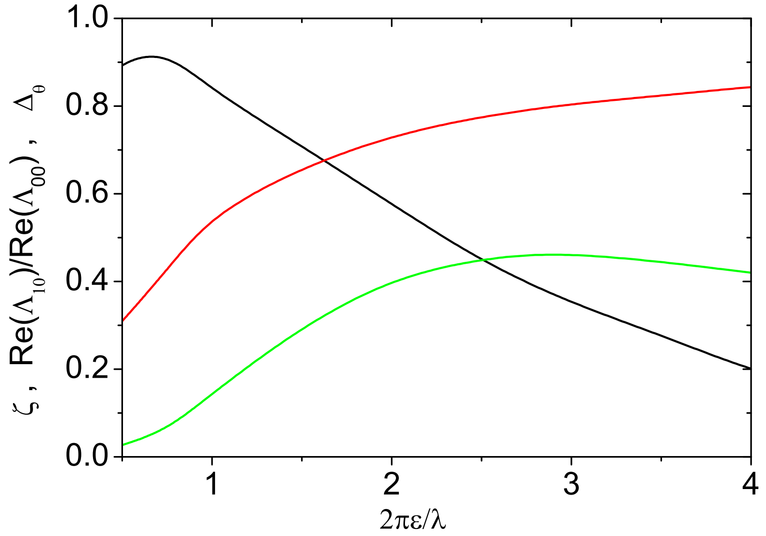


Fig. 52: Optimized SASE FEL: (black) degree of transverse coherence ζ, (red) ratio of gain $\mathrm{Re}(\Lambda_{10})/\,\mathrm{Re}(\Lambda_{00})$, and (green) r.m.s. deviation of photon beam centre of gravity Δ_θ in terms of r.m.s. size of photon beam. Simulations run with code FAST [77].

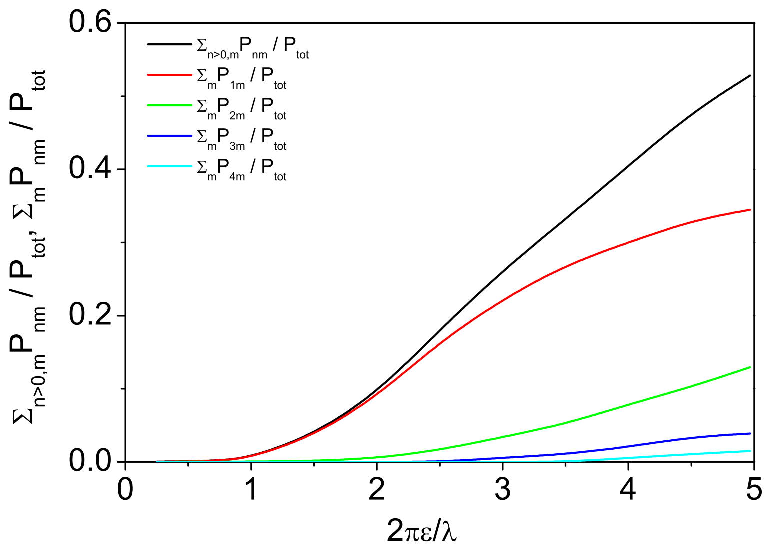


Fig. 53: Optimized SASE FEL: partial contributions of asymmetrical modes to total power versus emittance parameter $\hat{\epsilon} = 2\pi\epsilon/\lambda$; SASE FEL operates in saturation. Black curve is total contribution of asymmetrical modes, and colour curves correspond to azimuthal indices from 1 to 4. Simulations run with code FAST [77].

4.3.3 Mode degradation and pointing stability

The diffraction parameter scales with the emittance parameter as $B \simeq 13 \times \hat{\epsilon}^{5/2}$. Starting from $\hat{\epsilon} > 1$ the gain of the TEM_{10} mode approaches very close to the gain of the ground TEM_{00} mode (see Fig. 52). The contribution of the TEM_{10} mode to the total power progresses with the increase of the emittance parameter (see Fig. 53). Starting from $\hat{\epsilon} > 2$, the azimuthal modes TEM_{2n} appear in the mode contents, and so on. The maximum value of the degree of transverse modes (which occurs at the end of the linear regime) degrades gradually with the increase of the emittance parameter (see Fig. 52).

Mode degeneration has a significant impact on the pointing stability of a SASE FEL. Figure 53 shows the relative contribution to the total radiation energy of the modes with higher azimuthal indices. Typical intensity distributions in the far zone are shown in Fig. 54. Transverse intensity patterns in

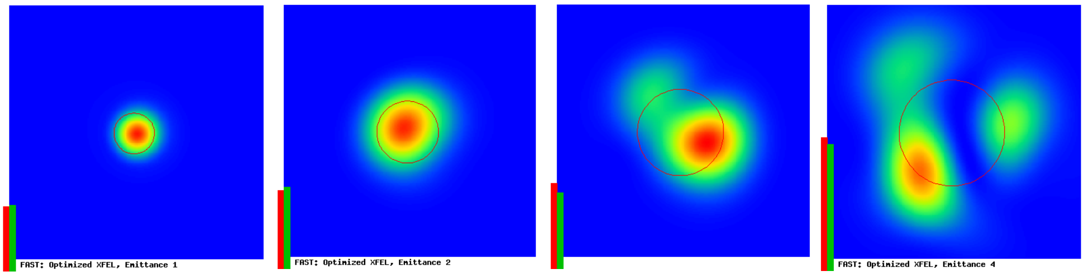


Fig. 54: Typical slice distribution of radiation intensity for optimized SASE FEL with (left to right) $\hat{\epsilon} = 1, 2, 3, 4$. Circle denotes r.m.s. spot size. SASE FEL operates at saturation. Simulations run with code FAST [77].

Table 1: Parameter space of X-ray FELs

	LCLS	SACLA	EXFEL	SWISS FEL	PAL XFEL
Energy [GeV]	13.6	8.0	17.5	5.8	10
Wavelength [A]	1.5	0.6	0.5	0.7	0.6
ϵ_n [mm-rad]	0.4	0.4	0.4	0.4	0.4
$\hat{\epsilon}$	1	2.7	1.5	3.4	2.1

slices have a rather complicated shape, owing to the interference of the fields of statistically independent modes with different azimuthal indices. These slice distributions are essentially non-Gaussian when the relative contribution of higher azimuthal modes to the total power approaches 10%. The shapes of the intensity and phase distributions change drastically on a scale of the coherence length, and the source point position and pointing jumps from spike to spike. Figure 52 presents a quantitative description of this phenomenon, using the notion of the r.m.s. deviation of the photon beam centre of gravity Δ_θ expressed in terms of the r.m.s. size of the photon beam. We see that there is no perfect pointing of the photon beam, and that for $\hat{\epsilon} \gtrsim 2$ fluctuations of the pointing exceed 40%. Averaging of slice distributions over a radiation pulse results in a smoother distribution. However, with a limited number of longitudinal modes, the centre of gravity of the radiation pulse (position) and its shape jitter from shot to shot; this is frequently referred to as poor pointing stability. This effect has been observed at FLASH [47] experimentally, and should evidently take place at other X-ray facilities. Only in the case of a long radiation pulse, or after averaging over many pulses, does the intensity distribution approach, asymptotically, an azimuthally symmetrical shape.

Table 1 presents a list of parameters of the X-ray FELs compiled for the shortest design wavelength [17–19,22,24]. We assume the normalized emittance to be the same for all cases ($\epsilon_n = 0.4$ mm-mrad). A lower electron beam energy results in a larger value of the emittance parameter, and the output radiation will have poor spatial coherence and poor pointing stability of the photon beam. Note that the spatial jitter is of a fundamental nature (shot noise in the electron beam), and takes place even for an 'ideal' machine.

There are very limited means of suppressing the mode degeneration effect by controlling the spread of longitudinal velocities (due to energy spread and emittance) [6, 51, 68]. Energy spread can be increased with a laser heater [84]. The price for this improvement is a significant increase in saturation length and reduction of the FEL power. Stronger focusing of the electron beam in the undulator helps to improve transverse coherence by reducing the diffraction parameter and increase of the velocity spread. However, this will also result in the increase in saturation length. Finally, with fixed energy of the electron beam, an available undulator length will define the level of a spatial coherence and spatial jitter of the photon beam.

References

[1] A.M. Kondratenko and E.L. Saldin, *Part. Accelerators* **10** (1980) 207.

[2] Ya.S. Derbenev *et al.*, *Nucl. Instrum. Methods* **193** (1982) 415. https://doi.org/10.1016/0029-554X(82)90233-6

[3] J.B. Murphy and C. Pellegrini, *Nucl. Instrum. Methods* **A237** (1985) 159. https://doi.org/10.1016/0168-9002(85)90344-4

[4] R. Bonifacio *et al.*, *Opt. Commun.* **50** (1984) 373. https://doi.org/10.1016/0030-4018(84)90105-6

[5] R. Bonifacio *et al.*, *Phys. Rev. A* **33** (1986) 2836. https://doi.org/10.1103/PhysRevA.33.2836

[6] E.L. Saldin *et al.*, *The Physics of Free Electron Lasers* (Springer-Verlag, Berlin, 1999).

[7] M. Hogan *et al.*, *Phys. Rev. Lett.* **81** (1998) 4867. https://doi.org/10.1103/PhysRevLett.81.4867

[8] S. V. Milton *et al.*, *Science* **292** (2001) 2037. https://doi.org/10.1126/science.1059955

[9] A. Tremaine *et al.*, *Nucl. Instrum. Methods* **A483** (2002) 24. https://doi.org/10.1016/S0168-9002(02)00280-2

[10] V. Ayvazyan *et al.*, *Phys. Rev. Lett.* **88** (2002) 104802. https://doi.org/10.1103/PhysRevLett.88.104802

[11] V. Ayvazyan *et al.*, *Eur. Phys. J.* **D20** (2002) 149. https://doi.org/10.1140/epjd/e2002-00121-4

[12] V. Ayvazyan *et al.*, *Eur. Phys. J.* **D37** (2006) 297. https://doi.org/10.1140/epjd/e2005-00308-1

[13] W. Ackermann *et al.*, *Nature Photonics* **1** (2007) 336. https://doi.org/10.1038/nphoton.2007.76

[14] K. Tiedtke *et al.*, *New J. Phys.* **11** (2009) 023029. https://doi.org/10.1088/1367-2630/11/2/023029

[15] S. Schreiber and B. Faatz, *High Power Laser Sci. Eng.* **3** (2015) e20. https://doi.org/10.1017/hpl.2015.16

[16] B. Faatz *et al.*, *New J. Phys.* **18** (2016) 062002. https://doi.org/10.1088/1367-2630/18/6/062002

[17] P. Emma *et al.*, *Nature Photonics* **4** (2010) 641. https://doi.org/10.1038/nphoton.2010.176

[18] T. Ishikawa *et al.*, *Nature Photonics* **6** (2012) 540. https://doi.org/10.1038/nphoton.2012.141

[19] H.S. Kang *et al.*, Current Status of PAL-XFEL Project, Proceedings of the 5th International Particle Accelerator Conference (IPAC 2014), Dresden, Germany, June 15-20, 2014, Christine Petit-Jean-Genaz, Gianluigi Arduini, Peter Michel, Volker R.W. Shaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper THPRO019 (2014).

[20] Jang-Hui Han for PAL-XFEL Beam Commissioning Team, Beam Commissioning of PAL-XFEL, Proceedings of the 7th International Particle Accelerator Conference (IPAC 2016), Busan, Korea, May 8-13, 2016, Kyung Sook Kim, In Soo Ko, Kyung Ryul Kim, Volker R.W. Schaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper MOXBA01 (2016).

[21] D.E. Kim *et al.*, Development of PAL-XFEL Undulator System, Proceedings of the 7th International Particle Accelerator Conference (IPAC 2016), Busan, Korea, May 8-13, 2016, Kyung Sook Kim, In Soo Ko, Kyung Ryul Kim, Volker R.W. Schaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper THPOW45 (2016).

[22] M. Altarelli, R. Brinkmann, M. Chergui, W. Decking, B. Dobson, S. Dusterer, G. Grubel, W. Graeff, H. Graafsma, J. Hajdu, J. Marangos, J. Pfluger, H. Redlin, D. Riley, I. Robinson, J. Rossbach, A. Schwarz, K. Tiedtke, T. Tschentscher, I. Vartaniants, H. Wabnitz, H. Weise, R. Wichmann, K. Witte, A. Wolf, M. Wulff, M. Yurkov (Eds.), The European X-Ray Free-Electron Laser: Technical design report, Preprint DESY 2006-097, DESY, Hamburg, 2006 (see also http://xfel.desy.de).

[23] F. Brinker *et al.*, Commissioning of the European XFEL Injector, Proceedings of the 7th International Particle Accelerator Conference (IPAC 2016), Busan, Korea, May 8-13, 2016, Kyung Sook Kim, In Soo Ko, Kyung Ryul Kim, Volker R.W. Schaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper TUOCA03 (2016).

[24] R. Ganter (Ed.), Swiss FEL, Conceptual Design Report, PSI Bericht Nr. 10-04 (April 2012).

[25] E. Allaria *et al.*, *Nature Photonics* **6** (2012) 699. https://doi.org/10.1038/nphoton.2012.233
[26] K.J. Kim, *Nucl. Instrum. Methods* **A250** (1986) 396. https://doi.org/10.1016/0168-9002(86)90916-2
[27] J.M. Wang and L.H. Yu, *Nucl. Instrum. Methods* **A250** (1986) 484. https://doi.org/10.1016/0168-9002(86)90928-9
[28] W.B. Colson, in *Laser Handbook*, Eds. W.B. Colson *et al.* (North-Holland, Amsterdam, 1990), Vol. 6, p. 115.
[29] R. Bonifacio *et al.*, *Phys. Rev. Lett.* **73** (1994) 70. https://doi.org/10.1103/PhysRevLett.73.70
[30] P. Pierini and W. Fawley, *Nucl. Instrum. Methods* **A375** (1996) 332. https://doi.org/10.1016/0168-9002(95)01221-4
[31] E. L. Saldin *et al.*, *Nucl. Instrum. Methods* **A393** (1997) 157. https://doi.org/10.1016/S0168-9002(97)00450-6
[32] E. L. Saldin *et al.*, *Opt. Commun.* **148** (1998) 383. https://doi.org/10.1016/S0030-4018(97)00670-6
[33] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A507** (2003) 101. https://doi.org/10.1016/S0168-9002(03)00847-7
[34] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A562** (2006) 472. https://doi.org/10.1016/j.nima.2006.02.166
[35] S. Krinsky and R.L. Gluckstern, *Phys. Rev. ST Accel. Beams* **6** (2003) 050701. https://doi.org/10.1103/PhysRevSTAB.6.050701
[36] S. Krinsky and Y. Li, *Phys. Rev.* **E73** (2006) 066501. https://doi.org/10.1103/PhysRevE.73.066501
[37] L.H. Yu and S. Krinsky, *Nucl. Instrum. Methods* **A285** (1989) 119. https://doi.org/10.1016/0168-9002(89)90436-1
[38] S. Krinsky and L.H. Yu, *Phys. Rev. A* **35** (1987) 3406. https://doi.org/10.1103/PhysRevA.35.3406
[39] E.L. Saldin *et al.*, *Opt. Commun.* **186** (2000) 185. https://doi.org/10.1016/S0030-4018(00)01061-0
[40] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A507** (2003) 106. https://doi.org/10.1016/S0168-9002(03)00848-9
[41] E.L. Saldin *et al.*, *Opt. Commun.* **281** (2008) 1179. https://doi.org/10.1016/j.optcom.2007.10.044
[42] E.L. Saldin *et al.*, *New J. Phys.* **12** (2010) 035010. https://doi.org/10.1088/1367-2630/12/3/035010
[43] E.L. Saldin *et al.*, *Opt. Commun.* **281** (2008) 4727. https://doi.org/10.1016/j.optcom.2008.05.033
[44] G. Geloni *et al.*, *New J. Phys.* **12** (2010) 035021. https://doi.org/10.1088/1367-2630/12/3/035021
[45] Z. Huang and K.-J. Kim, *Phys. Rev. ST Accel. Beams* **10** (2007) 034801. https://doi.org/10.1103/PhysRevSTAB.10.034801
[46] V. Ayvazyan *et al.*, *Nucl. Instrum. Methods* **A507** (2003) 368. https://doi.org/10.1016/S0168-9002(03)00946-X
[47] R. Ischebeck *et al.*, *Nucl. Instrum. Methods* **A507** (2003) 175. https://doi.org/10.1016/S0168-9002(03)00866-0
[48] M. Dohlus *et al.*, *Nucl. Instrum. Methods* **A530** (2004) 217. https://doi.org/10.1016/j.nima.2004.04.221
[49] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A562** (2006) 472. https://doi.org/10.1016/j.nima.2006.02.166
[50] E.A. Schneidmiller and M.V. Yurkov, *J. Mod. Opt.* **63** (2016) 293. https://doi.org/10.1080/09500340.2015.1066456
[51] E.A. Schneidmiller and M.V. Yurkov, Fundamental limitations of the SASE FEL photon beam pointing stability, Proceedings of the 37th International Free Electron Laser Conference (FEL 2015), Daejeon, Korea, Heung-Sik Kang, Dong Eon Kim, Schaa, R.W. Volker (Eds.), JACoW: Geneva, Switzerland, paper TUP021 (2015).

[52] E.A. Schneidmiller and M.V. Yurkov, Transverse coherence and fundamental limitatiton on the pointing stability of X-ray FELs, Proceedings of the 7th International Particle Accelerator Conference (IPAC 2016), Busan, Korea, May 8-13, 2016, Kyung Sook Kim, In Soo Ko, Kyung Ryul Kim, Volker R.W. Schaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper MOPOW012 (2016).

[53] A.J. Schmitt and J.H. Gardner, *Phys. Rev. A* **60** (1986) 6. https://doi.org/10.1063/1.337631

[54] R. Bonifacio *et al.*, *Nucl. Instrum. Methods* **A293** (1990) 627. https://doi.org/10.1016/0168-9002(90)90334-3

[55] W.M. Fawley *et al.*, Merits of a Sub-Harmonic Approach to a Single-Pass, 1.5-A FEL, Proceedings of the 1995 Particle Accelerator Conference, May 1-5, 1995, Dallas, Texas, USA, L.T. Gennari, R.H. Siemann (Eds.), IEEE, Piscataway, NJ, p. 219 (1995).

[56] Z. Huang and K. Kim, *Phys. Rev. E* **62** (2000) 7295. https://doi.org/10.1103/PhysRevE.62.7295

[57] Z. Huang and K. Kim, *Nucl. Instrum. Methods* **A475** (2001) 112. https://doi.org/10.1016/S0168-9002(01)01553-4

[58] H. Freund *et al.*, *Nucl. Instrum. Methods* **A445** (2000) 53. https://doi.org/10.1016/S0168-9002(00)00113-3

[59] H. Freund *et al.*, *IEEE J. Quant. Electr.* **36** (2000) 275. https://doi.org/10.1109/3.825873

[60] S. Biedron *et al.*, *Nucl. Instrum. Methods* **A483** (2002) 94. https://doi.org/10.1016/S0168-9002(02)00430-8

[61] S. Biedron *et al.*, *Phys. Rev. ST Accel. Beams* **5** (2002) 030701. https://doi.org/10.1103/PhysRevSTAB.5.030701

[62] A. Tremaine *et al.*, *Phys. Rev. Lett.* **88** (2002) 204801. https://doi.org/10.1103/PhysRevLett.88.204801

[63] W. Brefeld *et al.*, *Nucl. Instrum. Methods* **A507** (2003) 431. https://doi.org/10.1016/S0168-9002(03)00960-4

[64] E.L. Saldin *et al.*, *Phys. Rev. ST Accel. Beams* **9** (2006) 030702. https://doi.org/10.1103/PhysRevSTAB.9.030702

[65] E.A. Schneidmiller and M.V. Yurkov, Coherence properties of the odd harmonics of the radiation from SASE FEL with planar undulator, Proceedings of the 34th International Free Electron Laser Conference (FEL 2012), Nara, Japan, August 26-31, 2012, Toshinari Tanaka, Volker R. W. Schaa (Eds.), JACoW : Geneva, Switzerland, paper MOPD08 (2012).

[66] E.A. Schneidmiller and M.V. Yurkov, *Phys. Rev. ST Accel. Beams* **15** (2012) 080702. https://doi.org/10.1103/PhysRevSTAB.15.080702

[67] G. Moore, *Opt. Commun.* **52** (1984) 46. https://doi.org/10.1016/0030-4018(84)90071-3

[68] E.L. Saldin *et al.*, *Opt. Commun* **97** (1994) 272. https://doi.org/10.1016/0030-4018(93)90152-U

[69] M. Xie, *Nucl. Instrum. Methods* **A445** (2000) 59. https://doi.org/10.1016/S0168-9002(00)00114-5

[70] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A475** (2001) 86. https://doi.org/10.1016/S0168-9002(01)01549-2

[71] K.J. Kim, *Phys. Rev. Lett.* **57** (1986) 1871. https://doi.org/10.1103/PhysRevLett.57.1871

[72] J.M. Wang and L.H. Yu, *Nucl. Instrum. Methods* **A250** (1986) 484. https://doi.org/10.1016/0168-9002(86)90928-9

[73] S. Krinsky and L.H. Yu, *Phys. Rev. A* **35** (1987) 3406. https://doi.org/10.1103/PhysRevA.35.3406

[74] J. Goodman, *Statistical Optics* (John Wiley and Sons, New York, 1985).

[75] G. Geloni *et al.*, *Opt. Commun.* **271** (2007) 207. https://doi.org/10.1016/j.optcom.2006.10.019

[76] L. Mandel, *Proc. Phys. Soc. London* **74** (1959) 223. https://doi.org/10.1088/0370-1328/74/3/301

[77] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A429** (1999) 233.
https://doi.org/10.1016/S0168-9002(99)00110-2

[78] W.M. Fawley, *Phys. Rev. ST Accel. Beams* **5** (2002) 070701.
https://doi.org/10.1103/PhysRevSTAB.5.070701

[79] H. Wiedemann, *Synchrotron Radiation* (Springer-Verlag, Berlin, 2003).
https://doi.org/10.2172/815294

[80] E.L. Saldin *et al.*, *Opt. Commun.* **235** (2004) 415. https://doi.org/10.1016/j.optcom.2004.02.071

[81] D. Ratner *et al.*, *Phys. Rev. ST Accel. Beams* **14** (2011) 060701.
https://doi.org/10.1103/PhysRevSTAB.14.060701

[82] W.P. Robins, *Phase Noise in Signal Sources: Theory and Applications* (Peter Peregrinus, London, 1982).

[83] E.L. Saldin *et al.*, *Opt. Commun.* **202** (2002) 169. https://doi.org/10.1016/S0030-4018(02)01091-X

[84] E.L. Saldin *et al.*, *Nucl. Instrum. Methods* **A528** (2004) 355.
https://doi.org/10.1016/j.nima.2004.04.067

[85] E.A. Schneidmiller and M.V. Yurkov, Spatial properties of the radiation from SASE FELa at the European XFEL, Proceedings of the 34th International Free Electron Laser Conference (FEL 2012), Nara, Japan, August 26-31, 2012, Toshinari Tanaka, Volker R. W. Schaa (Eds.), JACoW : Geneva, Switzerland, paper MOPD06 (2012).

[86] C. Behrens *et al.*, *Phys. Rev. ST Accel. Beams* **15** (2012) 030707.
https://doi.org/10.1103/PhysRevSTAB.15.030707

[87] E.A. Schneidmiller and M.V. Yurkov, Application of statistical methods for measurements of the coherence properties of the radiation from SASE FEL, Proceedings of the 7th International Particle Accelerator Conference (IPAC 2016), Busan, Korea, May 8-13, 2016, Kyung Sook Kim, In Soo Ko, Kyung Ryul Kim, Volker R.W. Schaa (Eds.), JACoW conferences, Geneva, Switzerland: CERN, paper MOPOW013 (2016).

[88] L. Bittner *et al.*, MCP-based photon detector with extended wavelength range for FLASH, Proceedings of the 29th International Free Electron Laser Conference (FEL 2007), Novosibirsk, Russia, August 26-31, 2007, G.N. Kulipanov, N.A. Vinokurov (Eds.), JACoW : Geneva, Switzerland, paper WEPPH007 (2007).

[89] J. Wu *et al.*, LCLS X-ray pulse duration measurement using the statistical fluctuation method, Proceedings of the 32nd International Free Electron Laser Conference (FEL 2010), Malmo, Sweden, August 23-27, 2010, Leif Liljeby (Ed.), JACoW : Geneva, Switzerland, paper MOPC14 (2010).

Appendix

A Estimates of the radiation properties in the framework of the one-dimensional model

In this section, we combine the most essential formulae of the one-dimensional theory of the FEL amplifier, describing the main characteristics of the radiation. The main physical parameter of the problem is the FEL parameter ρ (see Eq. (11)) [4, 6]:

$$\rho = \frac{\lambda_{\mathrm{w}}}{4\pi}\left[\frac{4\pi^2 j_0 K^2 A_{\mathrm{JJ}}^2}{I_A \lambda_{\mathrm{w}} \gamma^3}\right]^{1/3} ,$$

where $j_0 = I/(2\pi\sigma^2)$ is the beam current density, $\sigma = \sqrt{\beta\epsilon_{\mathrm{n}}/\gamma}$ is the r.m.s. transverse size of the electron beam, and β is the external focusing beta function. The FEL parameter ρ relates to the efficiency parameter of the three-dimensional FEL theory as $\rho = \bar{\rho}/B^{1/3}$. Basic characteristics of the SASE FEL are estimated in terms of the parameter ρ and the number of cooperating electrons $N_{\mathrm{c}} = I/(e\rho\omega)$. Here we, present a set of simple formulae extracted from [6,29,32]:

$$
\begin{aligned}
&\text{Field gain length:} && L_{\mathrm{g}} \sim \frac{\lambda_{\mathrm{w}}}{4\pi\rho} \,, \\
&\text{Saturation length:} && L_{\mathrm{sat}} \sim \frac{\lambda_{\mathrm{w}}}{4\pi\rho}\left[3+\frac{\ln N_{\mathrm{c}}}{\sqrt{3}}\right] \\
&\text{Effective power of shot noise:} && \frac{P_{\mathrm{sh}}}{\rho P_{\mathrm{b}}} \simeq \frac{3}{N_{\mathrm{c}}\sqrt{\pi \ln N_{\mathrm{c}}}} \,, \\
&\text{Saturation efficiency:} && \rho \,, \\
&\text{Power gain at saturation:} && G \simeq \frac{1}{3} N_{\mathrm{c}}\sqrt{\pi \ln N_{\mathrm{c}}} \,, \\
&\text{Coherence time at saturation:} && \tau_{\mathrm{c}} \simeq \frac{1}{\rho\omega}\sqrt{\frac{\pi \ln N_{\mathrm{c}}}{18}} \,, \\
&\text{Spectrum bandwidth:} && \sigma_{\omega} = \sqrt{\pi}/\tau_{\mathrm{c}} \,.
\end{aligned}
\tag{A.1}
$$

The contribution of the higher odd harmonics to the total power of SASE FEL operating at saturation is described by Eq. (15). The coherence time at saturation is inversely proportional to the harmonic number, and the relative spectrum bandwidth remains constant with harmonic number.

B Estimates of radiation properties in the framework of the three-dimensional theory

Here, we present practical formulae that enable calculation of the parameters of a FEL amplifier optimized for maximum gain of the ground TEM_{00} radiation mode. The minimum gain length L_{g} of the fundamental mode and the corresponding optimum focusing beta function β_{opt} are [66,80]:

$$
\begin{aligned}
L_{\mathrm{g0}} &= 1.67\left(\frac{I_{\mathrm{A}}}{I}\right)^{1/2}\frac{(\epsilon_n\lambda_{\mathrm{w}})^{5/6}}{\lambda^{2/3}}\,\frac{(1+K^2)^{1/3}}{KA_{\mathrm{JJ}}}(1+\delta) \,, \\
\beta_{\mathrm{opt}} &\simeq 11.2\left(\frac{I_{\mathrm{A}}}{I}\right)^{1/2}\frac{\epsilon_n^{3/2}\lambda_{\mathrm{w}}^{1/2}}{\lambda K A_{\mathrm{JJ}}}\,(1+8\delta)^{-1/3} \\
\delta &= 131\,\frac{I_{\mathrm{A}}}{I}\,\frac{\epsilon_n^{5/4}}{\lambda^{1/8}\lambda_{\mathrm{w}}^{9/8}}\,\frac{\sigma_\gamma^2}{(KA_{\mathrm{JJ}})^2(1+K^2)^{1/8}} \,.
\end{aligned}
\tag{B.1}
$$

When (for technical reasons) focusing beta is $\beta > \beta_{\mathrm{opt}}$, the gain length is:

$$
L_{\mathrm{g}}(\beta) \simeq L_{\mathrm{g}}(\beta_{\mathrm{opt}})\left[1+\frac{(\beta-\beta_{\mathrm{opt}})^2(1+8\delta)}{4\beta_{\mathrm{opt}}^2}\right]^{1/6} \,, \qquad \text{for } \beta > \beta_{\mathrm{opt}}. \tag{B.2}
$$

Equation (B.1) provides an accuracy better than 5% in the range of parameters given by Eq. (30). For small energy spread, $\delta \ll 1$, the diffraction parameter B, the parameter of betatron oscillations $\hat{k}_\beta$, the reduced saturation length $\hat{L}_{\mathrm{sat}}$, the reduced FEL efficiency $\hat{\eta}$, the reduced coherence time τ_{c}, the radiation bandwidth σ_ω, and the degree of transverse coherence ζ are functions of the parameter $\hat{\epsilon} = 2\pi\epsilon/\lambda$ and the number of electrons in the volume of coherence $N_{\mathrm{c}} = IL_{\mathrm{g}}\lambda/(\lambda_{\mathrm{w}}c)$:

$$
\begin{aligned}
B &= \frac{2\bar{\Gamma}\sigma^2\omega}{c} \simeq 12.5\times\hat{\epsilon}^{5/2} \,, \\
\hat{k}_\beta &= \frac{1}{(\beta\bar{\Gamma})} \simeq 0.158\times\hat{\epsilon}^{-3/2} \,, \\
\hat{L}_{\mathrm{sat}} &= \bar{\Gamma}L_{\mathrm{sat}} \simeq 2.5\times\hat{\epsilon}^{5/6}\times\ln N_{\mathrm{c}} \,, \\
\hat{\eta} &= \frac{P}{(\bar{\rho}P_{\mathrm{b}})} \simeq 0.17/\hat{\epsilon} \,,
\end{aligned}
$$

$$\hat{\tau}_{\mathrm{c}} = \bar{\rho}\omega\tau_{\mathrm{c}} \simeq 1.16 \times \sqrt{\ln N_{\mathrm{c}}} \times \hat{\epsilon}^{5/6} ,$$
$$\sigma_{\omega} = \frac{\sqrt{\pi}}{\tau_{\mathrm{c}}} ,$$
$$\zeta_{\mathrm{sat}} \simeq \frac{1.1\hat{\epsilon}^{1/4}}{1 + 0.15\hat{\epsilon}^{9/4}} . \tag{B.3}$$

The gain parameter $\bar{\Gamma}$ and efficiency parameter $\bar{\rho}$ are given by:

$$\bar{\Gamma} = \left[\frac{I}{I_{\mathrm{A}}} \frac{8\pi^2 K^2 A_{\mathrm{JJ}}^2}{\lambda\lambda_{\mathrm{w}}\gamma^3} \right]^{1/2} , \qquad \bar{\rho} = \frac{\lambda_{\mathrm{w}}\bar{\Gamma}}{4\pi} .$$

Equations (B.1) and (B.2) should be used for the case of a finite value of the energy spread and non-optimum beta function. Correction to the coherence time and spectrum bandwidth scales as relative increase of the gain length. Spot size and angular divergence can be calculated following Ref. [85].

Properties of the higher odd harmonics for relative contributions to the total radiation power and spectral characteristics are similar to those found in the framework of the one-dimensional model. The contribution of the higher odd harmonics to the total power at the saturation is close to that given in Eq. (15) with relevant correction on effective energy spread. The coherence time at saturation is inversely proportional to harmonic number, and the relative spectrum bandwidth remains constant with harmonic number.

C Application of statistical methods for measurements of the coherence properties of the radiation from SASE FEL

Radiation from the SASE FEL operating in the linear regime has the properties of completely chaotic polarized light. Measurements of the SASE FEL gain curve enable determination of the saturation length, which is strictly connected with the coherence time. Statistical analysis of the fluctuations of the radiation energies measured with different spatial apertures allows one to determine the number of the longitudinal and transverse modes. Thus, with these simple measurements, it becomes possible to determine the degree of transverse coherence, coherence time, and photon pulse duration. In this section, we present the theoretical background and experimental results obtained at the FLASH FEL.

The amplification process in a SASE FEL starts from the shot noise in the electron beam. Initially poor, the coherence properties of the radiation are significantly improved in the exponential stage of amplification, and reach their best values at the onset of the saturation regime [41,42]. Radiation of SASE FEL consists of wavepackets (spikes) having durations of about the coherence time. Fields $\tilde{E}$ are well correlated within one spike, and are statistically independent for different spikes. Coherence properties of the radiation are described with temporal and spatial correlation functions (g_1 and γ_1), coherence time, and degree of transverse coherence (τ_{c} and ζ), given by Eq. (7). Radiation from a SASE FEL operating in the linear regime has properties of completely chaotic polarized light [6, 32], and the probability distribution of the radiation energy is a gamma distribution (Eq. (3)). This distribution is the function of the only parameter—the number of radiation modes, M. The parameter M is equal to the inverse squared value of the standard deviation of the radiation energy, $M = 1/\sigma_E^2$, and $\sigma_E^2 = \langle (E - \langle E \rangle)^2 \rangle / \langle E \rangle^2$.

We consider the electron bunch with Gaussian longitudinal profile of r.m.s. pulse duration σ_{z}. Figure 20 shows the evolution along the undulator of the radiation pulse energy, fluctuations, and r.m.s. photon pulse length. Normalized values of these parameters exhibit nearly universal dependencies for $\rho\omega\sigma_{\mathrm{z}} \gtrsim 1$. A maximum of fluctuations and a minimum of the pulse duration are obtained at the end of the exponential gain regime. The saturation point (corresponding to maximum brilliance of the radiation [41]) is defined by the condition for fluctuations to decrease by a factor of three with respect to the maximum value. In the framework of the one-dimensional model, the maximum value of the coherence time and saturation length,

$$(\tau_{\mathrm{c}})_{\max} \simeq \frac{1}{\rho\omega}\sqrt{\frac{\pi \ln N_{\mathrm{c}}}{18}} \, , \qquad L_{\mathrm{sat}} \simeq \frac{\lambda_{\mathrm{w}}}{4\pi\rho}\left(3 + \frac{\ln N_{\mathrm{c}}}{\sqrt{3}}\right) \, ,$$

are expressed in terms of the FEL parameter ρ [4] and the number of cooperating electrons $N_{\mathrm{c}} = I/(e\rho\omega)$ [6, 32]. Here, $\omega = 2\pi c/\lambda$ is the frequency of the amplified wave, I is the beam current, $-e$ is the charge of an electron, and λ_{w} is the undulator period. A practical estimate for parameter ρ comes from the observation that, in the parameter range of SASE FELs operating in the vacuum ultraviolet and X-ray wavelength ranges, the number of field gain lengths to saturation is about 10 [32]. Thus, the parameter ρ and coherence time τ_{c} relate to the saturation length as:

$$\rho \simeq \lambda_{\mathrm{w}}/L_{\mathrm{sat}} \, , \qquad \tau_{\mathrm{c}} \simeq \lambda L_{\mathrm{sat}}/(2\sqrt{\pi}c\lambda_{\mathrm{w}}) \, . \tag{C.1}$$

For the number of modes $M \gtrsim 2$, the r.m.s. electron pulse length and minimum FWHM radiation pulse length $\tau_{\mathrm{ph}}^{\min}$ at the end of the linear regime are given by [86, 87]:

$$\tau_{\mathrm{ph}}^{\min} \simeq \sigma_z \simeq \frac{M\lambda}{5\rho} \simeq \frac{M\lambda L_{\mathrm{sat}}}{5c\lambda_{\mathrm{w}}} \, . \tag{C.2}$$

The minimum radiation pulse duration expressed in terms of coherence time (Eq. (C.1)) is $\tau_{\mathrm{ph}}^{\min} \simeq 0.7 \times M \times \tau_{\mathrm{c}}$.

Lengthening of the radiation pulse occurs when the amplification process enters the saturation regime. This happens as a result of two effects. The first effect is lasing to saturation of the tails of the electron bunch, and the second effect is pulse lengthening due to slippage effects (one radiation wavelength per undulator period). The effect of lasing tails gives the same relative radiation pulse lengthening as illustrated in Fig. 20. At the saturation point, pulse lengthening is about a factor of 1.4 with respect to the minimum pulse for the linear regime given by Eq. (C.1), and it is increased by up to a factor of two in the deep non-linear regime. The slippage effect is more pronounced for relative lengthening of short pulses.

The total number of modes in the radiation pulse is the product of the number of longitudinal and transverse modes, $M_{\mathrm{Total}} = M_{\mathrm{Long}} \times M_{\mathrm{Trans}}$. This is the origin of an idea to use measurements of the fluctuations of the radiation pulse energy to derive the degree of the transverse coherence. Measurements of the fluctuations of the total pulse energy and of the radiation energy after a pinhole gives us the total number of modes M_{Total}, and the number of longitudinal modes M_{Long}, respectively. Their ratio gives the number of transverse modes $M_{\mathrm{Trans}} = M_{\mathrm{Total}}/M_{\mathrm{Long}}$. The degree of transverse coherence is equal to the inverse value of the number of transverse modes [41]:

$$\zeta = \frac{1}{M_{\mathrm{Trans}}} = \frac{M_{\mathrm{Long}}}{M_{\mathrm{Total}}} \, . \tag{C.3}$$

Numerical simulations with code FAST [77] confirm this physical consideration. We see from Fig. C.1 that, in the exponential gain regime, the squared ratio of the fluctuations exactly follows the degree of transverse coherence calculated with rigorous statistical definition (Eq. (7)). We should stress that simple statistical measurements give the fundamental quantity without making any additional assumptions. This happens as a result of the fundamental nature (Gaussian statistics) of the light produced by the SASE FEL in the exponential gain regime. Pinhole techniques enable the evolution of the degree of transverse coherence to be traced, up to the onset of the saturation regime.

The experimental technique is as follows. The gain curve of the SASE FEL is measured at the first step (average radiation energy and fluctuations versus undulator length), and the saturation length is determined. A practical hint for determination of the saturation point is the decrease of the fluctuations by a factor of three with respect to the maximum value. Using the value of the saturation length, we derive

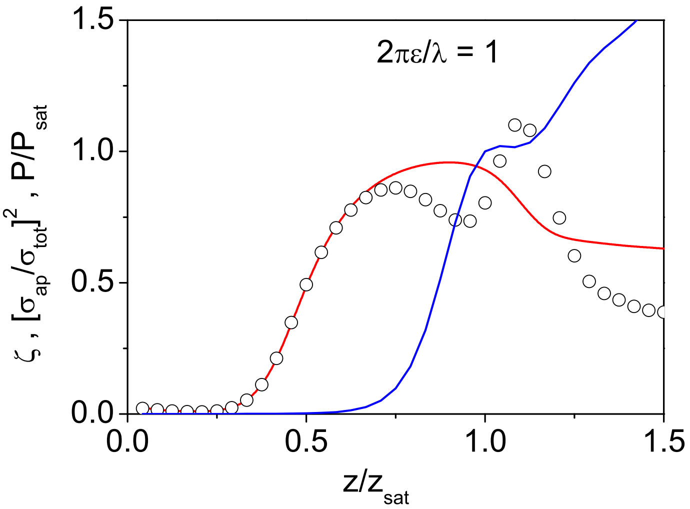


Fig. C.1: Evolution of (red) degree of transverse coherence ζ and (blue) FEL power P. Circles show ratio of fluctuations of radiation energy in pinhole to fluctuations of total energy, $\sigma^2_{E,\mathrm{ap}}/\sigma^2_{E,\mathrm{tot}}$. Simulations run with code FAST [77].

the FEL parameter ρ and coherence time τ_c (Eq. (C.1)). Then, the FEL process is stopped at the end of the high-gain linear regime (the FEL power is a factor of 20 less than the saturation level, see Fig. 20). Fluctuations are measured by using a pinhole aperture to select the central part of the photon pulse. Essential electron beam and machine parameters (charge, orbit, compression signal, RF parameters) for each shot are also recorded. The final step of the experimental procedure is the gating of the experimental results with machine parameters. The final data set contains mainly fundamental fluctuations of the radiation pulse energy related to SASE FEL process. The inverse squared value of fluctuations gives the number of longitudinal modes M_Long. The radiation pulse length is derived from Eq. (C.2). Additional measurements of the fluctuations for full radiation pulse energy enable determination of the number of total and transverse modes and the degree of transverse coherence (Eq. (C.3)).

Here we present the results of experimental measurements at FLASH of the number of the number of modes, coherence time, and degree of transverse coherence [86,87]. FLASH is equipped with a set of detectors for measurements of the energy in the radiation pulse: gas monitor detector, micro channel plate (MCP) based detector, photodiode, and thermopile [14,88]. Detectors are installed in several positions along the photon beam line. The MCP detector is installed in front of all the other detectors and is used for precise measurements of the radiation pulse energy. The MCP measures the radiation scattered by a metallic mesh (Cu, Fe, and Au targets are used) placed behind an aperture located 18.5 m downstream of the undulator. The electronics of the MCP detector itself have low noise, about 1 mV at the level of signal of 100 mV (1% relative measurement accuracy).

C.1 Measurements of the number of modes and pulse duration at FLASH [86]

The measurement procedure is organized as follows. We tune the SASE process to the maximum signal at a full undulator length of 27 m (six undulator modules). Then we apply an orbit kick (by means of switching on steerers) after the fourth undulator module, such that the FEL amplification process is suppressed in the last two undulator modules. The level of the radiation pulse energy after four undulator modules is about a factor of 20 less than the saturation level. This point corresponds to the end of the high-gain exponential regime with a minimum photon pulse length. To evaluate the number of longitudinal modes, we put a small (1 mm) aperture centred on the photon beam. Then we record radiation pulse energies (readings of the MCP detector) at this point. However, this is not all: fluctuations of the

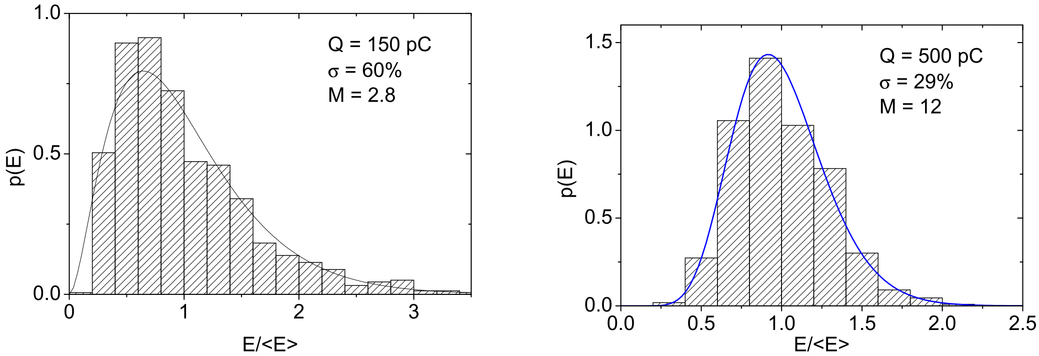


Fig. C.2: Experimental results from FLASH: probability distributions for energy in radiation pulses at end of exponential growth regime. Radiation wavelength, 13.5 nm. Left: 150 pC bunch charge. Right: 500 pC bunch charge. Solid lines show gamma distribution with $M = 2.8$ and $M = 12$ for 150 pC and 500 pC bunch charges, respectively.

electron beam and machine parameters may contribute to the fluctuations of the radiation pulse energy, but only fundamental SASE FEL fluctuations are essential for us. The final step of the experimental procedure is gating of the experimental results, enabling selection of only fundamental fluctuations of the radiation pulse energy related to the SASE FEL process. To do this, we record essential electron beam and machine parameters (bunch charge monitor readings, beam position monitor readings, readings of the pyroelectric detectors, read back values of RF parameters) for each shot, together with the readings of the MCP detector. Fluctuations of machine parameters are detected with this set of readings. If the machine (or electron beam) parameters deviate beyond a prescribed threshold, this event is excluded from data set. The number of events after the selection procedure must be sufficiently large, to provide the required statistical accuracy.

Experiments were performed for two different tunings of the beam formation system: a short pulse with 150 pC bunch charge, and a long pulse with 500 pC bunch charge. Figure C.2 shows the probability distribution of the radiation energy. Sets of raw data contained about 1400 (1700) measurements, and about 800 (550) measurements remained after the selection procedure for the 150 pC (500 pC) bunch charge. The measured number of radiation modes is 2.8 and 12 for the 150 pC and 500 pC bunch charges, respectively. Within current experimental conditions, the saturation length is estimated to be about 22 m for both cases of the bunch charge. Assuming the Gaussian shape of the lasing fraction of the electron bunch, we apply Eq. (C.2) to determine the r.m.s. electron pulse length, and obtain the following values: $6.2\,\mu\mathrm{m}$ (20 fs) and $26\,\mu\mathrm{m}$ (86 fs) for the 150 pC and 500 pC bunch charges, respectively. The FWHM pulse duration of the photon pulse in the end of the linear regime is approximately the r.m.s. electron pulse duration.

In parallel with statistical measurements, we recorded radiation spectra [86]. In the case of 150 pC, the single-shot spectra were dominated by two or three spikes. There were about ten spikes in the spectrum for the 500 pC bunch charge. Note that the spectrum is simply a Fourier transform of the temporal structure, and that the average number of spikes (modes) in the temporal domain should be about the same as the number of spikes in the spectral domain. These qualitative observations are in a good agreement with the measured number of modes.

C.2 Measurements of the degree of transverse coherence at FLASH [87]

Measurements were made in the framework of the experimental program at FLASH, with the aim of characterizing the transverse coherence of the radiation. Measurements were made in the same way as described before, except with one more statistical run with full pulse energy to define the total number

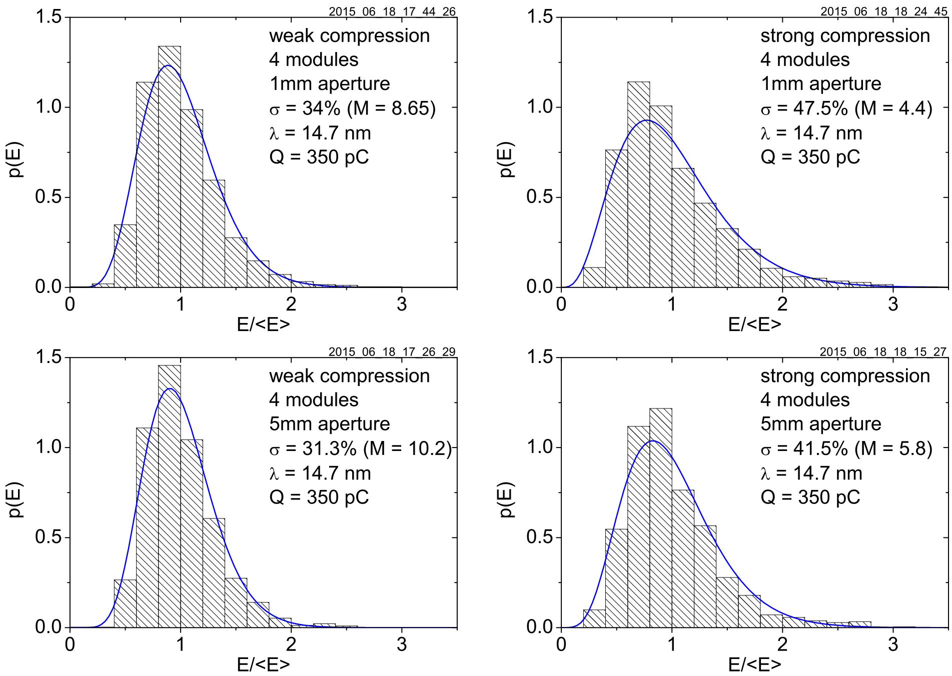


Fig. C.3: Experimental results from FLASH: probability distributions of the radiation energy. Left: weak compression. Right: strong compression. Upper and lower rows relate to pinhole and full pulse energy measurements.

of modes in the radiation pulse. Then the degree of transverse coherence is given by $\zeta = M_{\mathrm{Long}}/M_{\mathrm{tot}}$. Two regimes (weak and strong) of compression were characterized (see Fig. C.3). For weak and strong compression, with 1 mm aperture measurements, we find the number of longitudinal modes M_{Long}, 8.65 and 4.4, respectively. Measurements with full energy gives us the total number of modes, equal to $M_{\mathrm{tot}} = 10.2$ for weak compression, and $M_{\mathrm{tot}} = 5.8$ for strong compression. The ratio of these two measurements gives us the degree of transverse coherence, 85% and 75% for weak and strong compression, respectively. The obtained values are in reasonable agreement for the values expected at FLASH in these parameter range [50].

In conclusion to this section, we can state that statistical measurement is an extremely powerful tool for characterization of the main SASE FEL parameters: the FEL parameter ρ, saturation length, coherence time, photon pulse duration, and degree of transverse coherence. The method is based on the fundamental principles, and measured values have strict physical meaning (Eq. (7)). Statistical measurements have been used at FLASH since the start of its operation [10–13, 86]. There was also a trial experiment at LCLS [89]. However, FLASH is currently the only facility where statistical measurements are routinely used for SASE FEL characterization. Statistical measurements are conceptually simple, but rely on two important technical requirements. The first requirement is availability of a fast and precise radiation detector capable of measuring radiation energy of every pulse with high relative accuracy for a wide range of radiation intensities. At FLASH, we use an MCP detector with relative measurement accuracy better than 1%. The second requirement is small jitter of the machine parameters, much less than the fundamental SASE FEL fluctuations. Good phase stability of the superconducting accelerator FLASH helps a lot. In addition, the success of the technique depends on the quality of diagnostics enabling detection of jitters of the electron beam and machine parameters.

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

The European XFEL—Status and commissioning*†

H. Weise
Deutsches Elektronen-Synchrotron, Hamburg, Germany, on behalf of the European XFEL Accelerator Consortium

Abstract
The European XFEL under construction in Hamburg, Northern Germany, aims to produce X-rays in the range 260 eV to 24 keV using three undulators that can be operated simultaneously with up to 27,000 pulses per second. The FEL is driven by a 17.5 GeV superconducting linac. Installation of this linac is now finished and commissioning will take place next. First lasing is expected for spring 2017. This paper summarizes the status of the project. First results of the injector commissioning are given.

Keywords
European XFEL; superconducting linac; long bunch trains.

1 Introduction

The accelerator complex of the European XFEL [1] is being constructed by an international consortium under the leadership of DESY. Seventeen European research institutes contribute to the accelerator complex and to the comprehensive infrastructure. Major contributions are coming from Russian institutes. DESY co-ordinates the European XFEL Accelerator Consortium but also contributes with many accelerator components and the technical equipment of buildings, with associated general infrastructure. With the finishing of accelerator installation, the commissioning phase is now starting, with cool-down of the main linac scheduled for the end of November 2016.

2 Layout of the European XFEL

In the following, an overall layout of the European XFEL is given, with emphasis on the different sections of the accelerator complex.

2.1 Introduction to the accelerator

The European XFEL, with its total facility length of 3.4 km, follows the established layout of a high-performance single-pass self-amplified spontaneous emission (SASE) FEL. A high-bunch-charge, low-emittance electron gun is followed by some initial acceleration to typically 100 MeV. In the following, magnetic chicanes help to compress the bunch and therefore increase the peak current. This happens at different energies to take care of beam dynamic effects that would deteriorate the bunch emittance in the case of too early compression at too low energies. Thus, the linac is separated by several such chicanes. The European XFEL main linac accelerates the beam in three sections, following the first acceleration in the injector.

*This work is supported by the respective funding agencies of the contributing institutes; for details please see http:www.xfel.eu.
† This paper has already been published in the Proc. of RuPAC2016, St. Petersburg, Russia, 21—25 November 2016, http://accelconf.web.cern.ch/AccelConf/rupac2016/index.html, and it is reproduced with permission from the author.

2519-8041 – © the Author/s, 2018. Published by CERN under the Creative Common Attribution CC BY 3.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.597

2.2 Injector

The injector design of the European XFEL is visibly affected by the need for long bunch trains, which are required for the efficient use of superconducting linac technology. Like many other FELs, it starts with a normal-conducting 1.6 cell radio frequency (RF) electron gun but here the source has to deliver 600 µs long trains, i.e., the RF on time is equivalently long, and not just some few microseconds. The 6 MeV electron beam produced is almost immediately injected into the first superconducting accelerator section, which allows efficient acceleration of bunch trains. This first linac section consists of a standard eight-cavity XFEL module, followed by a harmonic 3.9 GHz module. The latter is needed to manipulate the longitudinal beam profile, together with the later bunch compression in magnetic chicanes. Beam diagnostics are used to verify the electron beam quality at an energy of about 130 MeV. The injector installation, which is 50 m long in total, ends with a beam dump that is able to take the full beam power.

The injector of the European XFEL was commissioned and operated during the installation period of the main linac sections. The first beam was accelerated in December 2015. At the end of the injector, 600 µs long electron bunch trains of typically 500 pC bunches are available, with measured projected emittances of 1–1.5 mm mrad. Most relevant for the FEL process is the slice emittance, which was found to be of the order of 0.5 mm mrad for 500 pC.

The next section downstream of the injector is a warm beamline including a so-called dogleg and the first bunch compressor, for historical reasons named BC0. The dogleg takes care of the vertical offset between the injector tunnel and the main linac tunnel.

Compression in all bunch compressors is obtained by creating different path lengths in a four-dipole magnet chicane. Electrons with slightly lower beam energy are deflected more strongly and thus pass the chicane on an ‘outward curve’. The acceleration in the injector section is slightly off-crest, i.e., the energy of the leading electrons in the bunch is intentionally smaller. The aforementioned 3.9 GHz harmonic system helps to achieve the proper energy modulation along the bunch. Since all electrons have essentially the same speed, the leading ones travel for slightly longer, and the bunch is compressed.

At the XFEL bunch compressor BC0 there is a first slight compression, by roughly a factor of two. The bunches ready for further acceleration reach 1 mm length, ≈100 A peak current, with an energy spread of 1.5% at 130 MeV beam energy.

At present, the European XFEL uses the lower of two injector tunnels. The second one was originally built to install a copy of the first injector—availability depending on reliable injector operation was the issue. Meanwhile it seems to be more adequate to aim for a different injector, favouring longer pulse or even continuous wave operation.

2.3 First linac section, L1

The first section of the main linac consists of four superconducting XFEL accelerator modules operated at 1.3 GHz. Since each module houses eight ≈1 m long superconducting structures, and since the required energy increase is only 470 MeV—the bunch compression scheme asks for ≈600 MeV at BC1—the accelerating gradient in the first linac section is very moderate and well below the XFEL design gradient of 23.6 MV/m. In fact, the failure of a few cavities could easily be compensated. With respect to the RF operation, the first four modules represent a standard XFEL unit, since all four are connected to a single 10 MW multibeam klystron [5]. While the injector klystrons are located outside the accelerator tunnel, the configuration of this first RF power station is identical to all other downstream stations: the modulator is installed outside the tunnel, the pulse transformer and the klystron, with its waveguide distribution, are located below the accelerator modules (see also Fig. 1). Special care is taken to improve the availability of the first linac section. The low-level RF control, installed in shielded compartments next to the klystron, is duplicated, with the possibility of switching between the two systems without tunnel access.

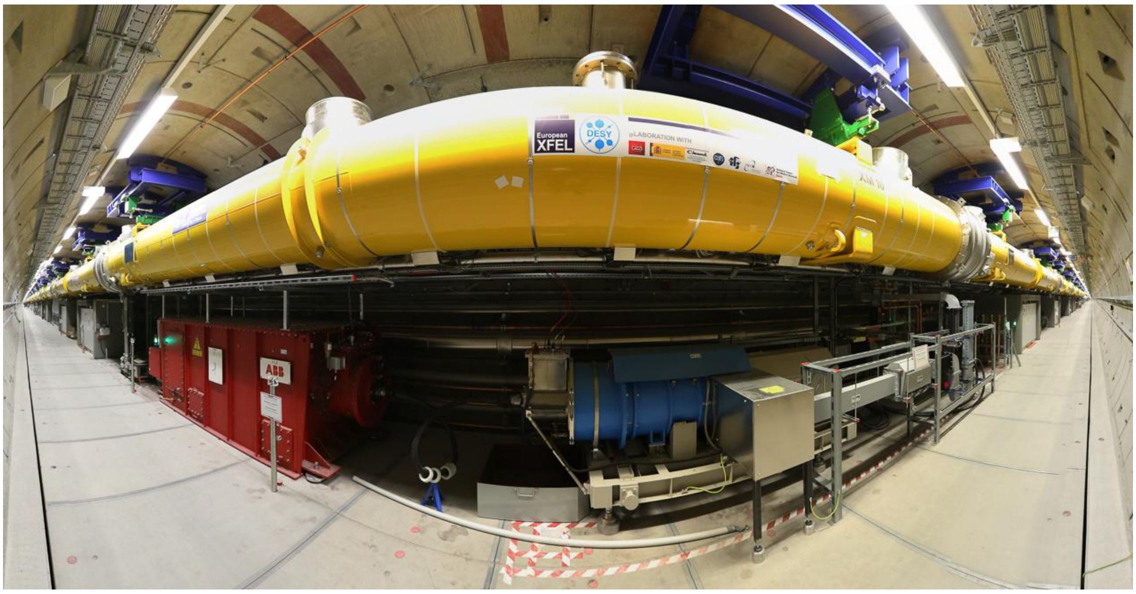


Fig. 1: Wide-angle photograph showing some few metres of the, in total, almost 1 km long superconducting linac of the European XFEL. The yellow accelerator module (length 12.2 m) is suspended from the ceiling. It houses eight superconducting structures. All installed 96 main linac modules are the result of a strong collaborative effort. Subcomponents were contributed by different partners, assembly [2, 3] was done at Saclay, France, and final cold testing [4] was carried out in the accelerator module test facility at DESY, Hamburg.

2.4 Bunch compression in BC1

The next section, starting at ≈100 m deep in the main linac tunnel (called XTL), is the bunch compression chicane, BC1.

The BC section needs four dipole magnets, further focusing elements, and beam diagnostics. Since this warm beamline section is close to the preceding, as well as to the succeeding, cold linac section, particle-free preparation of ultrahigh vacuum systems is essential. Here, the work had already started during the design phase of all respective beamline components. Cleaning methods had to be considered early on, and movable parts are to be avoided wherever possible. In consequence, the chicane vacuum chambers are wide and flat (in the vertical plane); changing the compression factor by shifting the beam to different paths does not involve moving the vacuum chambers mechanically. Here, the European XFEL design differs from normal-conducting linac designs, which are usually less restrictive with respect to particle cleanliness.

2.5 Second linac section, L2

The BC1 compressor is followed by a 12 accelerator module section (called L2). This altogether 150 m long superconducting linac is supposed to increase the electron beam energy to 2.4 GeV. The required average gradient is, at 18.75 MV/m, still moderate. Also here a conservative design gradient was chosen. Conversely, the installation of intentionally high-performance modules—accelerating gradients of around 30 MV/m were achieved in many module tests—can be and in fact has been done, again to increase the availability of a beam with sufficiently high energy, here at bunch compressor BC2. In addition, an energy increase at BC2 during parameter optimization becomes possible. From the point of view of the RF station, L2 consists of three identical RF stations with a pulse transformer and klystron every 50 m. Cryogenic-wise, L2 forms a standard unit. Altogether, 12 modules are connected to one cryogenic string, i.e., one long cryostat without intermediate separation valves. All linac sections have a cryogenic feed- and end-box, both connecting to the cryogenic bunch compressor bypass lines linking the different linac sections.

2.6 Final bunch compression in BC2

Downstream of L2, the last bunch compressor BC2 is installed, which basically repeats the functionality of BC1, here with the goal of producing the final electron bunch length required for lasing. A bunch length of 0.02 mm, corresponding to a 5 kA peak current, with a relative energy spread of 0.3% at 2.4 GeV beam energy will be produced. The section includes a transverse deflecting system as an essential beam diagnostic device. Single bunches are picked and deflected transversely to convert the short bunch length into a corresponding transverse beam size that can then be measured.

2.7 Main linac section, L3

Downstream of BC2, the linac L3 starts with a design length of more than 1 km. The actually installed length, including the cryogenic string connection and end boxes, is 984 m. Taking into account all installed main linac accelerator modules—4 in L1, 12 in L2, and 80 in L3—the achievable electron beam energy is greater than the European XFEL design energy of 17.5 GeV. The exact value will depend on the optimization of the low-level RF (LLRF) control, and here especially on the regulation reserve needed as a function of the electron beam current.

The main linac ends after 96 accelerating modules, corresponding to nine cryogenic strings, or 24 RF stations. The shortening by four accelerating modules was because of beamline vacuum leaks in two modules that could not be repaired in a timely manner. A third module suffers from a small leak in one of the cryogenic process lines. Thus, one RF station, equivalent to four modules, was omitted; this was legitimated by the excellent performance of many accelerator modules. A temporary transport beamline was installed, which is then followed by some further transport and a collimation beamline, protecting the downstream undulator beamlines from beam-halo and mis-steered beams in case of linac problems.

2.8 Beam transport, collimation, and distribution to the different undulators

Downstream of the linac, the electron beamline is also supported from the ceiling, over a length of 600 m. This keeps the tunnel floor free for transport and installation of electronics. Especially at the end of the 5.4 m diameter tunnel, where three beamlines (to SASEs 1 and 3, SASE 2, and into the linac dump) run in parallel, installation and maintenance of the components posed a considerable challenge. During accelerator operation, the electrons are distributed with a fast-rising flat-top strip-line kicker into one of the two electron beamlines. Another kicker system is capable of deflecting single bunches into a dump beamline. This allows for a free choice of the bunch pattern in each beamline, even with the linac operating with constant beam loading.

All undulators and photon beamlines are located in a fan-like tunnel. Figure 2 shows the arrangement of two hard X-ray undulators (SASE 1 und SASE 2), and a soft X-ray undulator (SASE 3), installed downstream of SASE 1. Each undulator provides X-ray photon beams for two different experiments. The time structure of the photon beams reflects the electron bunch pattern in the accelerated bunch trains, affected by the kicker systems.

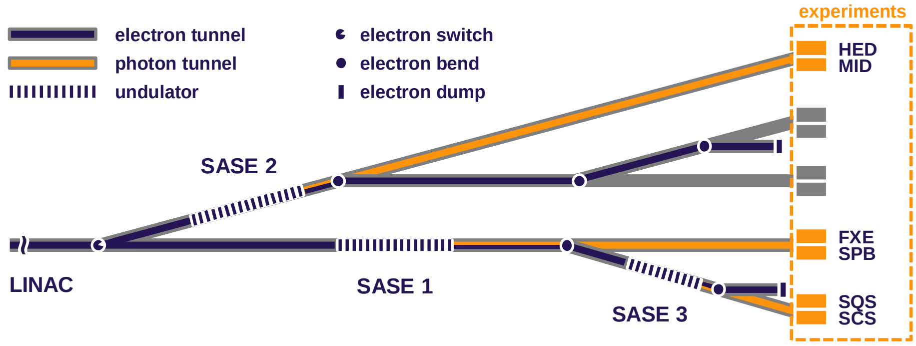


Fig. 2: Arrangement of two hard X-ray undulators (SASE 1 und SASE 2) and a soft X-ray undulator (SASE 3) installed downstream of SASE 1.

The fan-shaped tunnel system houses two electron beam dumps. Here the electrons are stopped after separation from the photon beams. Each dump can handle up to 300 kW beam power. An identical beam dump is located further upstream, at the end of the main linac tunnel (not shown in Fig. 2). Thus, accelerator commissioning and also beam operation is possible while installation or maintenance work in the undulator and photon beam tunnels is ongoing. All five photon beam tunnels end at the experimental hall. During initial operation, two experiments each are set up at three beamlines.

3 Overview of accelerator in-kind contributions

As described, the European XFEL project benefits from in-kind contributions provided by many partners. In the following, an overview is given, which allows understanding of the responsibilities within the project. The description essentially follows the project structure, i.e., contributions to the superconducting linac are listed first, followed by assignments related to the other sections of the accelerator complex. Infrastructure tasks are also described.

3.1 Cold linac contributions

Building the world's largest superconducting linac was only possible in collaboration. Sufficiently developed superconducting RF expertise was required. Major key players already working together in the TESLA linear collider R&D phase joined the European XFEL in an early phase. During the XFEL construction phase, DESY had several roles. The accelerator complex, including the superconducting linac, required coordination. At the same time, large in-kind contributions in the field of superconducting RF technology were made. Work packages contributing to the cold linac are, in all cases, co-led by a DESY expert and a team leader from the respective contributing institute. Integration into the linac installation and infrastructure was another task. The commissioning and operation of the accelerator complex is delegated to DESY.

The accelerator of the European XFEL is assembled from superconducting accelerator modules contributed by DESY (Germany), CEA Saclay, LAL Orsay (France), INFN Milano (Italy), IPJ Swierk, Soltan Institute (Poland), CIEMAT (Spain), and BINP, Russia. The overall design of a standard XFEL module was developed in the frame of TESLA linear collider R&D. Final modifications were made for the required large-scale industrial production. Further details of the contributions to the superconducting accelerator modules can be found in Ref. [3].

3.2 Contributions to the cold linac infrastructure

Operation of the superconducting accelerator modules requires the extensive use of dedicated infrastructure. DESY provided the RF high-power system, which includes klystrons, pulse transformers, connection modules and matching networks, high-voltage pulse modulators, preamplifiers, power supplies, RF interlocks, RF cables, and waveguide systems. During the design and development phase, the 10 MW multibeam klystrons used were developed together with industrial partners. In total, 27 klystrons were finally ordered from two vendors. Pulse transformers were procured as one batch from one company. The modules connecting klystrons and pulse transformers were developed and built in collaboration with BINP Novosibirsk. Each klystron supplies RF power for 32 superconducting structures, i.e., four accelerator modules. The waveguide system used takes care of sophisticated RF power matching [6]. The individual accelerating gradients, determined by module tests, are considered for a special tailoring of the distribution system. To optimize the RF control, both outputs of the multibeam klystron deliver roughly the same power, which is realized by a sorting of the accelerator modules before tunnel installation.

The LLRF system that controls the accelerating RF fields of the superconducting modules is another major DESY contribution. Precision regulation of the RF fields inside the accelerating cavities is essential to provide a highly reproducible and stable electron beam. The RF field regulation is achieved by measuring the stored electromagnetic field inside the cavities. This information is further processed by the feedback controller to modulate the driving RF source. Detection and real-time processing are performed using the most recent field programmable gate array (FPGA) techniques. Performance increase demands a powerful and fast digital system, which was realized with the Micro Telecommunications Computing Architecture (MicroTCA.4). Fast data transfer and processing is achieved by FPGAs within one crate, controlled by a CPU. In addition to the MicroTCA.4 system, the LLRF comprises external supporting modules, also requiring control and monitoring software. During the XFEL construction phase, DESY was operating the Free Electron Laser (FLASH), which is a user facility of the same type as the European XFEL but at a significantly lower maximum electron energy of 1.2 GeV. The LLRF system for FLASH is equal to that of the European XFEL, which allowed for testing, developing, and performance benchmarking in advance of the European XFEL commissioning [7].

BINP Novosibirsk produced and delivered major cryogenic equipment for the linac, such as valve boxes and transfer lines. The cryogenic plant itself was an in-kind contribution of DESY.

3.3 Contributions to the warm linac sections

The largest visible contributions to the warm beamline sections are the >700 beam transport magnets and the 3 km vacuum system in the different sections. While most of the magnets were delivered by the Efremov Institute, St. Petersburg, a smaller fraction were built by BINP Novosibirsk. Many metres of beamline, either simple straight chambers or quite sophisticated flat bunch compressor chambers, were also fabricated by BINP Novosibirsk. DESY made a careful incoming inspection, including particle cleaning when necessary.

State-of-the art electron beam diagnostics are of essential importance for the success of an FEL. Thus, 64 screens and 12 wire scanner stations, 460 beam position monitors of eight different types, 36 toroids, and 6 dark-current monitors are distributed along the accelerator. Longitudinal bunch properties are measured by bunch compression monitors, beam arrival monitors, electro-optical devices, and, most notably, transverse deflecting systems. Production of the sensors and readout electronics is basically finished. Prototypes of all devices have been tested at FLASH. BPM electronics were developed by the Paul-Scherrer-Institut, Villigen, and showed, together with the DESY built pick-ups, performance exceeding the specifications [8, 9].

4 Accelerator status at the start of commissioning

As of autumn 2016, the installation work in the main accelerator tunnel will be finished. All linac sections except for the last cryogenic strings (eight accelerator modules) will be ready for cold commissioning. The complete linac will be cooled to operating temperature. The last cryogenic string requires final actions, such as finishing the waveguide systems, commissioning of the technical interlock system, or, for a few components, even finishing installation of signal cables. The respective work will be done during maintenance access.

4.1 Cold linac status

Installation of, in total, 96 main linac accelerator modules was finished in September 2016. The original plan to get one module per week ready for tunnel installation was basically fulfilled. Modules assembled at CEA Saclay came to DESY and were tested. Test results were used to define the RF power distribution, which was then realized by a proper tailoring of the waveguide system. Sorting of modules helped to find an optimum in the grouping of four modules connected to each multibeam klystron. Finally, some prognosis with respect to the achievable linac energy can be made. Neglecting the working points of the bunch compressors, and looking only at the accelerator modules' usable gradients, as determined during the cold test after arrival at DESY, the sum of all individual accelerator modules' usable gradients is about 22 GeV. Respecting the constraints of the possible RF power distribution leads to a reduction to 21 GeV, corresponding to an average gradient of 27.5 MV/m. The European XFEL linac by far exceeds the design gradient of 23.6 MV/m. Details are given in Ref. [4].

It is expected that during cold commissioning some accelerator cavities or the respective associated systems (RF power coupler, waveguide, LLRF) will show some unforeseen limitations. The European XFEL design included one RF station (i.e., four modules) as spare. Thus, it is correct to state conservatively that the designed 17.5 GeV final energy can be safely reached. The excess in energy will give a higher availability.

The nominal working point of BC2 is 2.4 GeV, while the current highest possible working point is 3.3 GeV, which would bring the final energy to about 19.5 GeV, assuming that all systems are in operation and close to their limit.

Completing the picture of the accelerator module performance, the following can be stated.

- To make 808 superconducting cavities available for 101 accelerator modules, fewer than 1% extras were required. This is based on indispensable quality measures in the full production chain [10].
- Although many accelerator modules needed correction of non-conformities (component or assembly related), discovered either during assembly or even later during test at DESY, in the end, only three modules were not ready for installation in time. Nevertheless, sufficient expertise was required at all partner laboratories.
- Most challenging for the cold linac team was the availability of the RF power couplers. Quality issues, often, but not exclusively, related to the copper plating of stainless steel parts, and the resulting schedule challenges, were faced. The experienced supply chain risk required a large degree of flexibility and willingness to find corrective measures.

4.2 Other sections of the accelerator complex

The installation of all beamline sections from the injector to the end of the main linac tunnel (XTL) will be finished at the time of linac cool-down. Beam transport to the linac commissioning dump after 2.1 km will be possible.

After the linac, almost 3 km of electron beamlines distribute the beam through the SASE undulators to the three different beam dumps. In the northern branch, housing the SASE1 and SASE3 undulators, most of the beamline sections are ready. All undulators are in place. During the last quarter of 2016, the northern branch of tunnels will be completed. The southern branch, housing SASE 2, is scheduled for the first quarter of 2017.

5 Conclusion

The installation of the European XFEL accelerator complex comes to an end. While the linac sections are finished and cool-down and commissioning follows, the remaining beamline sections will be finalized in the next months. First lasing in the SASE 1 undulator is expected for spring 2017, about 6 months after the start of the linac cool-down

Acknowledgements

The European XFEL was built in a huge collaborative effort, accompanied by an immense team spirit of the involved partners. The author would like to thank all colleagues working as work package leaders, supervisors, key experts, field workers, or backstage helpers. Useful information of the project structure and an overview about in-kind contributions are available [11, 12].

References

[1] The European X-Ray Free-Electron laser, Technical design report, DESY 2006-097 (2007), https://doi.org/10.3204/DESY_06-097

[2] S. Berry, Assembly of XFEL cryomodules: lessons and results, Proc. LINAC’16, East Lansing, MI, USA, 2016, paper WE1A02.

[3] H. Weise, How to produce 100 superconducting modules for the European XFEL in collaboration and with industry, Proc. IPAC’14, Dresden, Germany, 2014.

[4] N. Walker, Performance analysis of the European XFEL SRF cavities, from vertical test to operation in modules, Proc. LINAC’16, East Lansing, MI, USA, 2016, paper WE1A04.

[5] V. Vogel *et al.*, Summary of the test and installation of 10 MW MBKs for the XFEL Project, Proc. LINAC’16, East Lansing, MI, USA, 2016, paper TUPLR017.

[6] S. Choroba *et al.*, Series production of the RF power distribution for the European XFEL, Proc. LINAC’16, East Lansing, MI, USA, 2016, paper THPLR067.

[7] C. Schmidt *et al.*, Performance of the microTCA.4 based LLRF system at FLASH, Proc. IPAC’14, Dresden, Germany, 2014.

[8] M. Stadler *et al.*, Beam test results of undulator cavity BPM electronics for the European XFEL, Proc. IPAC’14, Dresden, Germany, 2014.

[9] D.M. Treyer *et al.*, Design, and beam test results of button BPMs for the European XFEL, Proc. IPIC2013, Oxford, UK, 2013.

[10] W. Singer *et al.*, Production of superconducting 1.3-GHz cavities for the European X-ray Free Electron Laser, *Phys. Rev. Accel. Beams* **19** (2016) 092001. https://doi.org/10.1103/PhysRevAccelBeams.19.092001

[11] https://www.xfel.eu/organization/index_eng.html

[12] https://www.xfel.eu/facility/construction/in_kind_contributions/index_eng.html

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Crystallography and Molecular Imaging using X-ray Lasers

T.A. White
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract
A very successful application of X-ray free-electron lasers has been made in structural biology. Acquiring diffraction data using X-ray pulses with durations of a few tens of femtoseconds allows the conventional processes of radiation damage to be sidestepped, breaking limits that previously applied, while at the same time permitting experiments to probe chemical dynamics on short timescales. This contribution gives an introduction to applications of free-electron lasers in biochemistry, including potential future applications to single-molecule diffraction.

Keywords
Crystallography; biochemistry; free-electron laser.

1 The need for knowledge of the structure of biomolecules

The functioning of many biochemical systems relies primarily not on the chemical compositions of the molecules—the relative amounts of different elements making up the participating molecules—but rather on the structures of the molecules. Proteins, for example, are comprised primarily of carbon, hydrogen, nitrogen, and oxygen atoms in the ratios $40:62:10:12$ respectively, with remarkably little variation. Proteins are built from amino acids, which can be consecutively bonded together to form a *polypeptide* chain. A protein consists of one or more polypeptide chains, perhaps in addition to a few extra chemical groups. Stabilized primarily by hydrogen bonds, the polypeptide chains form specific structures consisting of structural motifs, such as helices, sheets, and loops. The specific structure of a protein gives it its particular properties, for example, by creating an open region or 'pocket' where another molecule can fit during catalysis of a bond-breaking reaction. Many proteins can be considered as 'molecular machines'.

Discovering the structure of a protein is one of the first steps towards understanding how it performs its task. This understanding is directly applicable in areas such as medicine, one possible application being to block a binding pocket in a pathogenic protein and therefore hinder its activity. A good example of this is shown in Fig. 1, which illustrates a protein known as HIV-1 protease. At a certain stage in the life cycle of the human immunodeficiency virus, this protein cuts a long polypeptide chain into smaller segments, which have specific activities of their own. The active side of HIV-1 protease is the hole that can be seen in the picture, which is where the polypeptide chain fits while being cut. The action of the protein can be inhibited by blocking the active site with some other molecule. A class of drugs known as protease inhibitors work in exactly this way, by fitting into the hole and binding there more tightly than the polypeptide chain. By acquiring and applying knowledge of the structure of the protease, including seeing how the inhibitor binds into the hole, we may be able to make better protease inhibitors, which bind more strongly and more selectively to the target molecules. Stronger binding, in this case, would mean the drug is more effective, while more selective binding may mean that the drug has fewer side-effects.

2 Diffraction from molecules and crystals

In an ideal experiment, we could determine the structure of a single protein molecule simply by placing it in an X-ray beam and measuring the intensity of the scattered X-rays in different directions with an

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.605

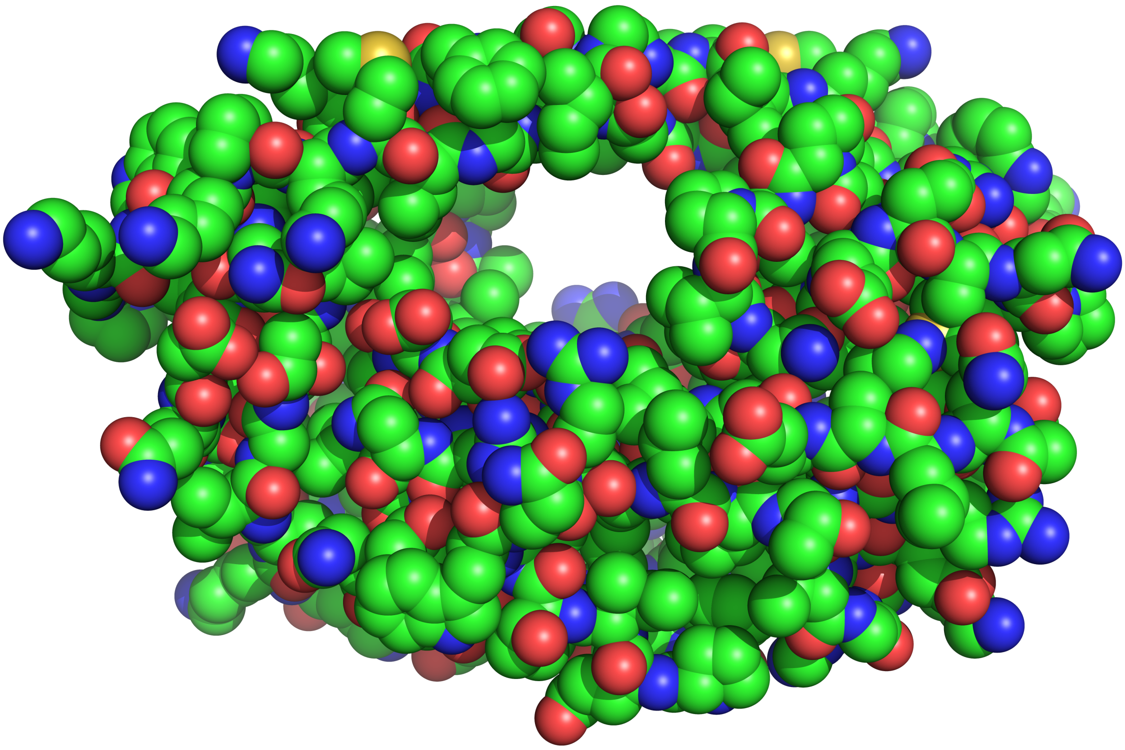


Fig. 1: Structure of HIV-1 protease, showing the hole where a long peptide chain fits while being cut into smaller segments. Atoms are represented by spheres. Carbon atoms are green; nitrogen, blue; oxygen, red; and sulphur, gold. Hydrogen atoms are not shown. Image generated from Protein Data Bank entry 3HVP [1].

area detector. Several features of a practical experiment have prevented us from doing this so far. First, the incident X-ray beam would have to be very intense to produce a measurable diffraction pattern. The required flux is of the order of 10^{13} photons per square micrometre, which corresponds to a radiation dose of the order of 10^{10} Gy [2]. The gray (Gy) is the unit of radiation dose, one gray being defined as one joule of radiation absorbed per kilogram of matter. These figures do not immediately say anything about the timescale over which the flux should be administered—it could be given using a very weak X-ray source over a long period of time—but 10^{10} Gy is a very large X-ray dose for a biological sample, several orders of magnitude greater than it would be able to withstand. The situation appears even more dire when we consider that the term 'measurable diffraction pattern', above, already takes into account that the detector should be able to measure individual scattered X-ray photons.

This problem can be solved by spreading the radiation dose among a large number of protein molecules. This can be done by using a protein crystal instead of a single protein molecule. A crystal consists of many copies of a translationally repeated unit cell, each of which contains one or more copies of the entire protein molecule. The regular arrangement of molecules within a crystal provides a very large increase in the scattered intensity, because the X-rays scattered by each unit cell of the crystal interfere constructively with those scattered by the others. This raises the signal to the point where a diffraction pattern can be measured from a protein crystal using an X-ray source of the type found in many laboratories. The increased signal, however, is compressed into sharp Bragg peaks, compared with the smoothly varying pattern that would be seen from a single molecule (albeit at a very low signal level). Figure 2 shows these two scenarios side by side. Reconstructing the structure in both cases would mean solving the well-known phase problem. The X-ray detector can measure only the intensities of

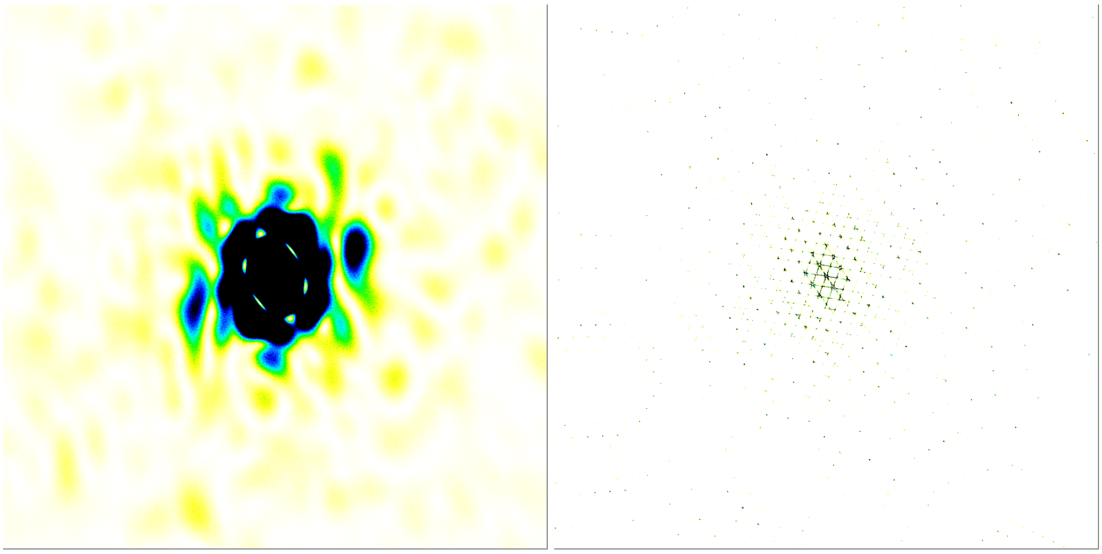


Fig. 2: Comparison of diffraction patterns from a single molecule (left) and a crystal (right). The colour scale, for low to high intensity, goes from white through yellow, green, cyan, blue, and, finally, black.

the diffraction signal, when in reality it is a complex-valued function and therefore has phase values as well. No atomic-resolution lens exists for X-rays, so we cannot build an imaging system to turn the amplitude and phase information directly into an image. Instead, we must reconstruct the phase information computationally.

It might appear that the crystal diffraction pattern, although it has the big advantage of being easily measurable, contains less information, and this is indeed the case. The crystal diffraction pattern in fact contains exactly half of the information that would theoretically be needed to reconstruct the structure [3]. The single-molecule pattern contains sufficient information to reconstruct the structure using a constraint satisfaction procedure [4], at least for the ideal experiment without background scattering and detector artefacts.

For the crystal case, a variety of structure solution techniques have been developed for making up the information shortfall. The most widely used of these is known as *molecular replacement*, and essentially involves comparing the measured intensities with those that would be produced by a molecule that we hypothesize is similar in structure to the protein under investigation. Other techniques exploit the variation of scattering power of heavy atoms with X-ray wavelength (the so-called anomalous diffraction techniques), or the changes to the intensities that arise when heavy atoms are embedded into the structure (the isomorphous replacement method). Notice that all of these techniques involve making further measurements or introducing external information. Direct methods, which are 'pure' solutions to the phase problem without doing either of these things, also exist. However, direct methods are usually not applicable to diffraction data from protein crystals because they require information to higher resolution than is usually permitted by their degree of order [5].

Is there any hope of measuring the diffraction from a single molecule? This would be possible if we could somehow suspend the usual rules of radiation damage and deliver an extremely large X-ray dose without the molecule being damaged. It turns out that this can be achieved by delivering the X-rays in a pulse with a very short duration indeed, and this is, of course, exactly the type of pulse provided by an X-ray free-electron laser. Before the first X-ray free-electron laser had been built, it was theorized, based on computer simulations, that radiation damage could be 'sidestepped' by delivering the entire X-ray dose in a single pulse with a duration of a few femtoseconds [6]. The molecule would be completely destroyed shortly afterwards, the electrons having been stripped from the atoms by the intense electric field, but the destruction would happen on a timescale longer than the pulse. Since the diffraction pattern is recorded during the pulse itself and not the time afterwards, the destruction of the molecule should not significantly affect the diffraction signal.

A few years after it was first proposed, this ‘diffraction before destruction’ principle was demonstrated in an experiment at FLASH, a soft X-ray free-electron laser facility in Hamburg, Germany [7]. Using a single radiation pulse, a diffraction pattern was recorded from a two-dimensional pattern etched into a silicon nitride membrane. The pattern was destroyed in the process, as could be seen from a pattern recorded using a subsequent pulse on the same sample.

3 Crystallographic data collection and processing

Interpretable diffraction signals from single protein molecules have not been achieved to date. In the meantime, the ability of X-ray free-electron laser pulses to sidestep the usual radiation damage processes has been put to good use on crystalline samples. In biomolecular crystallography, radiation damage is still a serious problem, although the dose is spread among a large number of molecules. Conventional crystallography, as has been practised for several decades, is based on the rotation method, where a series of diffraction patterns is recorded while the crystal is continuously rotated. This produces a three-dimensional dataset. However, it requires the crystal to be exposed to the X-ray beam for a relatively long period of time. The maximum tolerable dose must be apportioned over the entire rotation series.

Using X-ray pulses from a free-electron laser sidesteps the radiation damage limit, but unfortunately means that only one diffraction pattern can be recorded from a single crystal. Many different views of the crystal, in different orientations, are needed to reconstruct the three-dimensional structure. If a very large crystal is available, subsequent frames can be recorded from a different position on it, away from the region affected by damage from the first shot. Applications of this method have been described [8]. However, if this is not possible then many crystals will be required to form a complete dataset. The situation where only one diffraction ‘snapshot’ is recorded from each crystal has been dubbed ‘serial crystallography’. There are many ways to achieve this, but one of the most popular is to inject a ‘jet’ of crystal-laden liquid into the path of the X-ray beam [9]. Another popular method is to spread the crystals over a solid support, and then to raster the X-ray beam across it (although in practice the sample is moved, not the X-ray beam). When serial crystallography is performed using femtosecond X-ray pulses, it is known as *serial femtosecond crystallography*.

Data processing in serial crystallography is broadly similar to data processing for conventional rotation data. It is an active field of research in its own right, and has been described extensively in the literature [10]. The process is briefly described here. First, the Bragg peaks are found in each detector frame, and a decision made about whether the frame actually contains a usable diffraction pattern or not. Depending on the sample delivery method and the density of well-ordered crystals, typically only a small fraction of detector frames actually contain usable patterns. Once the ‘hits’ have been identified, the locations of the Bragg peaks are used to determine the orientation of the crystal, and hence to calculate the locations in the image where Bragg peaks should appear. This way, measurements can be made of all the peaks, even if they were not all found by the peak search, and even if some of them are very weak. Of course, the information that a particular Bragg peak is very weak, or even completely absent, is just as important as if it were very strong. Once the intensities have been measured from all the patterns, the measurements are combined, and the merged intensities used with a variety of algorithms to solve the structure. For serial crystallography, several data processing packages are now available, the most popular of which is CrystFEL [11, 12]. The subsequent steps—to solve the phase problem, determine the structure, and refine the structural model—are essentially the same for the merged data from free-electron laser experiments as in conventional rotation crystallography.

4 Human membrane proteins

All organisms must keep some volumes separated from others. For example, a cell contains different compartments, which allow many specialized processes to occur separately, and these compartments may contain very different chemical environments, such as varying levels of acidity. A completely sealed cell compartment would not be of much use, and some means of transmitting signals or controlling the movement of substances in and out of compartments is required. These movements and transmissions are controlled by proteins embedded in the membranes enclosing the compartments. Since so many biological processes involve this type of protein, they are the target of most pharmaceutical substances and consequently the proteins of which we would most like to know the structures. Unfortunately, extracting such proteins from the membrane is difficult, because often pressure from the membrane helps to maintain their structure and they become unstable once extracted. The proteins in this category, known as integral membrane proteins, are therefore some of the most difficult ones to study crystallographically.

This category of protein has proven to be a success story for serial femtosecond crystallography, with several examples now published. Crystals can be grown inside a *lipidic cubic phase*, in which the proteins remain embedded in a membrane. Rather than subsequently extracting the crystals from the lipidic cubic phase matrix, the whole thing can be injected into the X-ray beam using an injection device designed for viscous fluids [13], which greatly simplifies sample handling. In addition, using a viscous medium has the great advantage that the sample can flow very slowly, meaning that more of it is probed by the X-ray beam rather than flowing past the interaction point between X-ray pulses and being wasted.

Proteins that have been studied include the serotonin receptor $5\text{-}HT_{2B}$ [14], δ-opioid receptor [15], AT_1 receptor [16], A_{2A} receptor [17], and *smoothened* receptor [13]. All of these are human proteins, involved with the regulation of such things as pain and blood pressure. In all these cases, the protein was crystallized while containing a drug molecule or analogue of one. From a biochemical point of view, this is useful for precisely the reasons outlined in Section 1. From the point of view of technique development, it is also useful because it provides a test of quality of the data. If the structure is solved correctly and the intensity measurements are sufficiently accurate and precise, the drug molecule should be visible in the resulting electron density map. Indeed, this was the case for all these examples.

A recent success for serial femtosecond crystallography was the structure of a light-sensitive protein found in the retina, rhodopsin, bound to another protein called arrestin. The action of arrestin reverses the structural changes that take place in rhodopsin when light interacts with it, preparing it for a new cycle of light detection. The combination of both molecules is very delicate, making it very challenging to grow large, well-ordered crystals, and data could be acquired only to a resolution of 7.7 Å using conventional rotation crystallography at a synchrotron. In contrast, the structure was solved to a resolution of about 3.5 Å using serial femtosecond crystallography [18].

5 Time-resolved crystallography

Although the use of free-electron lasers for determining static structures has been the target of interest in structural biology, their potential for determining dynamic structures, or so-called ‘molecular movies’, is much greater. The short duration of the X-ray pulses means that the time resolution can be very high. Time-resolved serial femtosecond crystallography has now been applied to several systems where a protein responds to illumination in the visible region of the electromagnetic spectrum. In these cases, the reaction was triggered using a short pulse of light from a laser, a small fraction of a second before the arrival of the X-ray free-electron laser pulse at the sample. This is the so-called ‘pump-probe’ scheme, and is used in conjunction with many techniques other than X-ray crystallography.

The process of photosynthesis is a very important and interesting light-activated chemical reaction. It has therefore been (and continues to be) a high-profile target for time-resolved femtosecond crystallography [19, 20], even more so because of the large sizes of the protein complexes involved and the consequent difficulties with making large crystals.

Recently, the high time resolution achievable with an X-ray free-electron laser was demonstrated for two proteins, myoglobin [21] and photoactive yellow protein [22], in both cases achieving a time resolution close to 100 fs. Many other ways to start a chemical reaction have been proposed, such as to mix a protein with a substance it can act on [23], and acquire diffraction snapshots after a controlled amount of time for mixing and diffusion to occur.

6 Future outlook

It would be the dream of many structural biologists to determine the structure of biomolecules, at near atomic resolution, without crystallization, and progress is being made in that direction. Even though the problem of radiation damage is avoided, there are several other difficulties to be overcome. Since the signal level is still very low, only a few photons at high resolution, background scattering must be reduced to almost nothing, in turn, affecting how the protein molecules are delivered to the X-ray beam. The X-ray beam must be carefully characterized, and the detector must be well-calibrated and produce low noise. Alongside this, an interesting discovery was made recently; that certain types of disorder in a protein crystal can reveal the molecular scattering [24]. In this case, the signal is superimposed on the Bragg scattering of the crystal and extends to a higher resolution than the Bragg peaks.

It is clear that structural biology will continue to be a very important and productive application for current and future X-ray free-electron laser facilities.

Acknowledgements

I thank Henry Chapman, Dominik Oberthür, and Carolin Seuring for feedback on the manuscript, as well as funding from the Helmholtz Association through project oriented funds.

References

[1] A. Wlodawer *et al.*, *Science* **245** (1989) 616. https://doi.org/10.1126/science.2548279
[2] Q. Shen *et al.*, *J. Synchrotron Radiat.* **11** (2004) 432. https://doi.org/10.1107/S0909049504016772
[3] D. Sayre, *Acta Crystallog.* **5** (1952) 843. https://doi.org/10.1107/S0365110X52002276
[4] S. Marchesini, *Rev. Sci. Instrum.* **78** (2007) 011301. https://doi.org/10.1063/1.2403783
[5] C.J. Gilmore, *Acta Crystallogr.* **D56** (2000) 1205. https://doi.org/10.1107/S0907444900009355
[6] R. Neutze *et al.*, *Nature* **406** (2000) 752. https://doi.org/10.1038/35021099
[7] H.N. Chapman *et al.*, *Nat. Phys.* **2** (2006) 839. https://doi.org/10.1038/nphys461
[8] K. Hirata *et al.*, *Nat. Methods* **11** (2014) 734. https://doi.org/10.1038/nmeth.2962
[9] H.N. Chapman *et al.*, *Nature* **470** (2011) 73. https://doi.org/10.1038/nature09750
[10] T.A. White *et al.*, *Acta Crystallogr.* **D69** (2013) 1231. https://doi.org/10.1107/S0907444913013620
[11] T.A. White *et al.*, *J. Appl. Crystallogr.* **45** (2012) 335. https://doi.org/10.1107/S0021889812002312
[12] T.A. White *et al.*, *J. Appl. Crystallogr.* **49** (2016) 680. https://doi.org/10.1107/S1600576716004751
[13] U. Weierstall *et al.*, *Nat. Commun.* **5** (2014) 3309. https://doi.org/10.1038/ncomms4309
[14] W. Liu *et al.*, *Science* **342** (2013) 1521. https://doi.org/10.1126/science.1244142
[15] G. Fenalti *et al.*, *Nat. Struct. Mol. Biol.* **22** (2015) 265. https://doi.org/10.1038/nsmb.2965
[16] H. Zhang *et al.*, *Cell* **161** (2015) 833. https://doi.org/10.1016/j.cell.2015.04.011
[17] A. Batyuk *et al.*, *Sci. Adv.* **2** (2016) e1600292. https://doi.org/10.1126/sciadv.1600292
[18] Y. Kang *et al.*, *Nature* **523** (2015) 561. https://doi.org/10.1038/nature14656
[19] J. Kern *et al.*, *Science* **340** (2013) 491. https://doi.org/10.1126/science.1234273
[20] C. Kupitz *et al.*, *Nature* **513** (2014) 261. https://doi.org/10.1038/nature13453
[21] T.R.M. Barends *et al.*, *Science* **350** (2015) 445. http://dx.doi.org/10.1126/science.aac5492

[22] K. Pande *et al.*, *Science* **352** (2016) 725. https://doi.org/10.1126/science.aad5081

[23] M. Schmidt, *Adv. Condens. Matter Phys.* **2013** (2013) 167276. https://doi.org/10.1155/2013/167276

[24] K. Ayyer *et al.*, *Nature* **530** (2016) 202. https://doi.org/10.1038/nature16949

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Machine Protection

L. Fröhlich
DESY, Hamburg, Germany

Abstract
Conventional linacs used for modern free-electron lasers carry electron beams of unprecedented brightness with average powers ranging from a few watts to hundreds of kilowatts. Energy recovery linacs are already operated as radiation sources with nominal electron beam powers beyond 1 MW, and this figure can only be expected to increase in the future. This lecture discusses the scope of machine protection for these accelerators, reviews the parameters of existing and planned facilities, and gives an overview of typical hazards and damage scenarios. A brief introduction to the interaction of electron beams with matter is given, including a simple model for estimating some properties of electromagnetic cascades. A special problem common to most light sources—the field loss of permanent magnet undulators and its consequences for the emission of radiation—is discussed in the final section.

Keywords
Machine protection; linear accelerators; free-electron lasers; energy recovery linacs; radiation-induced demagnetization.

1 Introduction

Machine protection aspects have influenced the design and operation of particle accelerators for many decades. The storage ring community has recently seen a wave of activity in this field, owing to the unprecedented amount of energy stored in the beams and in the magnets of the Large Hadron Collider. In a similar fashion, the advent of high-gain free-electron lasers (FELs) and energy recovery linacs has led to a renewed interest in high-power electron linacs as drivers of radiation sources and in the special machine protection needs of these facilities.

All of this activity has produced countless reports, conference papers, and articles concerned with specific implementations and technical details. Alas, only a few publications attempt to provide a broader view of the machine protection field; interested readers will find a selection in Refs. [1–8]. The CERN Accelerator School undoubtedly deserves credit for inspiring some of the more extensive works.

This paper does not strive for generality. Prepared for the school on FELs and energy recovery linacs, it excludes any discussion of the specific issues of hadron machines, and is instead biased towards machine protection issues for linear accelerators in light sources. The initial section tries to establish the scope of what 'machine protection' means for these accelerators. Afterwards, an overview of typical hazards and damage scenarios is given and the parameters of existing and planned facilities are reviewed in light of their damage potential. The central part of the paper provides a brief introduction to the interaction of electron beams with matter, which is fundamental to an understanding of many problems in the field. A discussion of the field loss of permanent magnet undulators and its consequences for the emission of radiation sheds some light on a problem specific to light sources at the end of the paper.

2 The scope of machine protection

The term *machine protection* is often understood as a mere synonym for a system of protective interlocks and beam loss diagnostics. While such active systems play an important role, effective protection from damage involves many fields of accelerator engineering and physics. If we attempt to define the term

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-001.613

in a single sentence, we might say that *machine protection is the sum of all measures that protect an accelerator and its infrastructure from the beam.* Traditionally, the focus is on the charged particle beam, but the generated photons need to be considered as well, especially in light of X-ray FELs and energy recovery linacs with unprecedented peak and average power output. If we take this definition seriously, a number of fields must be regarded as integral parts of machine protection work, or at least as closely related.

Machine protection systems: a machine protection system implements interlocks on components that may interfere with the safe transport of the beam (e.g., magnets, screens). It monitors the beam with instrumentation that may be generic (beam position monitors, current monitors) or specifically designed for protection purposes (beam loss monitors, dosimetry systems). When excessive beam losses or other problems are detected, the machine protection system intervenes according to a mitigation strategy—it might simply inform the operator, reduce the repetition rate, or stop the beam production.

Collimators: collimators and scrapers are used to limit the extent of the electron bunch (and of possible dark currents) in phase space. If there are trajectory or focusing problems, they should intercept the electron beam before it reaches sensitive components. The electromagnetic cascades originating from the interaction of high-energy electron beams with matter are not easy to contain, so care must be taken to place suitable absorbers.

Shielding: the loss of a small fraction of an electron beam at the gigaelectronvolt level releases a dangerous amount of spontaneous radiation. Even if the average power of the beam is as low as a few watts, the radiation can quickly cause temporary or permanent damage to electronics in the vicinity of the beamline. Sustained exposure causes various types of radiation damage—cable insulation becomes brittle, optical components darken. Beam loss can also release sizable quantities of neutrons and activate materials in the process. Depending on the beam power, accelerator components may therefore require shielding against both electromagnetic dose and neutrons.

Beam physics: a loss-free transport of charge from the injector to the dump requires a good understanding of the optics and of the entire acceleration process. The higher the beam power, the more important it is to have good control over the optics matching and over collective effects that create emittance blow-ups, tails, or halos.

Robust systems: every system or software package that has a direct or indirect influence on the beam contributes to the protection of the machine by providing a certain level of robustness. Cardinal examples are beam-based feedback systems, low-level radio-frequency systems, or even high-level physics tools for optimization of the radiation output.

Procedures: well-defined procedures for typical linac operations, such as switch-on, change of energy, or ramp to full power, contribute to safety and make the machine state more reproducible. Automatization of these procedures can further help to avoid errors.

3 Beam power of existing and future facilities

When we examine the machine protection needs of an electron linac, the most important characteristic to consider is its average beam power P:

$$\begin{aligned}\text{average beam power} &= \frac{\text{energy}}{\text{charge}} \cdot \frac{\text{charge}}{\text{time}} \\ &= \frac{\text{`beam energy'}}{e} \cdot \text{average current}\,.\end{aligned}$$

Table 1: Maximum energy, bunch frequency, and average beam power of selected existing and planned FELs. The calculation of the beam power assumes typical parameters for minimum and maximum power operation for each facility.

	E (GeV)	ν (Hz)	P (W)
FERMI	1.4	10	14
SACLA	7	10–60	8–140
LCLS	15	120	36–360
FLASH	1.3	1M–3M pulsed	10–22k
European XFEL	17.5	4.5M pulsed	600k
LCLS-II	4	100k–1M cw	120k
NovoFEL	0.012	5.6M–22M cw	15k–60k
JLab FELs	0.2	75M cw	>1M
Future energy recovery linacs?	5	1.3G cw	500M

When the machine accelerates single bunches of charge Q at a fixed repetition rate ν, this becomes

$$P = \frac{\nu Q E}{e} ,$$

where E denotes the energy per electron and e the elementary charge.

Most existing and proposed single-pass FELs are based on normal conducting linacs using S- and C-band accelerating structures. The normal conducting technology permits only a short RF pulse so that, usually, only a single bunch is accelerated per pulse. The beam power is therefore limited by the repetition rate of the RF systems, of 5–120 Hz, and by the maximum usable bunch charge, which may vary between tens of picocoulombs and a few nanocoulombs. Depending on their individual parameters, normal conducting machines transport beams from a few watts to about 400 W (the first three entries in Table 1).

Superconducting linacs can sustain the RF pulse for a considerably longer time span. This makes it possible to accelerate long bunch trains with bunch frequencies in the megahertz range, which raises the average beam power considerably. FLASH, which is still the only working single-pass FEL based on a superconducting linac, has demonstrated the transport of 1800 bunches per pulse at a bunch charge of 3 nC with a repetition rate of 5 Hz, carrying an average power of 22 kW [9]. Facilities that are already under construction have design powers in excess of 100 kW (LCLS-II, with continuous-wave RF systems) or even above 0.5 MW (European XFEL, with pulsed RF systems). It is obvious that superconducting linacs, when operated at these power levels, have a serious damage potential.

Table 1 also lists the parameters for selected energy recovery linacs—although these are oscillators instead of single-pass FELs, they are an instructive point of reference for the typical problems associated with high beam powers. The Jefferson Lab FELs, when operated with a bunch frequency of 75 MHz (continuous wave), can carry a nominal electron beam power of more than 1 MW. This means that even the loss of a tiny fraction of the electron beam can cause serious problems including mechanical damage, and, consequently, machine protection aspects are a fundamental part of the operation of the accelerator. It is a safe assumption that future superconducting single-pass FELs operating in a similar power range will share many of the problems encountered in today's energy recovery linacs, while adding some of their own.

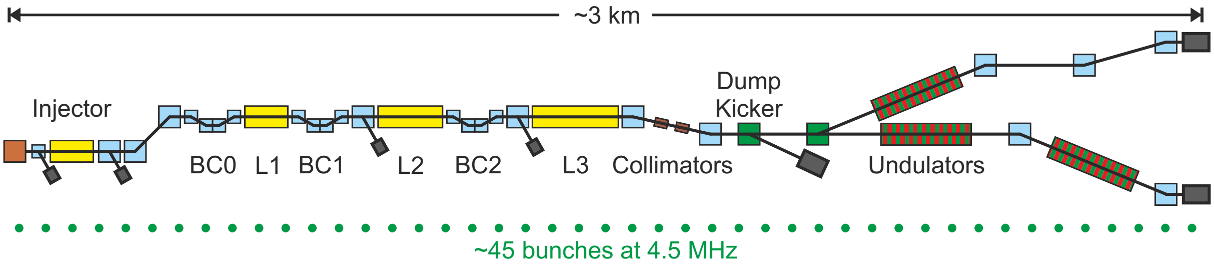


Fig. 1: Schematic of the European XFEL (not to scale). Main dipole magnets are shown as blue squares; accelerating sections as yellow rectangles.

4 Emergency reaction times: a case study

The high repetition rates of superconducting machines bring with them some complications for the design of machine protection systems. The European XFEL (Fig. 1) is an instructive example. In the machine, the distance from the injector laser to the last undulator is approximately 3 km. Hence, a signal needs about 10 µs to travel from one end of the accelerator to the other at the speed of light in vacuum (c). At the maximum bunch frequency of 4.5 MHz, up to

$$3 \text{ km} \cdot 4.5 \text{ MHz}/c \approx 45$$

bunches are simultaneously travelling through the beamline. Assuming that a beam loss occurs at the farthest position from the injector and is detected immediately, the signal still needs considerable time to reach the injector laser in order to switch it off; this time would be on the order of 15 µs for a fibre-optic transmission line with a signal propagation speed of $\frac{2}{3}\,c$. As a consequence, at least $45 + \frac{3}{2} \cdot 45 \approx 113$ bunches would be (partially) lost before a machine protection system could take any countermeasures. At a bunch charge of 1 nC and a final energy of 17.5 GeV, these bunches would carry a total energy of

$$113 \cdot 17.5 \text{ GeV} \cdot \frac{1 \text{ nC}}{e} \approx 2 \text{ kJ}\,,$$

enough to melt about 5 g of copper from room temperature.

In a nutshell, the efficiency of active systems for the protection of a high-power accelerator can already be limited by unavoidable signal propagation times. In such cases, greater emphasis needs to be put on passive protection measures, such as resilient and effective collimators. Sometimes it is also possible to make use of additional beam abort points—for the European XFEL, the fast dump kicker magnet can be fired when beam losses occur in the undulator sections.

5 Hazards

The complete or partial loss of the electron beam in a vacuum chamber can cause a number of detrimental effects. If we try to order these effects roughly by the local power deposition needed to cause them, we obtain a list like the one in Table 2. Any such overview can only be understood as an approximate indication of the orders of magnitude; obviously, each damage scenario needs to be assessed individually and, for special cases, very different numbers may be found.

Direct mechanical damage through melting or sublimation depends on power density rather than power; for typical scenarios, however, a substantial power deposition of hundreds of watts or kilowatts is necessary—hence, direct damage is of little concern for normal conducting machines, but needs to be protected against for superconducting ones. Single-bunch damage is not to be expected for the parameters of typical FELs or energy recovery linacs because of too low charge densities; however, for the design parameters of the International Linear Collider it is a clear possibility [10].

Table 2: Effects of beam loss. The table roughly relates the onset of various damaging effects to the local power deposition caused by a beam loss.

$P_{\min}$ (W)	Effects
100–1000	Thermal or mechanical damage
10–100	Mechanical failure of flange connections
1–100	Activation of components
1–100	Radiation damage to electronics, optical components, etc.
1–10	Excessive cryogenic load, quenches
0.01–0.1	Demagnetization of permanent magnets

The deposition of heat can also have indirect consequences—such as impairing the tightness of a flange connection once the metal starts to cool down after thermal expansion. This, again, is an unlikely scenario for the typical beam powers of normal conducting machines, but is a real danger once the beam power reaches the multikilowatt level.

The spontaneous radiation released by beam losses can lead to malfunctions in electronics or to various types of radiation damage. In fact, the radiation released by a single watt of electron beam dumped on a beam pipe is quite destructive to many types of electronics in the vicinity if no proper shielding is in place. Such a loss is, of course, easily diagnosed in a linac operating at low current, but it only corresponds to a fraction of 10^{-5} of a 100 kW beam. Similar considerations apply to the activation of components; generally, induced radioactivity at electron accelerators is relatively short-lived and substantially less than at hadron machines, but it can impair the maintainability of components and the accessibility of the beamline.

Superconducting accelerators have a special vulnerability to beam losses because any deposition of heat in the cold mass must be compensated for through the cryogenic system with a disproportionate amount of power. Beam losses can also cause superconducting cavities or magnets to quench (i.e., to become normal conducting), which in turn creates an immediate instability in the downstream beam transport. For cavities, a reduction of RF power is usually sufficient to stop a quench, whereas superconducting magnets need more intricate quench protection systems to protect them from damage.

Finally, light sources usually depend on undulators made of permanent magnets. These magnets are installed in the immediate vicinity of the beam axis and are susceptible to field loss under irradiation. This makes beam losses in insertion device sections a particular concern for machine protection. We will therefore revisit the topic in greater detail later on.

6 Interaction of electron beams with matter

For almost all studies related to machine protection, a good understanding of the interaction of the beam with matter is fundamental. In other words, what happens when the electrons or photons of our light source hit an obstacle?

6.1 How electrons lose energy

Electrons passing through matter lose kinetic energy and are deflected from their original direction. Several processes contribute to both effects, most importantly:

– elastic scattering with nuclei;
– inelastic scattering with atomic electrons;
– bremsstrahlung.

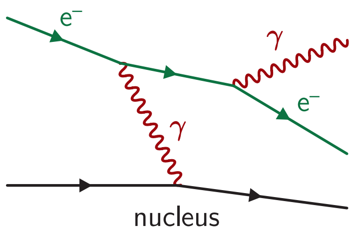


Fig. 2: Feynman diagram for the emission of a bremsstrahlung photon by an electron scattered at an atomic nucleus

All of these phenomena are caused by Coulomb interaction of the projectile with the atoms of the target material, but only the latter two contribute substantially to the energy loss of the electrons. In *elastic scattering* with a nucleus, the mass difference between both collision partners is so large that the electron loses only a tiny fraction of its kinetic energy. Multiple Coulomb scattering in the lattice of the target material can, however, deflect the electrons significantly from their incident direction and cause a broadening of the beam. A discussion of the angular distribution caused by multiple Coulomb scattering is found, e.g., in Ref. [11].

Inelastic scattering mainly takes place between the projectiles and the bound electrons of the target material. Some of the kinetic energy of the moving charge is transferred to the target atom in the form of electronic excitation or ionization. This is the only effect of any importance by which electrons can transfer energy *directly* to matter. The ESTAR online database [12] is an excellent resource for quantitative calculations and contains stopping power and range data for many materials. Readers interested in the original quantum mechanical treatment of inelastic scattering by Bethe and Bloch from the 1930s and in later corrections to the theory should consult Refs. [13–15].

Bremsstrahlung is the radiation emitted by fast electrons due to the interaction with the electric field of the positively charged nuclei of the target material (Fig. 2). By bremsstrahlung, the electrons lose energy without depositing it directly in matter—instead, the energy is carried away by photons, which may or may not interact with the material themselves. In the high-energy limit, these radiative losses scale almost linearly with the energy of the projectile E as

$$\frac{\mathrm{d}E}{\mathrm{d}x} \approx \mathrm{const.} \cdot E \cdot \frac{Z^2}{m^2}\,, \tag{1}$$

where $\mathrm{d}E/\mathrm{d}x$ is the energy loss per distance travelled inside the material, Z is the atomic number, and m is the mass of the projectile. The occurrence of m^2 in the denominator also indicates why bremsstrahlung is so much more important for electrons than for any other charged particles—they are light.

Figure 3 shows the contributions to the energy loss of an electron travelling through aluminium, copper, and lead. Inelastic scattering is most important at low energies, while emission of bremsstrahlung dominates the high-energy region.

For practical purposes, it is useful to know the particle energy at which the loss by inelastic scattering is equal to the radiative loss. To a good approximation, this quantity is a material constant called the *critical energy*. It can be estimated fairly well by the simple formula:

$$E_{\text{crit}} \approx \frac{800\ \text{MeV}}{Z + 1.2}\,.$$

Typical values are 51 MeV for aluminium, 25 MeV for copper, and 9.5 MeV for lead; more materials are listed in Ref. [16].

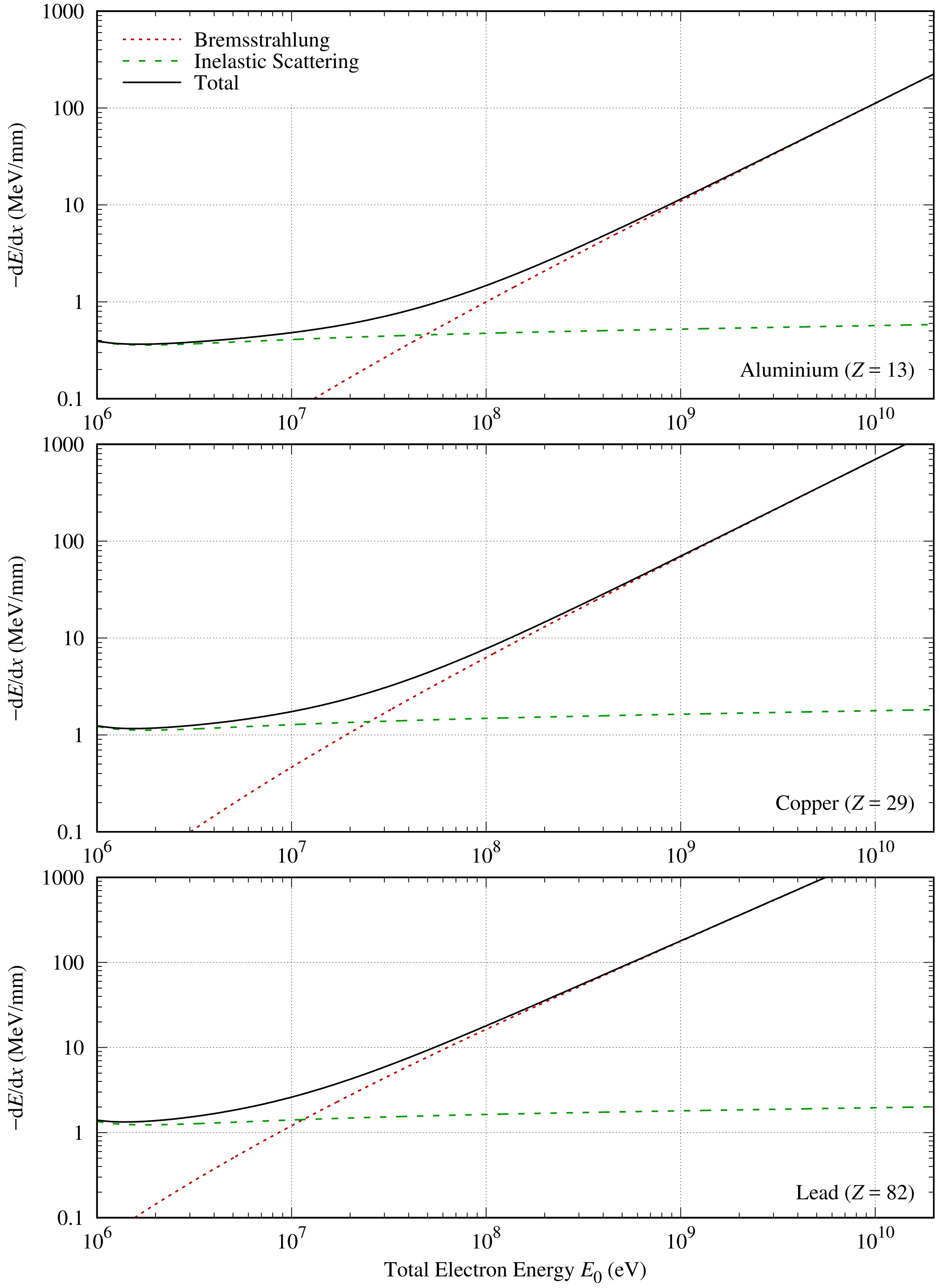


Fig. 3: Energy loss by electrons in aluminium, copper, and lead as a function of total electron energy

Table 3: Radiation length of selected materials [18]

Material	L_{rad} (cm)	X_0 (g/cm^2)
Aluminium	8.90	24.01
Titanium	3.56	16.16
Iron	1.76	13.84
Copper	1.43	12.86
Tungsten	0.35	6.76
Lead	0.56	6.37

6.2 Radiation length

As long as we are in the bremsstrahlung-dominated regime (well above the critical energy), we find that radiative losses are approximately proportional to the total energy of the electrons (Eq. (1)). Of course, this means that the energy decays exponentially with the distance x travelled in matter:

$$E(x) \approx E_0 \exp\left(-\frac{x}{L_{\text{rad}}}\right) .$$

The quantity L_{rad} is a material constant called the *radiation length*. It specifies the distance after which the energy of an ultrarelativistic electron has decreased to $1/e$ of its initial value. Some authors prefer to normalize the constant to the density ρ of the material ($X_0 = L_{\text{rad}} \cdot \rho$), although the resulting quantity is no longer a *length* in the literal sense. Table 3 shows values of L_{rad} and X_0 for some materials. A convenient method of calculating X_0 to a precision of a few per cent is given in Ref. [17]. Using the atomic number Z and the mass number A,

$$X_0 \approx \frac{A}{Z(Z+1)\ln(287 Z^{-0.5})} \cdot 716.4 \frac{\text{g}}{\text{cm}^2} .$$

6.3 Interaction of photons with matter

So far, we have treated only the direct interaction of electrons with the target material. Of course, the emitted bremsstrahlung photons can interact with the material, too. The most important processes for this are:

– the photoelectric effect;
– Compton scattering;
– pair production;
– photonuclear reactions.

The first two effects lead to ionization of the material—in Compton scattering, the incident photon transfers part of its energy to an atomic electron; in the photoelectric effect it is absorbed completely—while pair production creates a positron and an electron from a photon of sufficient energy (Fig. 4). Compared with these three effects, photonuclear reactions are extremely rare. Their main importance lies in the creation of free neutrons in the giant dipole resonance, which is the main source of beam-induced activation at electron accelerators.

The interaction cross-sections of photons in aluminium, copper, and lead are shown in Fig. 5. The photoelectric effect and Compton scattering are more important at lower energies, whereas pair production is clearly the dominant process above a few tens of megaelectronvolts. As a rule of thumb, the interaction cross-section for pair production scales with the square of the atomic number, $\sigma_{\text{pair}} \propto Z^2$. This is essentially the same proportionality as for the energy loss of electrons due to bremsstrahlung

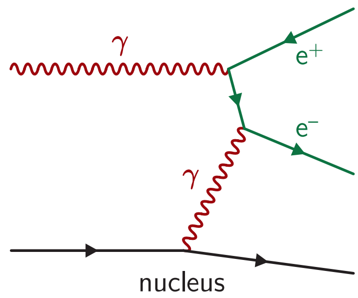


Fig. 4: Feynman diagram for electron–positron pair production

(Eq. (1)): unsurprisingly, heavier elements tend to provide better shielding against high-energy electron and photon beams.

In a pair production event, almost all of the photon energy is converted into the rest mass and kinetic energy of the electron–positron pair; the momentum transfer to the nucleus also participating in the interaction is negligible. Hence, the overall picture is the same as for electrons: high-energy particles do not transfer energy directly to matter; such energy absorption mainly takes place at the lower end of the energy spectrum.

A useful rule of thumb can be used to calculate the *mean free path length* L_{pair} of photons at high energies. It can be readily expressed in terms of the radiation length of the material as

$$L_{\text{pair}} \approx \frac{9}{7} L_{\text{rad}} \,.$$

Ignoring the difference of $\sim$30 %, this translates into the following, remarkably simple result:

> *The typical path length a photon can travel in matter until it is consumed in a pair production event is roughly the same as the radiation length of the material.*

On a final note, electrons and positrons are, of course, not the only particles produced in pair production events. For example, a channel for pair production of the next heavier particle, the muon, opens at photon energies of $2m_{\text{muon}} \approx 211$ MeV. However, the cross-section for muon production is several orders of magnitude less than that for electron–positron production. While muons *can* be of concern for general radiation protection (exposure to human beings), electrons and positrons are usually the only particles produced in sufficient quantities to be considered for machine protection purposes.

6.4 Electromagnetic cascades

At sufficiently high energies, the energy loss of electrons is dominated by bremsstrahlung, and the main interaction of photons with matter is the production of electron–positron pairs. Combined, these two effects create the phenomenon of an *electromagnetic cascade* or *shower*. As illustrated in Fig. 6, bremsstrahlung photons induce pair production, and the newly created electrons and positrons in turn generate bremsstrahlung when they interact with the nuclei of the material. These new photons can produce additional e^+/e^- pairs, and therefore the number of particles involved in the cascade increases exponentially until the energies are low enough to favour different processes. Hence, the effect of an electromagnetic cascade is the dispersal of transported energy from a few high-energy particles to many low-energy particles. These low-energy particles are mainly responsible for the energy transfer to the material.

6.5 Simplified cascade model

It is possible to derive a coarse estimate of the penetration depth of an electromagnetic shower from a very simple model of the cascade (see, e.g., Ref. [16]). Exploiting the fact that the characteristic length

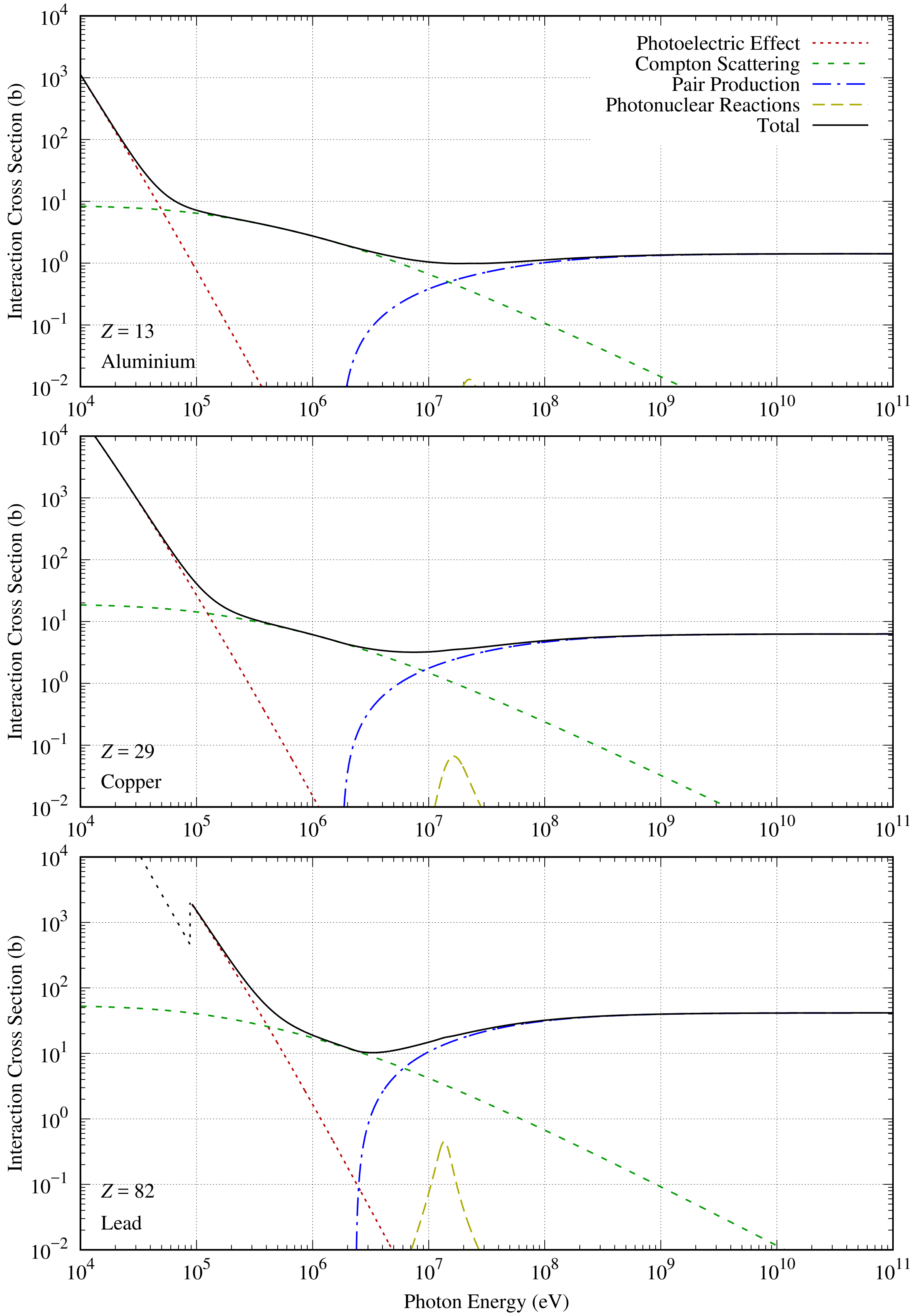


Fig. 5: Total cross-sections for photonic interactions in aluminium, copper, and lead

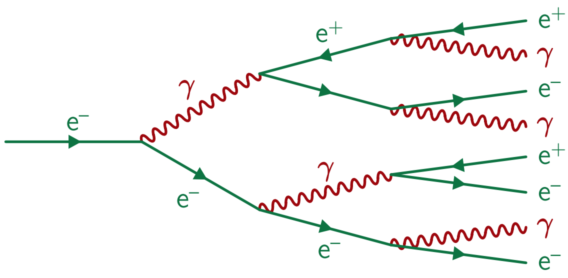


Fig. 6: Electromagnetic cascade as the sequence of emission of bremsstrahlung and pair production

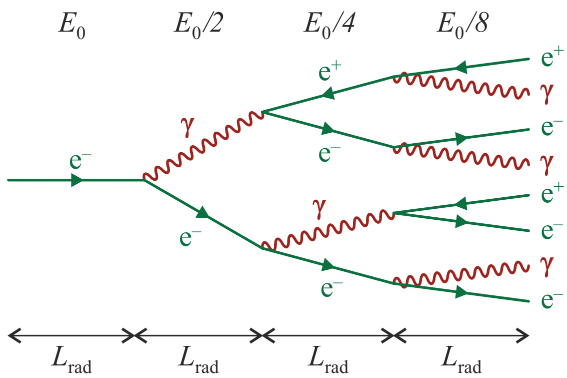


Fig. 7: Simplified model of electromagnetic cascade

scale of the problem is the radiation length of the material, three basic assumptions are made.

- An electron emits half of its energy as a single photon after travelling a distance L_{rad}.
- A photon is converted to an e^+/e^- pair, each carrying half of its energy, after L_{rad}.
- The shower stops when particle energies drop below the critical energy.

Apart from some general approximations, these assumptions basically reduce statistical statements to deterministic rules for individual particles.

Figure 7 shows the evolution of the cascade in this simplified model. The cascade starts from a single electron of energy E_0, which emits a photon of energy $E_0/2$ after one radiation length. More generally, after N radiation lengths, there are 2^N particles, each with energy $E_0/2^N$.

After a certain number N_{crit} of radiation lengths, the particle energy has decreased to the critical energy of the material, and the cascade stops. From

$$E_{\text{crit}} = \frac{E_0}{2^{N_{\text{crit}}}} ,$$

it is straightforward to calculate the number of particle generations in the shower:

$$N_{\text{crit}} = \frac{\ln(E_0/E_{\text{crit}})}{\ln(2)} . \tag{2}$$

It is not obvious how to interpret this remarkably simple result—clearly, a real cascade does not come to a sharp stop after a certain distance. To get a feeling for the physical meaning of Eq. (2), it is instructive to calculate the values of N_{crit} for a few scenarios and to compare them against the results of more sophisticated simulations. Considering an electron beam hitting a (large) copper block, Eq. (2) yields the following values for a critical energy of $E_{\text{crit}}(\text{Cu}) = 25$ MeV:

- for a beam energy of 100 MeV: $N_{\text{crit}} \approx 2$;

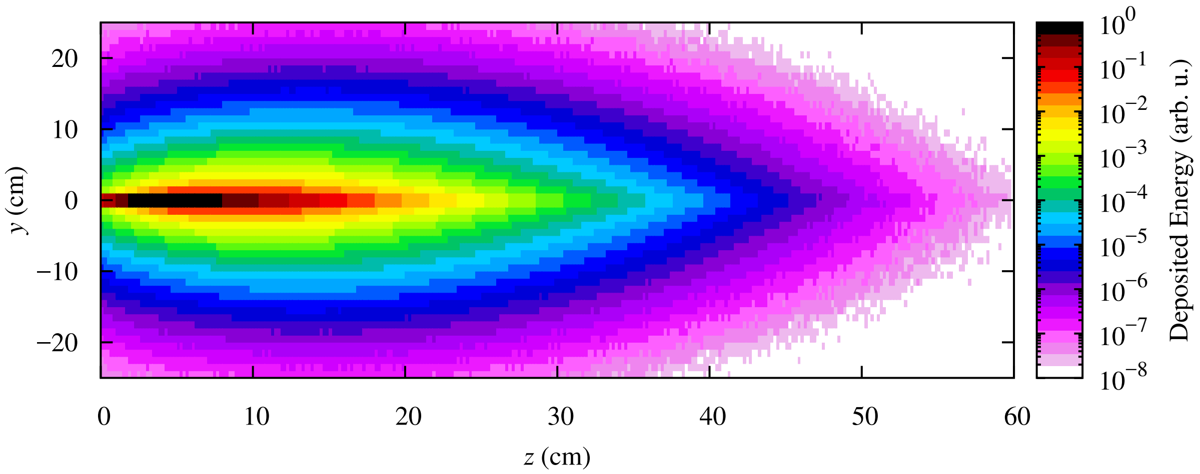


Fig. 8: Energy deposition by a 1 GeV pencil electron beam impinging on a big copper target. The deposited energy is averaged over the range -2 cm $< y <$ 2 cm.

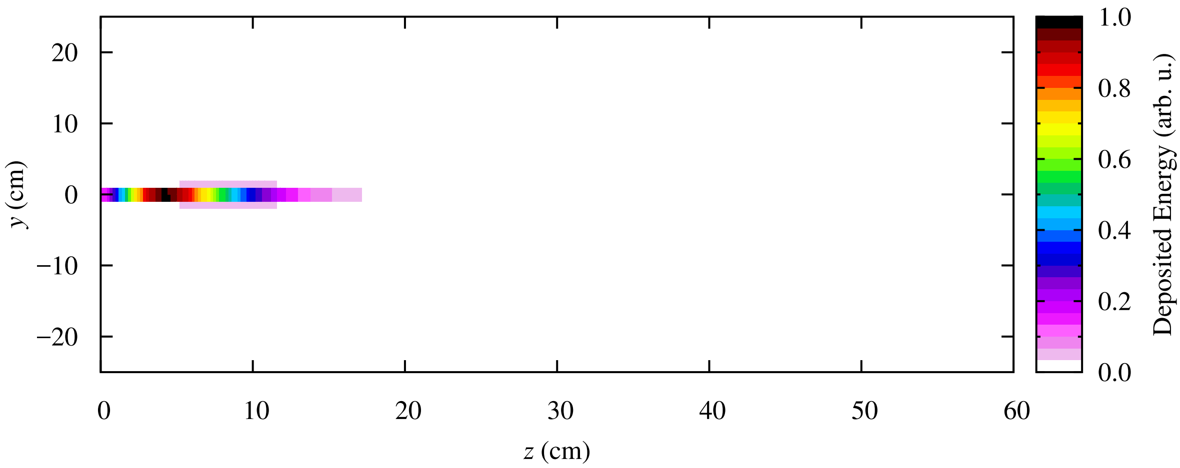


Fig. 9: The same plot as Fig. 8, but with a linear energy scale

- for a beam energy of 1 GeV: $N_{\rm crit} \approx 5.3$;
- for a beam energy of 10 GeV: $N_{\rm crit} \approx 8.6$.

6.6 Monte Carlo simulations

The propagation of electromagnetic cascades in the presence of obstacles or shielding can be modelled very well by Monte Carlo simulations. In such simulations, lots of particles are tracked, and their interactions with the material are modelled by random sampling from the actual physical probability distributions. The end result of such a simulation is always obtained by averaging over the contributions of all individual particles to the quantity of interest (e.g., energy deposition or fluence). Widely used and fairly complete simulation codes—not only for electromagnetic, but also for hadronic problems—are, for instance, FLUKA [19, 20] and Geant4 [21, 22].

We can continue the discussion of the simplified shower model by comparing its results with those obtained from a simple FLUKA simulation. In this simulation, we let a (pencil) electron beam impinge on a huge target made of copper. The deposited energy in the material is scored on a Cartesian three-dimensional grid. Figure 8 shows the result for a beam energy of 1 GeV. The deposited energy is averaged over a slice of the target geometry (namely, the range with -2 cm $< y <$ 2 cm) and indicated by false colours distributed along a logarithmic scale. This kind of plot is typical for all kinds of machine

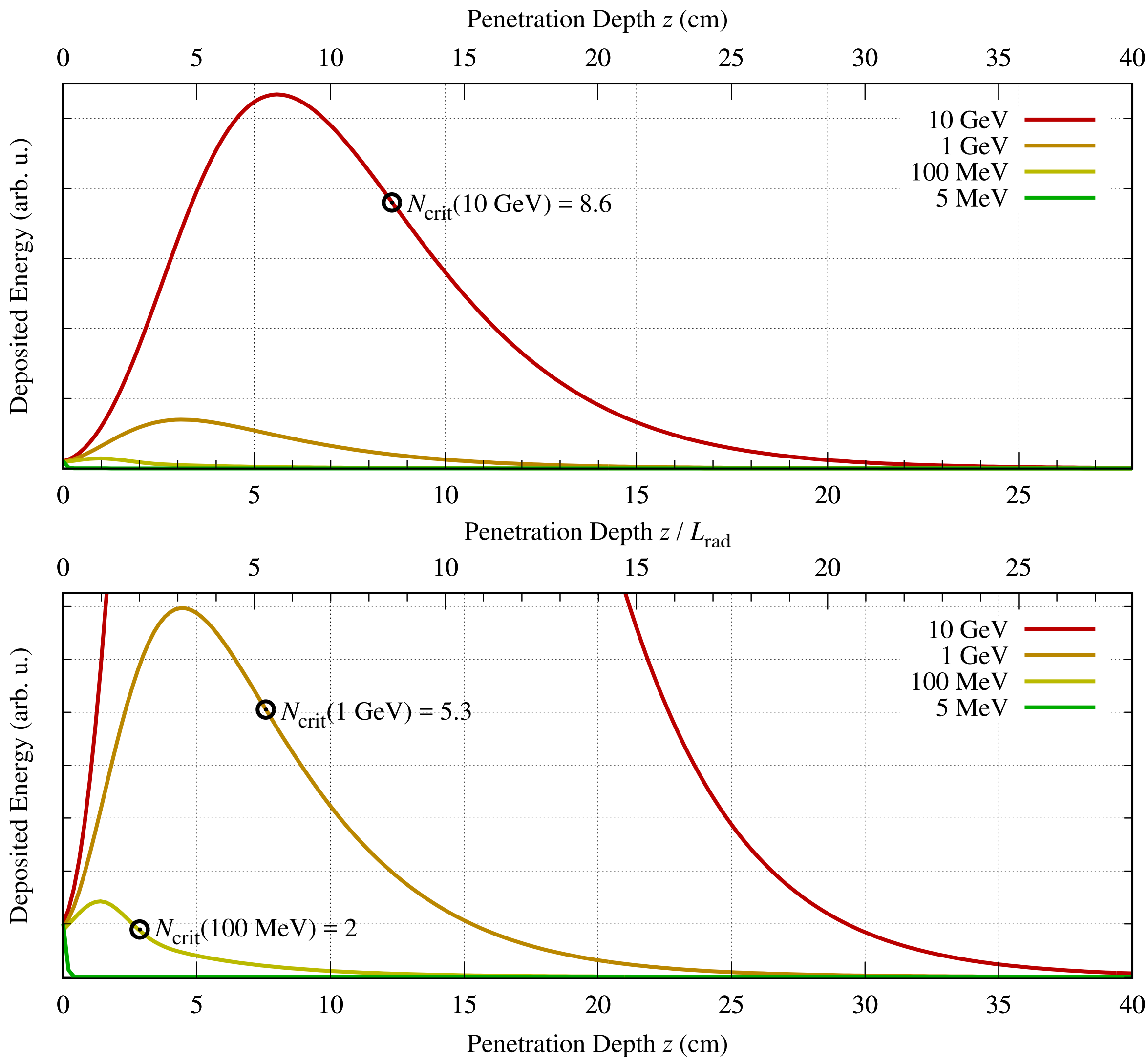


Fig. 10: Energy deposition by electron beams of various energies in a huge copper target. The beam impinges from the left at $z = 0$. Both plots show the same curves with different scales for the vertical axis.

and radiation protection studies where walls and other shielding implements attenuate radiation fields by many orders of magnitude; it can, however, be misleading because of the logarithmic colour scale. Where Fig. 8 seems to suggest that a substantial amount of energy is deposited far off-axis, the same plot with a linear energy scale (Fig. 9) shows clearly that this is not the case.

Figure 10 shows the projection of the deposited energy onto the z axis. This view is especially useful for judging the penetration depth of the shower in the material. Because the simulated copper target is huge, it absorbs all of the energy of the incoming beam. Therefore, a 10 GeV beam deposits ten times more total energy than a 1 GeV beam, which in turn deposits ten times more energy than a 100 MeV beam. All of these high-energy beams show similar curve shapes: the energy deposition rises steeply to a maximum and afterwards decreases ever more slowly, with a long tail to large penetration depths. This is simply a consequence of the statistical nature of an electromagnetic cascade: there are always some high-energy photons that traverse long stretches of the material before interacting with it. The behaviour of the electron beam at 5 MeV is different—because the beam energy is far less than the critical energy, its energy loss is dominated by inelastic electron–electron scattering, and very little bremsstrahlung is generated.

The figure also allows the results for N_{crit} from Eq. (2) to be put into context. The simple shower model reliably predicts a shower depth that is a good way beyond the actual shower maximum. For all its simplicity, the model can therefore be used for quick shielding estimates—at a thickness of about 2–3 N_{crit} radiation lengths, most of the shower energy has been absorbed by the material.

7 Damage to permanent magnets

Free-electron lasers and energy recovery linacs used as light sources usually depend on undulators or wigglers made of permanent magnets to extract synchrotron radiation from the electron beam. Unfortunately, the beam can also damage them: permanent magnets gradually lose their magnetization under irradiation (see e.g., Refs. [23–25]). This problem is of particular concern for machine protection at light sources because:

- it is cumulative (even small beam losses or dose rates can cause a deterioration of the field over longer time-scales).
- it is often not possible, or at least very expensive, to exchange an undulator.
- the undulators represent one of the smallest apertures in the accelerator (the SACLA in-vacuum undulators have a minimum gap of 3.5 mm [26]).
- in FELs, the lasing process itself depends crucially on a high precision of the magnetic field.

As a general rule, designers of insertion devices already prefer magnetic materials of higher coercivity because they are more radiation-resistant. Nonetheless, measurements after a few years of operation in an accelerator sometimes reveal significant loss of field. For example, a sacrificial permanent magnet structure installed at FLASH lost 3% of its initial field after 3 years of operation [27], and the first period of an undulator from the Petra-II storage ring was reduced to almost half of its magnetization in the lifetime of the machine, with visible signs of demagnetization continuing at least until the 20th period of the device [28,29]. These measurements are alarming enough that we should examine the effect of a partially demagnetized undulator on the emission of synchrotron radiation in some detail.

7.1 Effect of demagnetization in an undulator

For a number of reasons, typical beam loss scenarios cause a very inhomogeneous dose deposition along the longitudinal axis of an undulator. The strongest demagnetization is usually to be expected in the first periods at the upstream end of the magnet structure.

To understand the effect of a partially demagnetized undulator, we can track a single electron through the centre of a perfect undulator field with a simple two-dimensional tracking code. At each turning point n of the undulating trajectory (where the transverse velocity is zero), we note the longitudinal slippage Δz_n between the electron and a photon emitted at the undulator entrance. In an ideal undulator, this slippage simply increases by one radiation wavelength λ_{r} for each full period of the undulating motion, $\Delta z_{n+2} - \Delta z_n = \lambda_{\mathrm{r}}$.

If the field amplitude at the undulator entrance is reduced, the electron motion is no longer synchronous with the nominal radiation wavelength—the particle effectively takes a straighter trajectory and therefore gets ahead of where it should be. This effect can be described by a phase error $\Delta\phi$. At each trajectory turning point n, we define:

$$\Delta\phi_n = 2\pi \cdot \frac{n\,\lambda_{\mathrm{r}}/2 - \Delta z_n}{\lambda_{\mathrm{r}}} + \phi_0\,,$$

where the starting phase ϕ_0 can be chosen at will (as by a phase shifter chicane in a real-world FEL).

We are going to use the parameters of an undulator in the final stage of the FEL-2 line of FERMI (Table 4) to study a real-world example. To simulate radiation-induced demagnetization, the ideal undu-

Table 4: Undulator and electron beam parameters for phase error calculation

Number of periods	66
Period length	3.48 cm
Field amplitude	1.105 T
Electron energy	1.25 GeV
Wavelength (fundamental)	21.7 nm

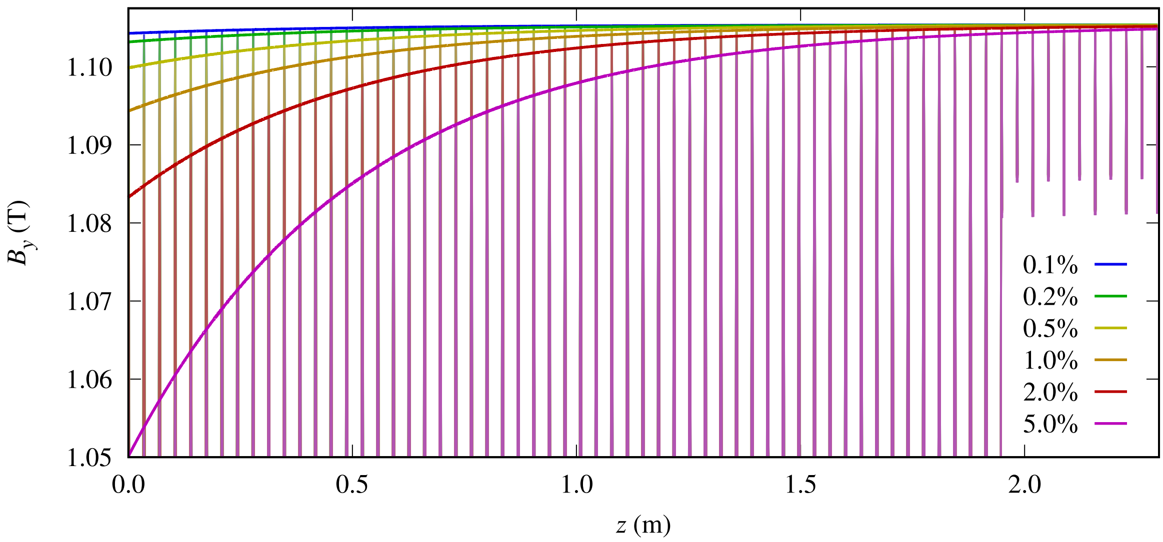


Fig. 11: Magnetic field profiles used for phase error calculations. Thick lines show the tapering of the field amplitude; the actual oscillating magnetic field is indicated by thinner lines below.

lator field $B_y(z)$ is tapered exponentially according to

$$B'_y(z) = B_y(z) \cdot \left(1 - d\mathrm{e}^{(-z/L)}\right) ,$$

with $L = 0.5$ m and a factor d specifying the relative demagnetization at $z = 0$. Field profiles for values of d between 10^{-3} and $5 \cdot 10^{-2}$ are shown in Fig. 11.

The resulting phase errors are displayed in Fig. 12. For ease of comparison, they have been adjusted (via ϕ_0) to coincide at zero at the exit of the undulator. It can be seen that the phase errors quickly reach large values: already at a tapered demagnetization of 1%, the electron bunch is out of phase by 90° at the undulator entrance. For higher values of d, the electron bunch in the first part of the structure effectively cancels out a part of the radiation through destructive interference. It should be noted that inhomogeneous phase errors like this can only partially be compensated by adjusting the undulator gap—after all, opening the undulator gap is equivalent to the introduction of a linear phase slope.

Obviously, the effect on the microbunching and on the final output power of an FEL needs to be studied in the context of the whole system of insertion devices and electron beam optics. It is clear, however, that even a small loss of magnetic field can have a big influence on the performance of an undulator.

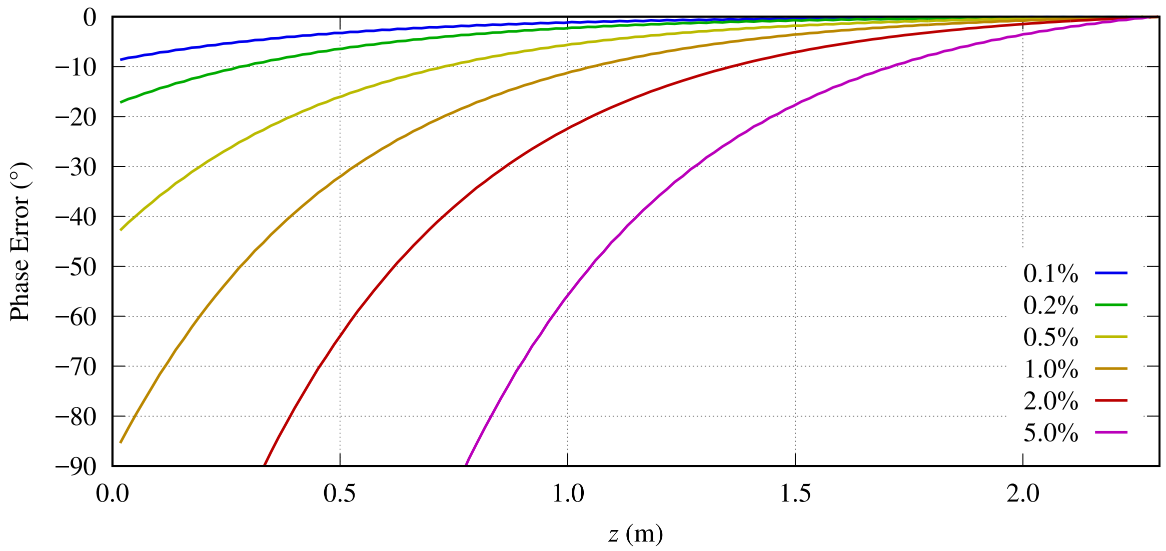


Fig. 12: Resulting phase errors for various demagnetization profiles. The reference point for all phases is the exit of the undulator.

8 Conclusions

All synchrotron light sources share a common set of machine protection problems: the limitation of induced activation, the protection of components from radiation damage, and the protection of permanent magnet undulators from demagnetization. The latter is of particular concern for FELs because magnet damage can directly impair the FEL process itself. The high beam power of superconducting or energy recovery linacs makes all of these problems much more challenging and adds the potential for direct or indirect mechanical damage. In this context, it is quite natural to see catastrophic, sustained losses of the entire beam as the main danger in high-powered accelerators. However, while such an incident can obviously have dramatic consequences, it is also relatively easy to detect and to avoid. The more serious problem with powerful beams is that even tiny fractional losses can represent huge absolute power depositions: a loss of 10^{-4} of a 1 MW beam is not easily measurable by beam current monitors, but it still represents 100 W of power. To control beam losses to this level, possible causes have to be understood and many aspects of the accelerator operation have to be optimized. Ultimately, the goal of machine protection is to avoid damage to expensive components and to prevent the loss of beam time—one of the most precious resources at any light source. The best approach to this goal is not to reduce machine protection to a mere system of interlocks, but to make safety considerations an integral part of the design and operation of an accelerator.

References

[1] C. Sibley, Machine protection strategies for high power accelerators, Proc. PAC'03, Portland, USA, 2003 (IEEE, New York, 2003), p. 607, http://doi.org/10.1109/pac.2003.1288989.

[2] R. Schmidt *et al.*, *New J. Phys.* **8** (2006) 290, http://doi.org/10.1088/1367-2630/8/11/290.

[3] K. Wittenburg, in Proc. CERN Accelerator School on Beam Diagnostics, Dourdan, France, 28 May–6 June 2008, edited by D. Brandt (CERN, Geneva, 2009), p. 249, http://doi.org/10.5170/CERN-2009-005.249

[4] L. Fröhlich, Machine protection for FLASH and the European XFEL, Ph.D. thesis, Hamburg University, 2009, http://doi.org/10.3204/DESY-THESIS-2009-012.

[5] L. Fröhlich, Machine protection for single-pass FELs, Proc. FEL'12, Nara, Japan, 2012 (JACoW, Geneva, 2012), p. 345.

[6] R. Schmidt, in Proc. CERN Accelerator School on Advanced Accelerator Physics, Trondheim, Norway, 18–29 August 2013, edited by W. Herr (CERN, Geneva, 2014), p. 221, http://dx.doi.org/10.5170/CERN-2014-009.221

[7] J. Wenninger, State-of-the-art and future challenges for machine protection systems, Proc. IPAC'14, Dresden, Germany, 2014 (JACoW, Geneva, 2014), p. 4060.

[8] I. Strašík, Machine protection, CERN Accelerator School: Introduction to Accelerator Physics, Prague, Czech Republic, 2014, http://cas.web.cern.ch/cas/CzechRepublic2014/Lectures/Strasik.pdf.

[9] N. Walker *et al.*, Operation of the FLASH linac with long bunch trains and high average current, Proc. PAC'09, Vancouver, Canada, 2009 (JACoW, Geneva, 2009), p. 2766.

[10] *The International Linear Collider technical design report vol. 3.II: Accelerator baseline design*, edited by C. Adolphsen *et al.*, 2013, http://www.linearcollider.org/ILC/Publications/Technical-Design-Report.

[11] G.R. Lynch and O.I. Dahl, *Nucl. Instrum. Methods Phys. Res. B* **58** (1991) 6, http://doi.org/10.1016/0168-583X(91)95671-Y.

[12] M.J. Berger *et al.*, Stopping-power & range tables for electrons, protons, and helium ions, NISTIR 4999, National Institute of Standards and Technology, 2005, http://www.nist.gov/pml/stopping-power-range-tables-electrons-protons-and-helium-ions.

[13] H. Bethe, *Ann. Phys.* **397** (1930) 325, http://doi.org/10.1002/andp.19303970303.

[14] F. Bloch, *Ann. Phys.* **408** (1933) 285, http://doi.org/10.1002/andp.19334080303.

[15] S.M. Seltzer and M.J. Berger, *Int. J. Appl. Radiat. Isot.* **35** (1984) 665, http://doi.org/10.1016/0020-708X(84)90113-3.

[16] W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments: A How-To Approach* (Springer, Berlin, 1994), http://doi.org/10.1007/978-3-642-57920-2.

[17] W.-M. Yao *et al.* (Particle Data Group), *J. Phys. G* **33** (2006) 1, http://doi.org/10.1088/0954-3899/33/1/001.

[18] K.A. Olive *et al.* (Particle Data Group), *Chin. Phys. C* **38** (2014) 090001, http://doi.org/10.1088/1674-1137/38/9/090001.

[19] T.T. Böhlen *et al.*, *Nucl. Data Sheets* **120** (2014) 211, http://doi.org/10.1016/j.nds.2014.07.049.

[20] A. Ferrari *et al.*, FLUKA: a multi-particle transport code, CERN-2005-10 (2005), INFN/TC_05/11, SLAC-R-773.

[21] S. Agostinelli *et al.*, *Nucl. Instrum. Methods A* **506** (2003) 250, http://doi.org/10.1016/S0168-9002(03)01368-8.

[22] J. Allison *et al.*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, http://doi.org/10.1109/TNS.2006.869826.

[23] A.B. Temnykh, *Nucl. Instrum. Methods A* **587** (2008) 13, http://doi.org/10.1016/j.nima.2008.01.002.

[24] Y. Asano *et al.*, *J. Synchrotron Radiat.* **16** (2009) 317, http://doi.org/10.1107/S0909049509008899.

[25] T. Bizen, Brief review of the approaches to elucidate the mechanism of the radiation-induced demagnetization, ERL Workshop 2011, Tsukuba, Japan.

[26] C. Pellegrini, The challenge of 4th generation light sources, Proc. IPAC'11, San Sebastián, Spain, 2011 (JACoW, Geneva, 2011), p. 3798.

[27] J. Skupin *et al.*, Undulator demagnetization due to radiation losses at FLASH, Proc. EPAC'08, Genoa, Italy, 2008 (JACoW, Geneva, 2008), p. 2308.

[28] H. Delsim-Hashemi *et al.*, Status of the sFLASH undulator system, Proc. FEL'09, Liverpool, UK, 2009 (JACoW, Geneva, 2009), p. 500.

[29] P.V. Vagin *et al.*, *J. Surf. Invest.* **5** (2011) 1055, http://doi.org/10.1134/S1027451011110218.

Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Different Applications of Energy Recovery Linacs

K. Aulenbacher
Institut für Kernphysik der Johannes Gutenberg-Universität, Mainz, Germany

Abstract
Besides their application for radiation generation, energy recovery linacs may become a unique tool for scattering experiments in nuclear and particle physics. Applications for fixed target and also for collider experiments are discussed. Spin polarized operation is an essential feature which requires additional attention.

Keywords
Energy recovery linac; electron ion collider; Mainz energy-recovering superconducting accelerator.

1 Introduction

The proposal to use energy recovery for electron-collider experiments by Maury Tigner dates back to the 1960s [1]. However, it soon became apparent that synchrotrons would offer better conditions, thus the idea was not pursued intensely until the end of the century. Nowadays, the limitations of the storage ring concept are well known and scientists have revisited the advantages linac-based experiments would offer. The advantages of Energy Recovery Linacs (ERLs) for radiation production have been discussed in another contribution to this school [2]. In this paper I will mention the possibilities for scattering experiments with ERLs that serve particle or nuclear physics studies.

Such experiments have been pursued in the past with linacs and also storage rings, the Stanford Linear Collider (SLC) and the Large Electron Positron collider (LEP) being typical large-scale examples. In Section 2 I will address the specific advantages that ERLs may offer compared to the established systems. Typical experiments are discussed together with their physics goals in Section 3. Present ERL designs always incorporate recirculations, therefore their energy range is limited to $\approx$ 100 GeV due to synchrotron radiation losses. In this energy region, the investigation of spin-dependent interactions is an important possibility. The requirements of spin operation are discussed in the final Section 4, together with some specific hardware needed for such experiments.

2 Applications for ERLs in particle physics

ERLs have hitherto not been used for particle physics experiments, although the initial suggestion of Tigner pointed in exactly this direction. The reason is of course that other accelerator types have served the purposes of particle physics with extraordinary success. However, accelerator-based experiments are presently facing tremendous challenges. One of these challenges, increasing the beam energy, cannot be efficiently addressed with present-day electron ERLs due to the fact that the recirculation system will produce intense synchrotron radiation and consequently limit the luminosity at high energies. The advantages of ERLs have therefore to be sought at low energies ($<$100 GeV or even much lower) where unprecedented experimental conditions can be achieved which are not accessible with the established accelerators. Two such regimes have been proposed, which we discuss in the following subsections.

2.1 Low background in fixed target experiments at low energies

If a beam hits a target at rest which has dimensions larger than the beam itself, the reaction rate is given by

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.
https://doi.org/10.23730/CYRSP-2018-001.631

$$R = L \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \Delta\Omega \tag{1}$$

The symbols denote L: luminosity, $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}$ differential cross-section, $\Delta\Omega$: solid angle of the detector system. If the interest of the experimenter is in achieving high statistical accuracy—for instance if the cross-section to be measured is very low—a high luminosity is desired. In a fixed target experiment the luminosity can be increased by a high areal density which represents the product of the density ρ and the thickness (d_{target}) of the target:

$$L = \frac{N_{\mathrm{A}}}{A} \frac{I_{\mathrm{beam}}}{e} \rho_{\mathrm{Target}} d_{\mathrm{target}} \tag{2}$$

where e = electron charge, N_{A} = Avogadro's number, and A = atomic mass unit of the target.

Conventional fixed target experiments have the advantage of a potentially very high luminosity, for instance, the planed P2 experiment at the Mainz Energy-recovering Accelerator (MESA) facility in Mainz will be operated with a 60 cm long liquid hydrogen target and an external beam current (without energy recovery) of $150\,\mu$A [3], yielding a luminosity of $>10^{39}$ cm^{-2} s^{-1}. In this experiment, the areal density of the scattering centres which contribute to the reactions is approximately 2.5×10^{24} cm^{-2}.

The large number of scattering centres in the target creates limitations for precision experiments in several ways. One of them is the uncertainty caused by multiple scattering in the target which sets limits to the precision of the determination of scattering angle, energy loss, etc. Furthermore, multiple scattering leads to tails in the angular distribution which can cover large angular intervals. In consequence, the background created from such halo particles when they hit structural components in the vicinity of the experiment creates another source of systematic uncertainty. Yet another contribution of this kind is the enclosure of the target, which represents another background source. Moreover, the scattered signal particles have to penetrate the enclosure too. This will straggle their angular and energy distribution and hence further reduce the measurement accuracy. These effects become more and more pronounced with decreasing energy.

Whereas these effects are not limiting for experiments such as P2, other high-precision experiments call for a target without enclosure and low areal density. These demands are met, for instance, by gas jets—such targets are also called 'windowless gas targets'. The application of such a target, together with an ERL, was proposed by Heinemeyer et al. [4]. They observed that a thin target would only lead to a small deterioration of the beam quality which could allow for energy recovery of the beam. Figure 1 shows the schematic setup of an ERL employing such a target.

Experiments of this type have already been performed at storage rings for a long time, for instance with the HERMES target at HERA [5]. In a storage ring, precision experiments are burdened by the fact that the luminosity is varying with time and that the injection periods interrupt the data acquisition. Again, these disadvantages become more severe the lower the energy of the operation. On the other hand, in an ERL, each beam particle passes the target only once, which leads to stationary beam conditions. In order to distinguish this from the storage ring, where beam particles may pass the target billions of times, we call the ERL target a 'pseudo internal target' (PIT).

Therefore, in a beam energy range <1 GeV, a window of opportunity may exist in which operating an ERL can enable experiments with fixed targets which have hitherto been difficult or impossible. The existing Jefferson Laboratory ERL and the MESA facility at Mainz, which is currently under construction, both operate at energies in the 100 MeV range, where the shortened beam lifetime in a storage ring would make internal target experiments difficult. Experiments at these facilities aim to demonstrate the advantages of this new type of experimental regime. I discuss several such experiments below. A very encouraging result has already been obtained by a team formed by MIT/Bates and JLAB[1]. They

[1] MIT=Massachusetts Institute of Technology, Boston, USA; JLAB= Jefferson Laboratory, Newport News, USA

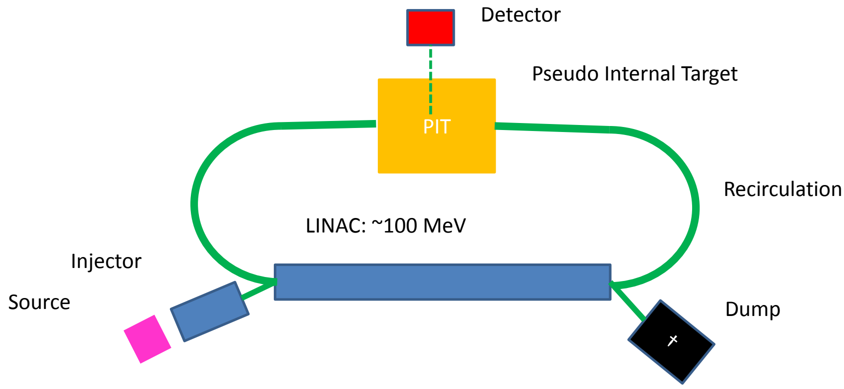


Fig. 1: Schematic of an ERL setup with 'pseudo internal target' (PIT)

successfully demonstrated transmission of a high intensity beam through a 127 mm long tube with a 2 mm aperture with negligible losses [6].

2.2 Electron ion colliders

Electron ion colliders (EICs) are operating in the multiple GeV energy range. They may be designed as ring/ring (RR) or as linac/ring (LR) machines. Figure 2 shows a schematic layout of such a LR configuration. Two examples which are currently under discussion for the LR configuration are the eRHIC project at BNL [7] and the LHeC [8] at CERN[2]. In both cases, a high energy ion accelerator already exists and will be complemented by an ERL. In the BNL design, the ERL will be integrated into the already existing tunnel and at CERN a dedicated electron ring separated from the LHC tunnel will be built.

Again it is useful to discuss what the substantial advantages are that can be achieved with respect to the established RR ansatz. A first advantage is the enormously reduced complexity with respect to the spin degree of freedom. In a ring setup, depolarizing resonances have to be avoided and space consuming spin rotators are difficult to integrate. This is one of the main reasons why the LEP ring did not directly exploit spin degrees of freedom for its particle physics program, although important information was gained by using the depolarizing resonances as a tool for absolute energy calibration [9]. Such problems are virtually absent in a linac-based approach that offers high flexibility of spin orientation with very modest effort [10]. Due to the fast acceleration in the linac-type accelerator, depolarization is virtually absent.

As in the fixed target case, the luminosity is a salient ingredient to obtain sufficient reaction rates. For the case of a collider, a simplified formula under the assumption of equal emittances (ϵ) and beta functions at the interaction point (β^*) is

$$L = f_{\mathrm{coll}} \frac{N_{\mathrm{el}} N_{\mathrm{Ion}}}{\epsilon \beta^*}, \tag{3}$$

where N_{Ion} is the number of ions per bunch. The bunch collision rate, f_{coll}, times the number of particles per electron bunch, $e f_{coll} N_{el}$, is the beam current which can surpass 1 A in a storage ring. The virtual

[2]eRHIC=electron Relativistic Heavy Ion collider at Brookhaven National Laboratory (BNL), Upton, USA, LHeC=Large electron Hadron Collider at Centre des Etudes des Recherches Nucleaires (CERN), Geneva Switzerland.

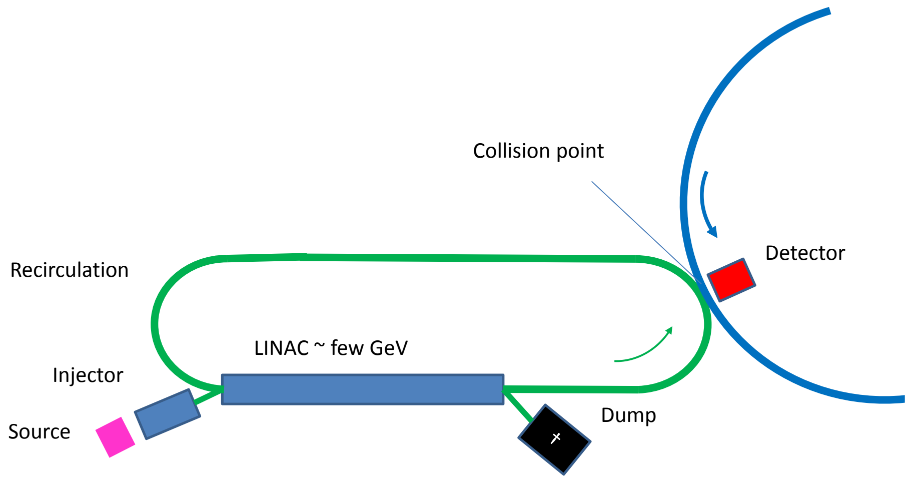


Fig. 2: Schematic layout of a linac/ring (LR) configuration

beam power at the collision point (beam current times energy) is ≈GW which is of no concern in a storage ring but makes energy recovery mandatory for a linac.

The fact that the emittance (ϵ) of the lepton beam [2] can be smaller than the equilibrium emittance in the ring can create an advantage for the generation of high luminosities. This of course also calls for a similar emittance of the ion beam. Recently, promising concepts such as coherent cooling [11] have been suggested for ion beams which might help to increase the luminosity even further. Another potential advantage can be realized if one takes into account that stable operation of a ring collider may be limited by the beam–beam tune shift which is proportional to the bunch charge of the other beam species [12].

$$\Xi_{\mathrm{Ion}} \propto \frac{N_{\mathrm{el}}}{\gamma_{\mathrm{Ion}}} \; ; \; \Xi_{\mathrm{El}} \propto \frac{N_{\mathrm{Ion}}}{\gamma_{\mathrm{el}}} . \tag{4}$$

This offers a means to circumvent the beam tune shift by reducing the number of electrons per bunch at the expense of a higher N_{Ion}. The increased tune shift of the electrons can be handled, since the electrons pass the target only once. A further advantage is that the difficulties of producing a high average current of spin polarized electrons (see Section 4 below) can be mitigated in this regime.

It should be noted that for experiments using (polarized) *positron* beams the RR concept is better. A linac-based experiment needs a particle source with an average current in the mA range which is presently not feasible for positrons. But the majority of experiments suggested for such colliders aim at hadronic observables which are probed with the lepton beam. In this case positrons do not give much of an advantage.

3 Particle physics experiments at ERLs

As an example of possible fundamental physics applications we discuss the experimental portfolio for the MESA ERL. The ERL will be operated with a beam energy of 105 MeV. The machine is currently being built at the Johannes Gutenberg-University in Mainz, Germany. Figure 3 gives an overview of the accelerator and its experiments.

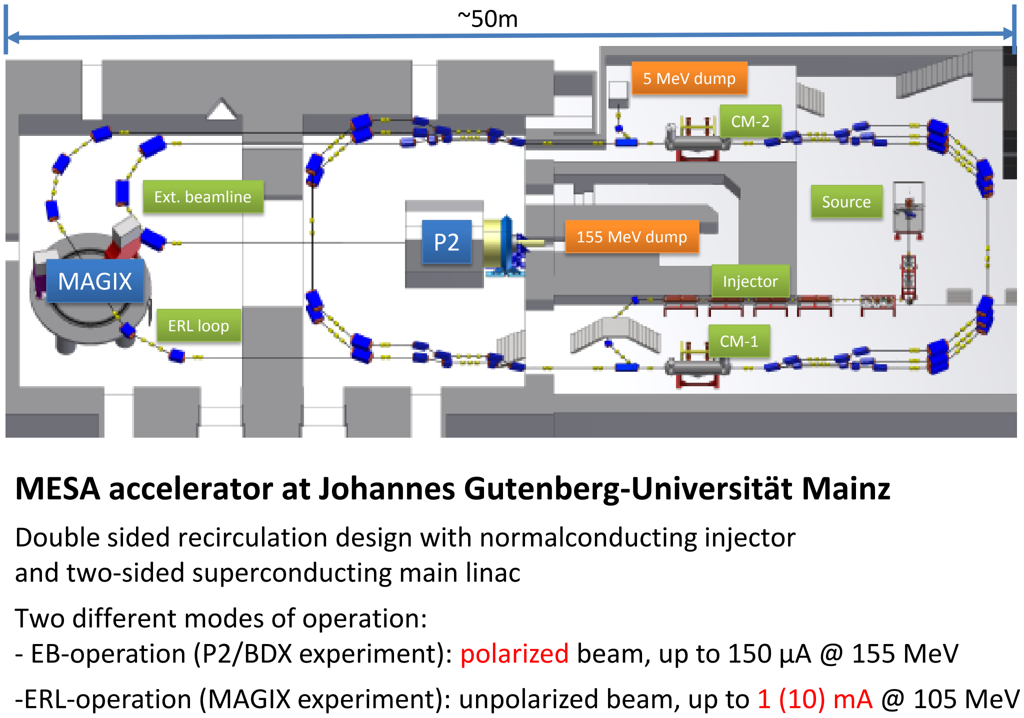


Fig. 3: Schematic layout of the MESA accelerator

3.1 Brief description of the MESA accelerator

MESA (**M**ainz **E**nergy-recovering **S**uperconducting **A**ccelerator) will use a spin polarized photo source with currents up to 1 mA. The beam time structure will be 1.3 GHz c.w. (continuous wave) which minimizes the bunch charges and therefore the beam dynamical issues associated with the charge. Nevertheless, space charge related effects are still important. The P2 experiment requires sophisticated spin manipulation techniques which lead to a relatively long low energy beam transport system between the source and the injector linac. Therefore larger currents than 1 mA will be difficult to handle, hence for experiments with even higher demands, we plan to install another source for an unpolarized beam closer to the injector linear accelerator (ILAC), which will allow us to achieve 10 mA of beam current.

After passing the injector the beam will have 5 MeV and can be injected in the first cryomodule where an energy gain of 25 MeV is achieved by 2 slightly modified TESLA-type cavities. The beam is then sent through a spreader, a 180 degree deflection and a recombiner which latter is almost identical to the spreader, and then enters the second cryomodule. The different recirculation arcs can be used to pass the cryomodules 3 times, yielding 155 MeV for external beam operation with the P2 experiment. This experiment is not of direct relevance here since it operates with the conventionally extracted beam. The "MAGIX" (**MA**inz Internal **G**as target **E**xperiment) experiment is integrated in another recirculation loop where 105 MeV will be the beam energy. To operate the experiment, the recirculation loop is extended into an additional hall in which the beam passes the MAGIX target and is sent back to the main linac. Since the loop represents a net 180 degree phase shift the beam is decelerated again through 2 recirculations. Afterwards, the beam leaves at the opposite side with respect to the injection point at 5 MeV.

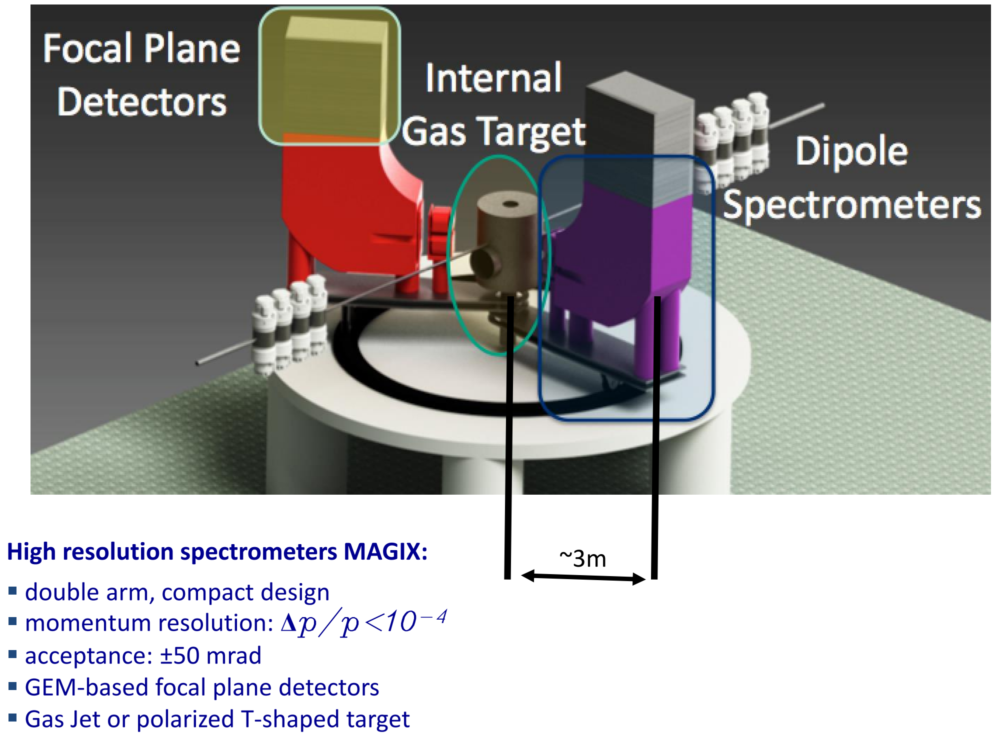


Fig. 4: Artists view of the MAGIX setup [13]

3.2 The MAGIX experiment

Figure 4 gives an impression of the planned MAGIX setup. Two magnetic spectrometers are employed in order to perform coincidence experiments. The spectrometers operate with a bending radius $\approx 1\,\text{m}$ which indicates that the setup will be quite compact. They can achieve a momentum resolution of $\Delta p/p < 10^{-4}$. This sets a corresponding requirement for the energy definition of the MESA beam which should be at least the same or preferentially have a lower value. Specific detector systems are currently being designed which take into account the low energies of the scattered particles.

The target region is separated from the beam line vacuum by a differential pumping stage. Modern jet or cluster targets allow high areal densities $>5 \times 10^{18}\text{cm}^{-2}$ (see Section 4.1 below). With the planned beam current of MESA stage-1 (1 mA) this results in a luminosity of $>3 \times 10^{34}\ \text{s}^{-1}\ \text{cm}^{-2}$. Figure 5 shows the angular distribution after a beam with no angular spread has hit such a hydrogen target. The distribution has been produced by Monte Carlo simulation with Geant 4. The resulting widening of the angular distribution is negligible in terms of the natural divergence of an electron beam, at least as far as the root-mean-square value of the angular straggling is concerned.

However, this does not mean that operation of such a target with an ERL beam is straightforward. It is obvious that trajectories at arbitrary angles will exist due to elastic scattering. For very large angles ($\theta > 5\,\text{deg}$) such particles may reach the detectors and can be considered as a signal. For small angles ($\theta < 10\,\text{mrad}$) particles may fit into the acceptance of the beamline and the subsequent deceleration system. Particles in the interval between these regions are target-induced halo and must be absorbed (collimated) at well-defined positions. The stopping process should ideally not produce background in the detectors or produce radiation levels that could become harmful for hardware installed in the areas behind the target.

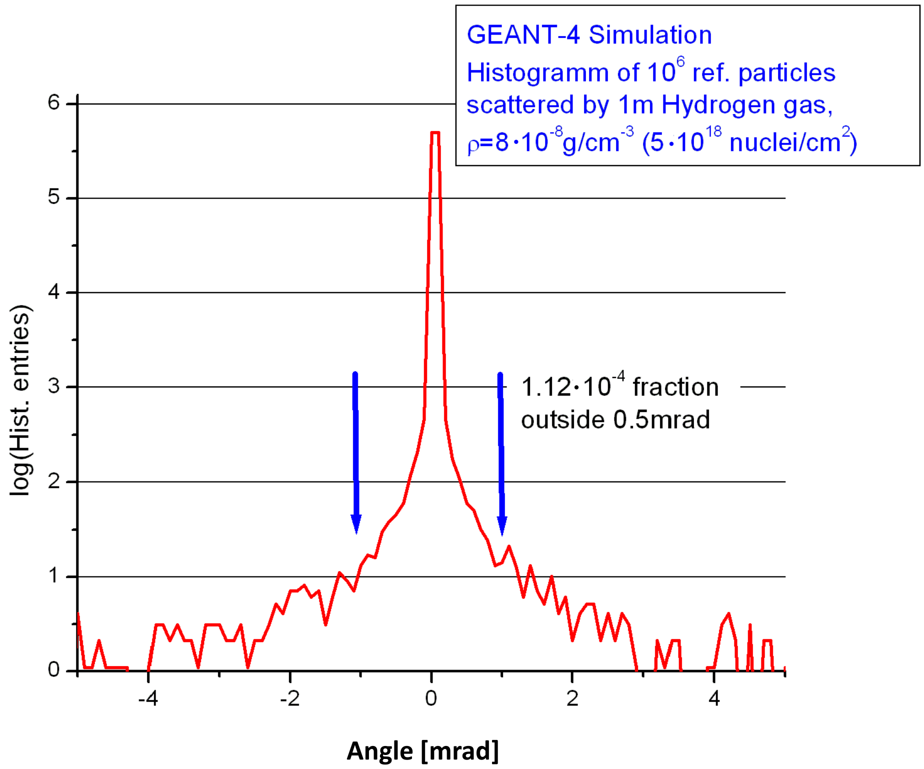


Fig. 5: Angular electron distribution behind target

3.3 Low background reactions: dark photons

The 'dark photon' (denoted here by γ') is a hypothetical gauge particle that could explain several anomalies related to astrophysical observations or the (g-2) anomaly of the muon [14]. Such a particle would behave like a photon but would have rest mass. The observations hint at a particle mass of the γ' of between 1 and 1000 MeV.

Since a gauge particle represents a force carrier, a charge is attributed to it. Though its effective interaction with ordinary matter will be very small, it is not zero. This is expressed via the Feynman graph in Fig. 6. In a scattering process radiation occurs, the vast majority of which is photon bremsstrahlung, but in very rare cases the dark photon could be produced due to the charge which causes a coupling ϵ.

If the dark photon preferentially decays into electron/positron pairs, the total energy of the pair will correspond to the rest mass of the γ'. Hence, if one observes such pairs and measures their momentum, a continuous background spectrum will be reconstructed on which the sharp peak resulting from the mass of the γ' is superimposed. The sensitivity of such a discovery increases with the resolution of the detectors, which motivates the use of magnetic spectrometers. Such measurements have been performed for instance at the Mainz Mikcrotron (MAMI), but also at many other accelerators, although a γ' has so far not been discovered [15, 16].

These measurements allow us to exclude certain ranges of charge and mass for the dark photon. The current (2016) status of this exclusion is presented in Fig. 7. MAGIX can cover hitherto uncovered area in the parameter space for 2 MeV $\leq m_{\gamma'} < 80$ MeV and for couplings $10^{-3} \geq \epsilon > 5 \times 10^{-5}$. Besides this, there is an even more interesting argument to use MAGIX in a slightly modified way.

In a specific region of the parameter space (red hatched area in Fig. 7) the observed (g-2) anomaly

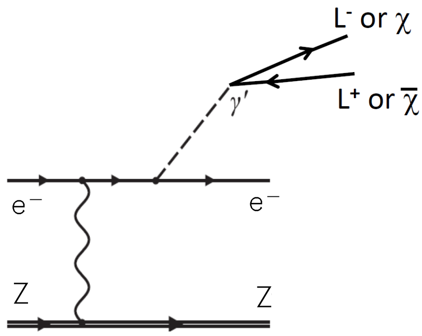


Fig. 6: Feynman graph for dark photon (γ') production during electromagnetic scattering on a nucleus of charge Z. The γ' could decay either into a lepton/antilepton pair (L^+/L^-) or into dark matter particles ($\chi/\overline{\chi}$).

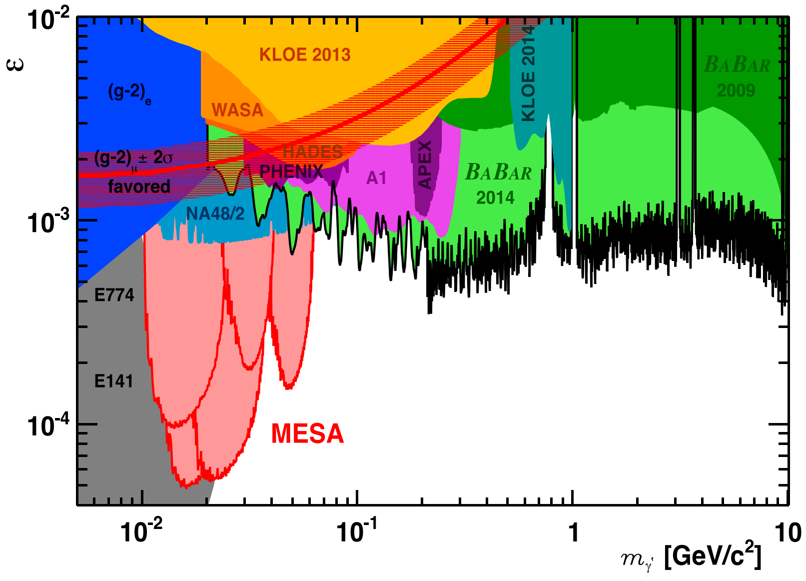


Fig. 7: Exclusion plots for the dark photon based on the assumption of a dominant reaction channel $\gamma' \rightarrow e^+ + e^-$

of the muon could be explained by the existence of a γ', but it has not been discovered there. However, the exclusion limits have been achieved under the assumption that the decay into e^+/e^--pairs is the dominant reaction channel. This may not be the case if dark particles $\chi, \overline{\chi}$ exist into which the γ' could decay preferentially (Fig. 6). In this case, a wide region of the interesting area remains unexplored. The $\chi-$ dark matter particles are invisible to detectors. Nevertheless, the rest mass of the γ' can still be reconstructed if the energy of the outgoing scattered electron and the recoil energy of the nucleus are measured. Due to the low recoil energy this is difficult in conventional targets but can be achieved in the windowless gas target of MAGIX.

3.4 Precision observables: form factors

Form factors are ground state properties of composite systems, for instance the proton. They depend on the four momentum transfer Q^2 in elastic scattering. Of particular interest is the extrapolation of the form factor towards $Q^2 = 0$. The slope of this extrapolation at $Q^2 = 0$ defines the charge radius of the particle. For the proton an exciting situation has occurred recently which has been dubbed the ‘proton

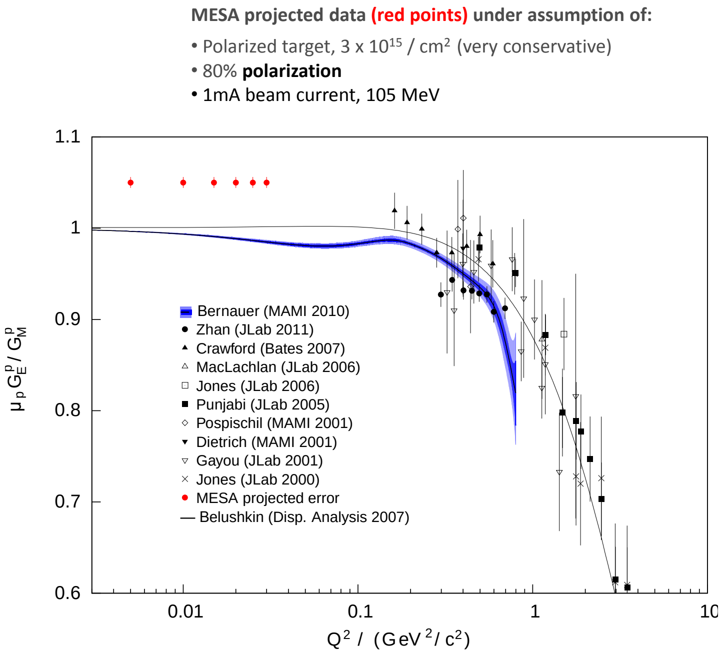


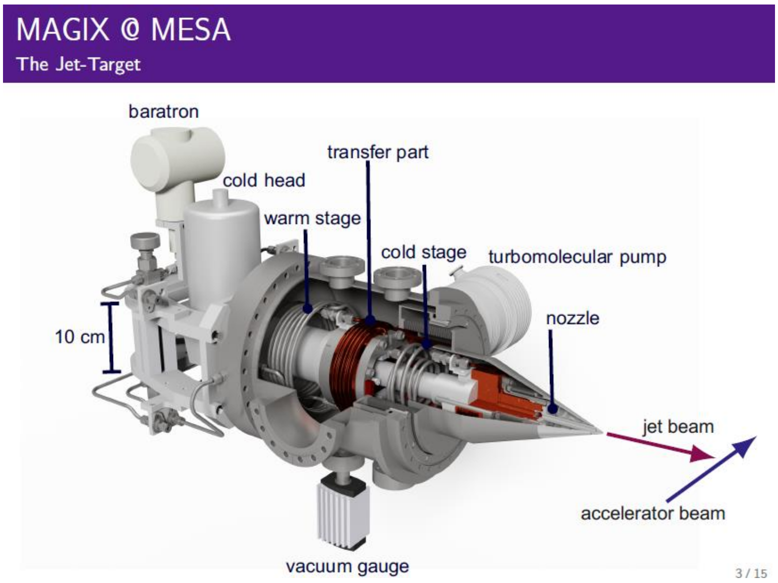
Fig. 8: Existing magnetic form factor data can be extended to much lower momentum transfers with the MAGIX setup using a 1 mA spin polarized beam. The red points stand for the expected accuracy of the data in the hitherto unexplored region of momentum transfers. The references given correspond to different experiments sand extrapolations that where undertaken during the last two decades [18–27].

radius puzzle'. This has been caused by the availability of high precision spectroscopic data from muonic hydrogen atoms which do not agree with the data obtained by electron scattering [17, 18]. The situation calls for data at lower momentum transfer in the electron scattering experiments, which so far have not been achievable due to the aforementioned strong beam target interaction in conventional experiments. Figure 8 presents the situation for the ratio of the electric to the magnetic form factor. The theoretical extrapolation (solid line) can be checked by MAGIX data points at the indicated level of precision, if an intense spin polarized beam is available .

It should be mentioned that MAGIX can also be used to observe other ground state properties such as the electromagnetic polarizabilities of the nucleons. To summarize, it can be stated the outlook for a rich physics program in low energy hadron physics at MAGIX looks very promising.

3.5 Nuclear physics at MAGIX

The possibility of exploring low momentum transfers opens new windows also for nuclear physics. For instance, in nuclear astrophysics, the reaction $^{12}\mathrm{C}+\alpha \rightarrow\ ^{16}\mathrm{O}+\gamma$ is highly important for stellar evolution but is so far not accessible experimentally due to the low energies relevant here. This can be changed by using MAGIX if the inverse reaction $^{16}\mathrm{O} + \gamma^* \rightarrow\ ^{12}\mathrm{C} + \alpha$ is investigated. Here, electron scattering creates exchange of a virtual photon γ^*. Again, the extrapolation to the photon point (mass of exchanged particle $= Q^2 = 0$) will deliver the desired information.



S. Grieser ;https://indico.mitp.uni-mainz.de/event/66/session/5/contribution/27/material/slides/0.pdf

Fig. 9: The jet target foreseen for MAGIX at MESA

4 Instrumentation for ERLs operating for particle physics

4.1 Windowless targets

In Fig. 9 we present the target setup which is presently discussed for MESA. Present day designs allow for unpolarized densities of 10^{19} cm^{-2} for various gas species [28], whereas polarized gas targets are much more difficult to handle. They need the application of a storage cell—typically about one 0.3–1 m long with a small aperture to increase the gas density inside. Such targets can achieve only $\approx 10^{16}$ cm^{-2} with high nuclear polarization. Experiments with such targets therefore require high electron currents, moreover, since single spin asymmetries are not particularly interesting in this energy range, a spin polarized electron beam is needed to perform double polarization experiments.

4.2 Spin polarized beams at ERLs

Spin polarized beams are produced by photoemission from so-called semiconductor superlattices [29]. Typically GaAs/GaAsP layers of a few nanometres thickness form a basic period which is repeated many times. This lifts the light hole/ heavy hole degeneracy in the semiconductor and allows therefore the efficient transfer of the angular momentum of the photons towards the spin of the electrons. In reality, this means that circularly polarized light can be transformed to spin polarized electrons inside the conduction (mini-) band of the superlattice. After photoexcitation, the electrons diffuse towards the surface from where they can escape if the work function of the photocathode is lowered by about a mono-atomic layer of caesium plus a certain amount of fluorine or Oxygen atoms [30]. The degree of spin polarization may reach almost 90% at typical laser wavelengths in the near infra-red (IR) (780 nm). The quantum efficiency at this wavelength can approach 1% which is in more practical units about 5 mA of electron current per watt of incident laser power. Modern laser systems can produce enough laser power to allow the production of a 100 mA electron beam which is presently a typical goal for light source ERLs [31] and spin polarized currents for the electron–ion collider projects are typically a little bit lower.

However, the weak point of the chain is the sustainability of the beam current, which is expressed by the so-called cathode lifetime τ. The lifetime during which the quantum efficiency drops to $1/e$ of its initial value is electron current dependent. This is due to the ionization of the residual gas which causes back bombardment of the cathode and corresponding radiation damage. The highest average currents that have been achieved for reasonable operation times ($>$100 hours) are a few milliampere only. The product $I \cdot \tau$ is called the charge lifetime [32]. There are indications that it is not current, but the electron fluence, which limits the lifetime, i.e. the extracted charge per square centimetre. Since the beam emittance is proportional to the beam radius at the cathode, one cannot increase the extractable charge during one lifetime arbitrarily without increasing the beam emittance. However, for a large acceptance injector, which could accept a normalized emittance of $\approx$20 μm in theory, charge lifetimes of 10 000 coulombs or more can be expected which could make operation for a collider feasible. Nevertheless, there are remaining challenges here. They can for instance be addressed by further lowering the base pressure and in consequence the amount of ion back bombardment. This could, for instance, be achieved by a cryogenic system and there is hope that, for instance, higher lifetimes could be observed in superconducting Radio-Frequency (RF-)guns. Another approach is being pursued at BNL where the beam time structure of the proposed eRHIC collider makes it possible to funnel beams from several cathodes onto a common orbit by interleaving them with a time-dependent deflector. This scheme has been called the 'Gatling-gun' and is currently under development at BNL [33].

In summary, the problem of lifetime should be tractable at the level desired for the collider projects but the effort and the resources needed should not be underestimated.

Acknowledgements

I wish to thank the organizers of the CAS workshop for giving me the opportunity to present these contents. This work has been supported by the Deutsche Forschungsgemeinschaft DFG through the cluster of excellence PRISMA.

References

[1] M. Tigner, *Nuovo Chim.* **37**(3) (1965) 1228. https://doi.org/10.1007/BF02773204

[2] A. Jankowiak, these proceedings.

[3] N. Berger *et al.*, *J. Univ. Sci. Tech. China* **46** (2016) 481.

[4] S. Heinemeyer *et al.*, arXiv:0705.4056v2 (2007).

[5] E. Steffens and W. Haeberli, *Rep. Prog. Phys.* **66**(11) (2003) 1887. https://doi.org/10.1088/0034-4885/66/11/R02

[6] R. Alarcon *et al.*, *Phys. Rev. Lett.* **111**(16) (2013) 164801. https://doi.org/10.1103/PhysRevLett.111.164801

[7] E. Metral (editor), ICFA Beam Dynamics Newsletter **58** (2012) 52.

[8] E. Metral (editor), ICFA Beam Dynamics Newsletter **58** (2012) 86.

[9] A. Assmann *et al.*, Energy calibration with resonant depoalrization at LEP, Proceedings third European Accelerator Conference EPAC 1994, London 935-937 (1994)

[10] B.S. Schlimme *et al.*, *Nucl. Instrum. Meth. A* **850** (2017) 54. https://doi.org/10.1016/j.nima.2017.01.024

[11] Y.H. Wu *et al.*, Beam dynamics studies for coherent electron cooling experiment, Proceedings IPAC2016, WEPOY023 (2016), http://accelconf.web.cern.ch/AccelConf/ipac2016/papers/wepmw027.pdf.

[12] A. Chao and M. Tigner (editors), *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 2002), p. 128.

[13] A. Denig, Workshop New Vistas in Low Energy Particle Physics, Mainz, 2016, https://indico.mitp.uni-mainz.de/event/66/overview

[14] G.W. Bennett *et al.*, *Phys. Rev. D* **73**(7) (2006) 072003. https://doi.org/10.1103/PhysRevD.73.072003

[15] J.P. Lees *et al.*, BaBar Collaboration, *Phys. Rev. Lett.* **113**(20) (2014) 201801. https://doi.org/10.1103/PhysRevLett.113.201801

[16] H. Merkel *et al.*, A1 Collaboration, *Phys. Rev. Lett.* **106**(25) (2011) 251802. https://doi.org/10.1103/PhysRevLett.106.251802

[17] R. Pohl et al., *Nature* **466**(7303) (2010) 213. https://doi.org/10.1038/nature09250

[18] J.C. Bernauer *et al.*., *Phys. Rev. Lett*. **105**(24) (2010) 242001. https://doi.org/10.1103/PhysRevLett.105.242001

[19] X. Zhan *et al*, *Phys. Lett B* **705**(49) (2011). https://doi.org/10.1016/j.physletb.2011.10.002

[20] J.B. Crawford *et al.*, *Phys. Rev. Lett.* **98** (2011) 052301. https://doi.org/10.1103/PhysRevLett.98.052301

[21] G. MacLachlan *et al.*, *Nucl. Phys. A* **764** (261) (2006). https://doi.org/10.1016/j.nuclphysa.2005.09.012

[22] M. K. Jones *et al.*, *Phys. Rev. C* **74** (2006) 242301. https://doi.org/10.1103/PhysRevC.74.035201

[23] V. Punjabi *et al.*, *Phys. Rev. C* **71** (2005) 055202. https://doi.org/10.1103/PhysRevC.71.055202

[24] T. Pospichil *et al.*, *Eur. Phys. J. A* **12** (125) (2001). https://doi.org/10.1007/s100500170046

[25] O. Gayou *et al.*, *Phys Rev C* **64** (2001) 038202. https://doi.org/10.1103/PhysRevC.64.038202

[26] M. K. Jones *et al.*, *Phys. Rev. Lett.* **84** (1398) (2000). https://doi.org/10.1103/PhysRevLett.84.1398

[27] M. A. Belushkin *et al.*, *Phys. Rev. C* **75** (2007) http://dx.doi.org/10.1103/PhysRevC.75.035202

[28] S. Grieser, Workshop New Vistas in Low Energy Particle Physics, Mainz, 2016, https://indico.mitp.uni-mainz.de/event/66/overview

[29] T. Maruyama *et al.*., *Appl. Phys. Lett.* **85**(13) (2004) 2640. https://doi.org/10.1063/1.1795358

[30] W.E. Spicer and A. Herrera-Gomez, Modern theory and applications of photocathodes, SLAC-PUB-6306 , (1993), http://slac.stanford.edu/pubs/slacpubs/6250/slac-pub-6306.pdf

[31] G.H. Hoffstatter, S.M. Gruner, M. Tigner (editors), Cornell Energy Recovery Linac, Project definition report (2013). https://www.classe.cornell.edu/rsrc/Home/Research/ERL/ErlPubs2013/PDDR.pd

[32] J. Grames *et al.*., Lifetimes Measurements of high polarization strained superlattice gallium arsenide at beam current > milliamp using a new 100 kV load lock photogun, Proceedings PAC2007, THPMS064 (2007). http://accelconf.web.cern.ch/AccelConf/p07/PAPERS/THPMS064.PDF

[33] X. Chang *et al.*., A multiple cathode gun design for the eRHIC polarized electron source, Proceedings PAC 2011, WEP263 (2011) 1969. http://accelconf.web.cern.ch/AccelConf/PAC2011/papers/wep263.pdf

Participants

ANDRE, T.	Soleil Synchrotron, Saint Aubin, FR
BETTONI, T.	Psi, Villigen, CH
BIRKHAN, J.	Tu Darmstadt, Darmstadt, DE
BOEDEWADT, J.	Desy, Hamburg, DE
BORISSENKO, D.	Desy, Hamburg, DE
BRYNES, A.	Science and Technology Facilities Council Daresbury Laboratory, Warrington, GB
CELCER, T.	Psi, Villigen, CH
D'ALESSIO, A.	Tu Darmstadt, Darmstadt, DE
DINTER, H.	Desy, Hamburg, DE
DITTERT, D.	Tu Darmstadt, Darmstadt, DE
DOMONT-YANKULOVA, D.	Accelerator Technology Group(temf), Darmstadt, DE
DORDA, U.	Desy, Hamburg, DE
ENGLISCH, U.	Xfel Gmbh, Hamburg, DE
FABICH, A.	CERN, Geneva, CH
FANG, S.	Chinese Academy of Sciences, Beijing, CN
FELBER, M.	Desy, Hamburg, DE
GENG, Z.	Psi, Villigen, CH
GOBBO, A.	Psi, Villigen, CH
GRUEN, C.	Desy, Hamburg, DE
HEIL, P.	Johannes Gutenberg Universitaet, Mainz, DE
HEIN, L.	Johannes Gutenberg Universitaet, Mainz, DE
HELLERT, T.	Desy, Hamburg, DE
HERBERT, M.	Tu Darmstadt, Darmstadt, DE
HESSLER, C.	CERN, Geneva, CH
HUANG, X.	Institute of High Energy Physics, Beijing, CN
KAERTNER, F.	Desy, Hamburg, DE
KAMMERING, R.	Desy, Hamburg, DE
KARABEKYAN, S.	European Xfel Gmbh, Hamburg, DE
KINJO, R.	Riken, Sayo-gun, JP
KOSTIN, D.	Desy, Hamburg, DE
KRUGER, H.	Desy, Hamburg, DE
LANG, T.	Desy, Hamburg, DE
LAZZARINO, L.	University of Hamburg, Hamburg, DE
LECHNER, C.	Desy, Hamburg, DE
LEDROIT, B.	Johannes Gutenberg Universitaet, Mainz, DE
LI, P.	European Xfel, Hamburg, DE
LI, Y.	European Xfel Gmbh, Hamburg, DE
LIU, S.	Desy, Hamburg, DE
LOCKMANN, N.	Delta/Tu Dortmund, Dortmund, DE
LORBEER, B.	Desy, Hamburg, DE
MAAG, C.	Desy, Hamburg, DE
MAI, C.	Tu Dortmund University, Dortmund, DE
MUELLER, F.	Desy, Hamburg, DE
PANOFSKI, E.	Helmholtz-zentrum Berlin, Berlin, DE
PETRALIA, A.	Enea Frascati, Frascati, IT
PFINGSTNER, J.	University of Oslo, Oslo, NO
PFORR, J.	Tu Darmstadt, Darmstadt, DE
PLATH, T.	University of Hamburg, Hamburg, DE
RODRIGUEZ FERNANDEZ, A.	Psi, Villigen, CH

SAA-HERNANDDEZ, A.	Psi, Villigen, CH
SALEHI, E.	Amirkabir University of Technology, Tehran, IR
SANAYE HAJARI, S.	Ipm, Tehran, IR
SCHMEISSER, M.	Helmholtz-zentrum Berlin, Berlin, DE
SHAYEGANRAD, G.	Center For Synchrotron Radiation Technische Universitaet Dortmund, Dortmund, DE
SICHAU, J.	Center For Hybrid Nanostructures University, Hamburg, DE
SIMON, D.	Institut Fuer Kernphysik, Mainz, DE
STEINHORST, M.	Tu Darmstadt, Darmstadt, DE
STENGLER, T.	Johannes Gutenberg Universitaet Mainz, Mainz, DE
STOLL, C.	Johannes Gutenberg Universitaet Mainz, Mainz, DE
SULIMOV, A.	Desy, Hamburg, DE
TARKESHIAN, R.	Psi, Villigen, CH
USHAKOV, A.	Helmholtz-zentrum Berlin, Berlin, DE
WEI, T.	European Xfel Gmbh, Hamburg, DE
WINKELMANN, L.	Desy, Hamburg, DE
WISSMANN, J.	Institut Fuer Kernphysik, Darmstadt, DE
WOLFF-FABRIS, F.	European Xfel Gmbh, Hamburg, DE
YARMOHAMMADI, M.	Ipm, Tehran, IR
ZHAO, Q.	Pitz Desy, Zeuthen, DE
AHOU, C.	Center For Free-electron Laser Science, Hamburg, DE