CERN Yellow Reports: School Proceedings
Volume 3/2017

CERN-2017-006-SP

The CERN Accelerator School

Intensity Limitations in Particle Beams

Geneva, Switzerland
2–11 November 2015

Editor: W. Herr

CERN

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

ISBN 978–92–9083–462–5 (paperback)
ISBN 978–92–9083–463–2 (PDF)
ISSN 2519-8041 (Print)
ISSN 2519-805X (Online)
DOI https://doi.org/10.23730/CYRSP-2017-003

Accepted for publication by the CERN Report Editorial Board (CREB) on 24 July 2017

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

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

This report should be cited as:

Proceedings of the CAS–CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol.3/2017, CERN-2017-006-SP (CERN, Geneva, 2017). https://doi.org/10.23730/CYRSP-2017-003

A contribution in this volume should be cited as:

[Author name(s)], in Proceedings of the CAS–CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017), pp. [first page] – [lastpage], https://doi.org/10.23730/CYRSP-2017-003. [first page]

Abstract

This report presents the proceedings of a specialized course organized by the CERN Accelerator School (CAS). The topic this time being 'Intensity Limitations in Particle Beams'. The course was held at CERN, Switzerland, from 2-11 November 2015. The last course on this topic was organized in the framework of the Joint US-CERN School on Particle Accelerators in November 1990. It was felt that the progress in the field justified a revised course. The lectures addressed fundamental theory as well as experimental results. The indispensable beam diagnostics and mitigation measures were covered in dedicated lectures. The lectures were complemented by several sessions of exercises and the discussion of the solutions.


CERN
CERN

Preface

The aim of the CERN Accelerator School is to collect, preserve and disseminate the existing knowledge accumulated in the world's accelerator laboratories and universities. In addition to courses on general accelerator physics, specialized courses are organized to deepen the knowledge and increase technical competencies on specific topics in accelerator science. While most specialized courses treat related sub-systems and acccelerator technology, this course focuses on an advanced topic in beam dynamics. Intensity limitations do not only occur in High Energy Particle Accelerators but are of increasing importance in other types of accelerators. Accelerators for medical and industrial applications have an increasing need for high intensity and high quality beams.

The last course on this topic was organized in the framework of the Joint US-CERN School on Particle Accelerators in November 1990. It was felt that the rapid progress in this field motivated a revised course. Therefore the organization of such a course was fully supported by the CAS Advisory Committee.

This course was held at CERN, Switzerland from 2-11 November 2015 and its proceeding are compiled in the present volume. The backing of the CERN management and the provision of the necessary infrastructure have made this course possible.

The programme of the course was elaborated with the help of a dedicated Scientific Programme Committee, composed of experts in this field. They deserve our sincere thanks for their effort to ensure a good coverage of this very demanding topic. It was possible to attract world-reknowned experts as lecturers at this course and this ensured the high level of the presented material. The tremendous amount of work in preparing, presenting and writing-up of their topics for the present proceedings deserves the thanks of the organizers as well as the participants who came from all over the world to attend this course.

Finally, the quality of the contributions to these proceedings and the professional preparation will be highly appreciated by many people who will use the proceedings in the future.

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/37.

Werner Herr, Editor
CERN Accelerator School

Draft Programme
Intensity Limitations in Particle Beams, CERN, Geneva, Switzerland, 2-11 November, 2015

Time	Monday 2 November	Tuesday 3 November	Wednesday 4 November	Thursday 5 November	Friday 6 November	Saturday 7 November	Sunday 8 November	Monday 9 November	Tuesday 10 November	Wednesday 11 November
8:30		Opening Talks	Bench Measurements and Simulations of Beam Coupling Impedance	Beam Instabilities in Linear Machines II	Observations and Diagnostics in High Brightness Beams	Space Charge in Linacs		Electron Cloud I	High Brightness Photo Injectors	
9:30	A		U. Niedermayer	M. Ferrario	A. Cianchi	I. Hofmann		G. Rumolo	E. Chiadroni	DEPARTURE
9:30	R	Introduction and Needs for High Intensity and High Brightness	Beam Dynamics with High Intensity II	Beam-Beam Effects in Hadron Colliders I	Sources and and Low Energy Beam Transfer	Intrabeam Scattering	E	Beam-Beam Effects in Linear Colliders	Electron Cloud II	
10:30	R I V A L	L. Rivkin	A. Chao	T. Pieloni	R. Scrivens	M. Martini	X C U R S I O N	D. Schulte	G. Rumolo	
		COFFEE	COFFEE	COFFEE	COFFEE	COFFEE		COFFEE	COFFEE	
11:00		Overview of Limitations	Beam Based Impedance Measurements	Effects near Transition	Space Charge and Impedances	Space Charge in Circular Machines		Passive Mitigation	Active Mitigation	AFTER
12:00		W. Herr	E. Shaposhnikova	E. Metral	O. Boine-Frankenheim	G. Franchetti		V. Kornilov	H. Schmickler	
12:00	D	Wakefield and Impedances I	Beam Instabilities in Circular Machines II	Beam-Beam Effects in Hadron Colliders II	Numerical Methods I	Coherent Beam-Beam Effects		Machine Protection	Beam Loss Consequences	
13:00	A Y	M. Dohlus	A. Chao	T. Pieloni	K. Li	X. Buffat		R. Schmidt	F. Cerutti	BREAKFAST
		LUNCH	LUNCH	LUNCH	LUNCH	LUNCH		LUNCH	LUNCH	
14:30		Beam Dynamics with High Intensity I	Beam Instabilities in Linear Machines I	FREE	Study	CERN		Study	Ions	
15:30		A. Chao	M. Ferrario						R. Nagaoka	
		TEA	TEA		TEA			TEA	TEA	
16:00		Wakefield and Impedances II	Observations and Diagnostics in High Intensity Beams	AFTERNOON	Beam-Beam Effects in Circular Lepton Colliders	VISIT		Vacuum Issues	Numerical Methods II	
17:00	**Registration**	R. Wanzenberg	V. Kornilov		C. Milardi			P. Chiggiato	K. Li	
17:00		Beam Instabilities in Circular Machines I	Study		Tutorial			Tutorial	Seminar	
18:00		A. Chao							D. McGinnis	
		Welcome Drink								
19:00	DINNER	DINNER	DINNER	DINNER	DINNER	DINNER	Special Dinner	DINNER	DINNER	

Contents

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Overview—Intensity Limitations in Particle Accelerators

W. Herr
CERN, Geneva, Switzerland

Abstract
A brief and qualitative overview is given of the various intensity limitations that occur in particle accelerators. The aim is to make the participants aware of what is happening in this field in terms of observation, mitigation, and understanding. Such an overview cannot be rigorous nor complete. It serves as an introduction to the lectures on the different topics.

Keywords
Intensity limitations, instabilities, impedance, wakefields, beam-beam effects.

1 Introduction

1.1 Basic considerations

Many applications of accelerators require beams of high intensities [1]. High beam intensities imply large electromagnetic fields generated by the beams, by interactions within the beams, interactions with other beams, and interaction with beam surroundings. The result is a reaction back onto the fields that may lead to unstable beams. High intensities are therefore a strong source of instabilities and limitations and eventually limit the performance of accelerators [2–5]. Dedicated schools on these topics have been published [6–8].

High-intensity effects may affect the stability of single particles subjected to the generated or perturbed fields as well as the entire beam, leading to collective instabilities. Typical effects are direct (free) space charge effects, i.e., self-generated fields acting back on the beam. In a vacuum pipe, image charges are created, which act back on the beam. These space charge effects lead to coherent and incoherent tune shifts. The interactions with the environment are described by the impedance and the generated wakefields.

Another important aspect is that non-relativistic beams lead to collective effects.

1.2 Topics of this school

The limitations discussed in this school can be divided into different categories:

- incoherent single-particle and multiparticle effects;
- collective *coherent* effects and instabilities;
- collective *incoherent* effects;
- more than one beam, e.g., beam–beam effects, electron cloud, beam–gas interactions;
- practical considerations.

Related topics to be discussed are:

- diagnostics: high-intensity and high-brightness diagnostics;
- limitations in low- and high-energy beams;
- passive mitigation (e.g., Landau damping);
- active mitigation (e.g., feedback systems);
- numerical and simulation tools.

https://doi.org/10.23730/CYRSP-2017-003.1

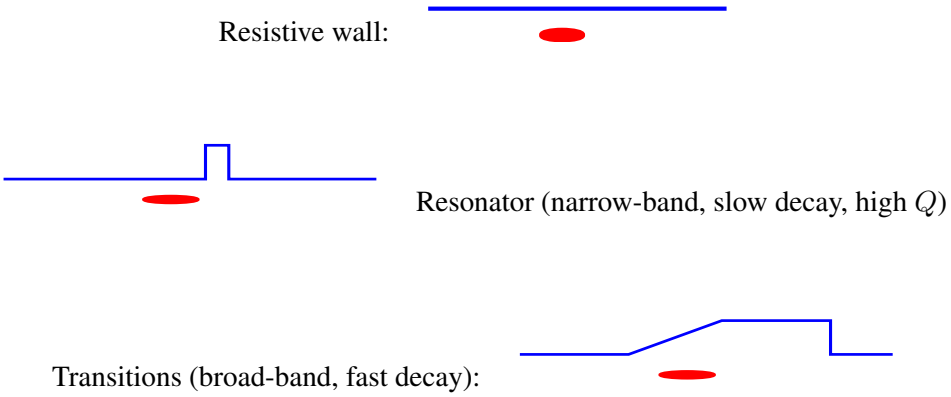


Fig. 1: Different sources of impedance

2 Impedances and wakefields

Wakefields are generated by moving charges and affect other charges in the beam [9–12]. A smooth, perfectly conducting wall will not generate wakefields. Wakefields can have different origins, the most relevant being:

- resistive wall;
- discontinuities.

Depending on the type of discontinuity, the frequency content of the wakefields can be very different: a resonator with a high Q value generates a narrow frequency band and decays over a longer time scale. A vacuum chamber with different structures induces wakefields with a rather broad frequency spectrum, also known as broad-band wakefields. The impedances are the Fourier transforms of the wakefields and are a direct measurement of the frequency content. A sketch is shown in Fig. 1.

The wakefields are stronger as the intensities increase and will eventually lead to instabilities and limit the obtainable intensities. Of large practical importance is the Panofsky–Wenzel theorem, which relates the longitudinal and transverse impedances.

2.1 Resistive wall

In the case of a resistive wall, a charge will leave a wakefield behind it. No field exists in front of the charge. Wakefields due to resistive walls can have a slow decay, affecting trailing particles; in some cases the wakefield decay is sufficiently slow to perturb the motion of the charge after a full turn in the machine (in circular accelerators), i.e., the charge is affected by its own wakefield. Such short- and long-range wakefields, therefore, cover a large frequency spectrum of the corresponding impedance.

2.2 Discontinuities

In general, abrupt changes of the wall boundaries generate rather complex wakefields and impedances. For smooth transitions, a simplified treatment can be applied. Figures 2 and 3 show the real and imaginary parts of the impedance for resonators with different Q values. The exact wakefields (and impedances) for complex structures are rather difficult to compute and numerical methods are required [13]. Beam-based measurements are a rather reliable method [14].

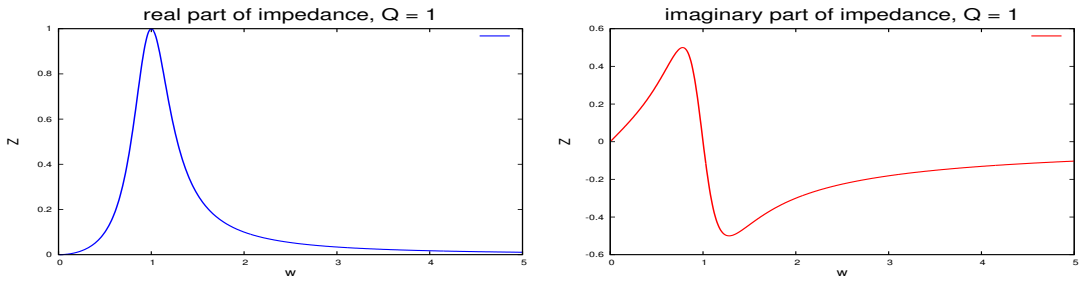


Fig. 2: Impedances: real and imaginary part of a narrow-band resonator with a rather low quality factor Q

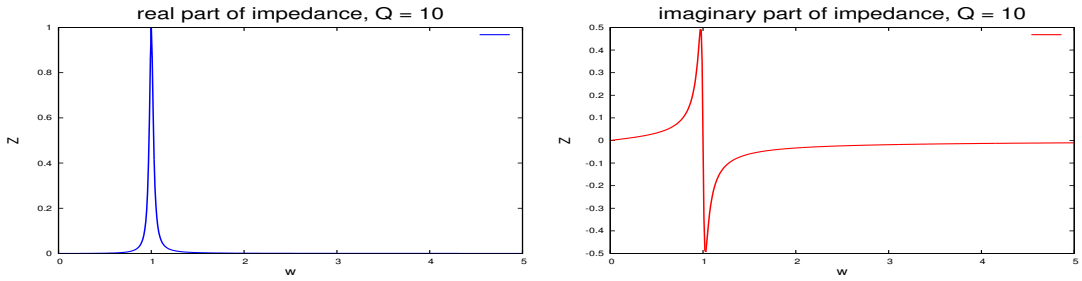


Fig. 3: Impedances: real and imaginary part of a resonator with a high quality factor Q

3 Single-particle effects

The fields generated by high-intensity beams perturb the electromagnetic fields in the surroundings of the beam and can lead to single-particle effects, such as reduced dynamic aperture, emittance growth, and particle loss.

3.1 Non-linear effects and resonances

Space charge effects are a significant source of non-linearities [15, 16]. They lead to strong detuning with amplitudes and particles are subjected to resonance effects. Above a threshold of the space charge effects, the losses may become unacceptable. Beam–beam effects in colliders are another important source of non-linear effects caused by high-intensity beams [17–21].

3.2 Touschek effects and intrabeam scattering

Touschek effects and intrabeam scattering are typical single-particle effects in high-intensity beams and lead to particle loss and emittance growth [22]. The physical origin of these effects is collisions between particles. The collision strengths are directly related to intensity and beam sizes. Collisions where the momentum transfer is large can lead to the instantaneous loss of a particle (Touschek effect). Smaller momentum transfer typically results in emittance growth. An example of the emittance growth as a function of different intensities and different beam sizes is shown in Fig. 4. The observed dependencies can be easily seen.

4 Low-energy effects and transitions

An important concern in hadron accelerators during acceleration is passing through transition [23, 24] at the transition energy. The main effects at low energy are space charge effects and the crossing of transition. The main objective is to maintain the longitudinal emittance during the transition crossing. The main effects at the transition energy are:

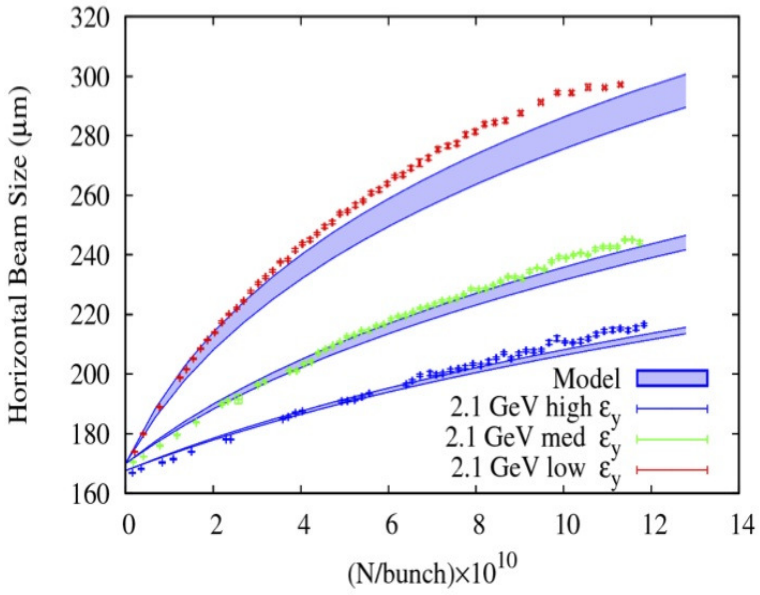


Fig. 4: Horizontal emittance growth due to intrabeam scattering for different beam parameters ([25])

– change of stable phase;
– relative momentum spread $\Delta p/p$ goes to infinity;
– bucket area gets large;
– bunch length gets short;
– Synchrotron frequency goes to zero.

These consequences are not negligible when high-intensity beams pass through transition.

A resistive impedance can cause a microwave instability near transition crossing. If the transition is not crossed fast enough, other slowly growing instabilities may be excited.

5 Linear accelerators

5.1 Space charge in linacs

Particles, in particular hadrons in linacs, are often non-relativistic and high intensities lead to a very significant space charge [15]. An important consequence is emittance growth along the line.

5.2 Energy spread in linacs

Wakefields generated by a charge in a linac produce longitudinal forces on the particles. This leads to an energy change. Not all particles lose the same energy; this uneven energy loss results in an energy spread. In linear colliders, this can become a problem for final focusing.

5.3 Parasitic heating

By interacting with the longitudinal impedance, the charged particles lose energy. This parasitic loss leads to a heating of the vacuum chamber. A large fraction is lost in the presence of sharp discontinuities, e.g., cavities. Trapped wakefields are often the source of instabilities [26].

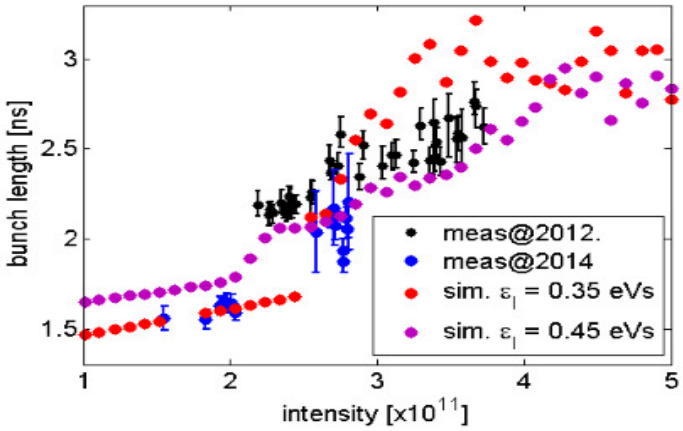


Fig. 5: Bunch lengthening observed in the SPS

5.4 Instabilities in linacs

As a typical example of an instability in linacs, we can mention the beam break-up. A wakefield generated by the head particles of a bunch can cause deflections of the tail; for large wakefields, i.e., high intensities, this can lead to an instability.

6 Longitudinal effects—circular accelerators

6.1 Longitudinal space charge forces

Longitudinal space charge forces can be very large at low energies and introduce longitudinal defocusing below the transition energy. This can limit the acceptance and a careful adjustment of the RF voltage may be necessary.

6.2 Longitudinal effects in coasting beams

6.2.1 Longitudinal microwave instability

In unbunched beams, very short-range wakefields can lead to a longitudinal density modulation. This is often accompanied by high-frequency signals and is therefore termed microwave instability. A criterion linking machine impedance with the maximum intensity is given by the Keil–Schnell criterion [27].

6.3 Longitudinal effects in bunched beams

6.3.1 Longitudinal microwave instability

Increased bunch intensities can lead to longitudinal instabilities of single bunches. A key criterion to characterize this instability is the peak intensity in the bunches [28]. In beams with bunches, we must distinguish between short- and long-range wakefields, or, equivalently, with high- and low-frequency components of the impedance. Typical effects are, e.g., energy loss and bunch lengthening. The bunch lengthening observed in the SPS is shown as an example in Fig. 5. On increasing the intensity, the measured bunch length is increased.

6.3.2 Longitudinal instability

Longitudinal instabilities may be single or coupled bunch effects [2, 29]. Possible different single bunch modes from a simulation are shown in Fig. 6. An observation [30] of the beam profile using a wall

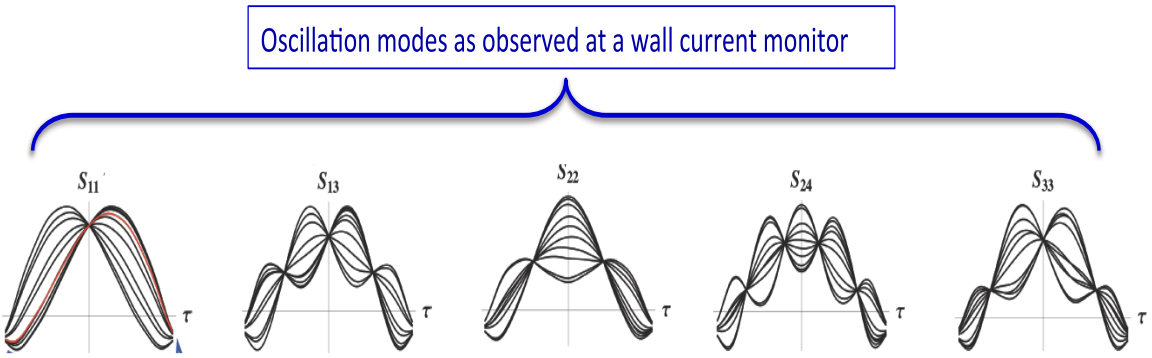


Fig. 6: Longitudinal oscillation modes, simulated. Horizontal axis time τ

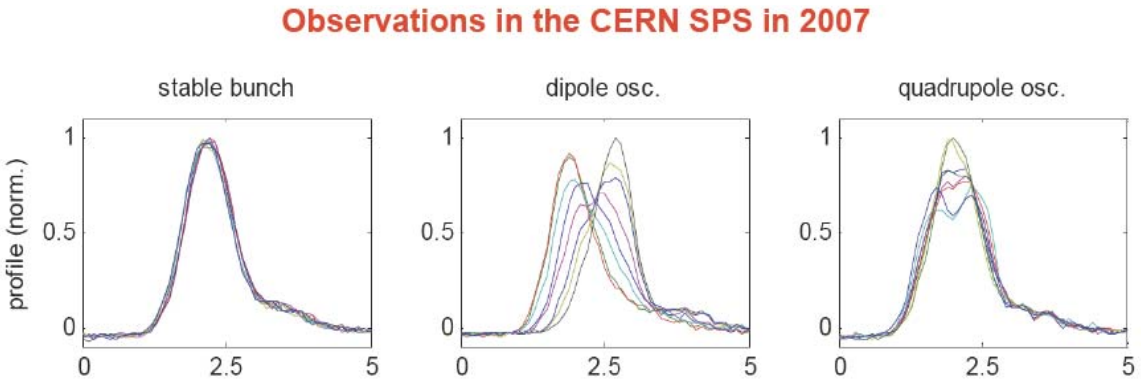


Fig. 7: Profile of longitudinal oscillation modes, observed in the SPS

current monitor in the SPS is shown in Fig. 7, which shows very clearly the different modes of stable beams and the basic modes of oscillation.

A large number of bunches can result in a large number of oscillation modes. For quantitative treatment, the bunches are usually considered as rigid objects [29].

6.3.3 Robinson instability

A longitudinal instability that can occur in circular accelerators is the Robinson instability. The fundamental frequency of cavities is tuned to values of the revolution frequency ω_0 multiplied by the harmonic number. The cavities are a source of impedance; for a value of slightly above or below ω_0 (depending on whether one works above or below the transition), the motion is stable or unstable. Since the quality factor of a cavity is usually very large, fine tuning is required to avoid this instability. The Robinson instability is often considered one of the most fundamental forms of instability.

6.3.4 Potential well distortion

Impedances at high frequencies, i.e., corresponding to short-range wakefields, have an effect on longitudinal focusing. This change in longitudinal focusing is usually called potential-well distortion. It has a strong impact on the charge distribution within a bunch.

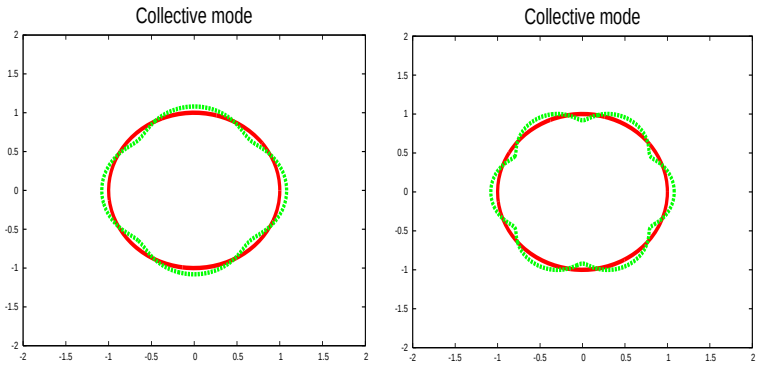


Fig. 8: Transverse collective modes in unbunched beams

7 Transverse effects—circular accelerators

7.1 Transverse space charge forces

Direct space charge effects are the result of the interaction of particles within a beam. Transverse space charge forces are the dominant effect in low-energy hadron machines. Space charge forces are responsible for both coherent and incoherent effects and introduce incoherent and coherent tune shifts. The shifts are always defocusing. By nature, the effects are strongly non-linear and excite resonances or losses driven by different mechanisms. In addition, one must expect an interplay between the space charge effects and machine non-linearities. This makes it difficult to predict the consequences.

7.2 Transverse effects in coasting beams

7.2.1 Transverse microwave instability

Unbunched beams can execute collective modes with different mode indices, depending on the pattern. Examples of two different modes are shown in Fig. 8. One has to distinguish between *fast waves* and *slow waves*, related to orbital harmonics. The beams exhibit a different behaviour concerning possible instabilities; above the intensity threshold they can become unstable. At a fixed observation point, the beams oscillate at high frequencies. This type of instability is usually called 'transverse microwave instability in unbunched (coasting) beams'.

7.3 Transverse effects in bunched beams

7.3.1 Transverse microwave instability

Originally derived and used for unbunched beams, this instability also occurs in bunched beams. The intensity limit is again determined by the peak density of the bunched beam. Typically, the transverse microwave instability is accompanied by a fast increase of the transverse emittance.

7.3.2 Coupled bunch instability

In a beam with a large number of bunches, wakefields can excite coupled bunch oscillations. Here, the wakefields are the sum of those generated by all bunches in the ring. This instability can exhibit a rather large number of different modes with a very different pattern. In general, for a beam with N bunches, one can have N possible modes of oscillation. The growth rate of certain modes depends on the intensity and may become a limitation.

Longitudinal coupled bunch instabilities also exist, but are much weaker, in general.

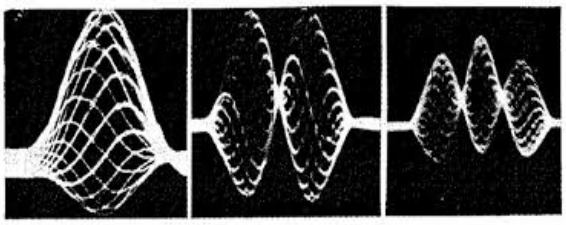


Fig. 9: Head–tail modes, observed in CERN-PS, [31]

Fig. 10: Head–tail mode $m = 1$, observed in the Large Electron–Positron Collider. Screen shot with streak camera

7.3.3 Head–tail instability

This instability is a single bunch instability. Short-range transverse wakefields from particles at the 'head' of a bunch can excite oscillations at its 'tail'. The synchrotron motion exchanges the particles and the new head particles continue to excite the particles behind. The motion becomes unstable if the oscillation grows, depending on the chromaticity and whether the instability is not suppressed by Landau damping.

Head–tail modes as observed with a longitudinal profile monitor in the CERN-PS are shown in Fig. 9. A nice demonstration is shown in Fig. 10, where the head–tail mode $m = 1$ is shown turn by turn, as measured with a streak camera.

7.3.4 Transverse mode coupling instability

This instability, also known as 'fast head–tail instability', appears when two neighbouring modes approach each other as a result of the frequency detuning with increasing bunch intensity. This instability shows a very prominent threshold behaviour. This behaviour was studied with a simulation program; Fig. 11 shows the results. The tune change and the merging of modes as the intensity is increased can be clearly seen. An instability develops at the intensity where the modes have merged.

8 Beam–beam effects

While space charge effects are strongly suppressed for ultrarelativistic beams, the interaction between colliding beams does not vanish. It is by far the strongest source of non-linearities in high-energy particle colliders [7, 8, 17, 18]. This strong non-linearity makes it very difficult to predict the exact beam behaviour; in particular, it is a very complex problem when there are many bunches and collisions. While synchrotron damping helps significantly in avoiding detrimental effects in lepton colliders, such damping does not exist in hadron colliders and the underlying mechanisms and problems are very different.

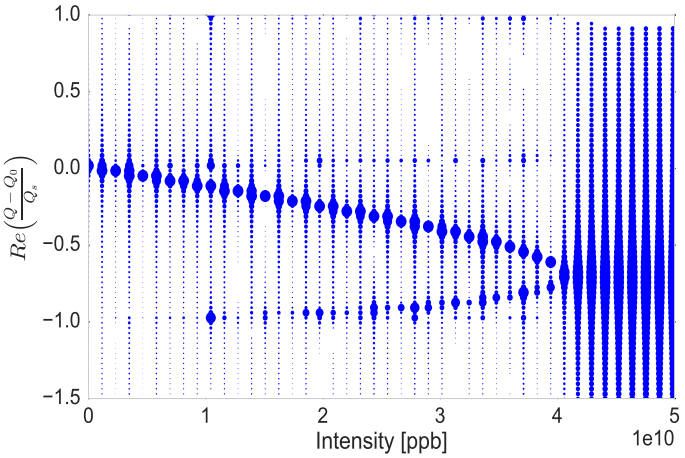


Fig. 11: Transverse mode coupling, simulation

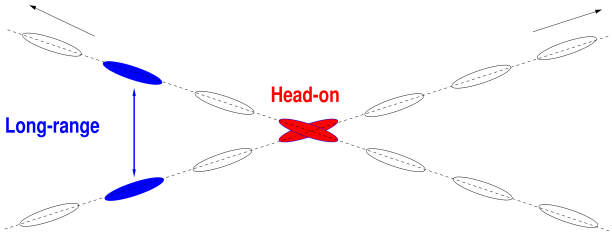


Fig. 12: Beam crossing in the Large Hadron Collider. Head-on (HO) and long range (LR) collisions are indicated

8.1 Hadron beams

In hadron colliders with many bunches, crossing angles (Fig. 12) are required to separate the beams at unwanted collision points and lead to the further complication of long-range interactions [18].

Beam losses due to long-range beam–beam effects as a function of crossing angle are shown in Fig. 13. The crossing angle was reduced in steps from 142 μrad to 72μrad and the lines correspond to bunches with different numbers of long range (LR) interactions. A sufficiently large crossing angle is necessary to guarantee small losses, but leads to further complications [18]. The losses depend strongly on the number of long range interactions.

Other effects are coherent beam–beam motions, which can lead to very fast beam loss within a few turns [20].

8.2 Lepton beams

Colliding lepton beams have a strong interplay with radiation damping and can accept significantly stronger beam–beam effects [19]. However, they show a very distinct threshold behaviour (Fig. 14). Above the threshold (beam–beam limit), the beam–beam tune shift remains constant when the intensities are increased. This is related to an increase in the vertical emittance caused by the beam–beam effects, which are responsible for a strong coupling to the horizontal motion. The interplay between excitation by beam–beam effects and damping leads to an equilibrium and a constant beam–beam tune shift.

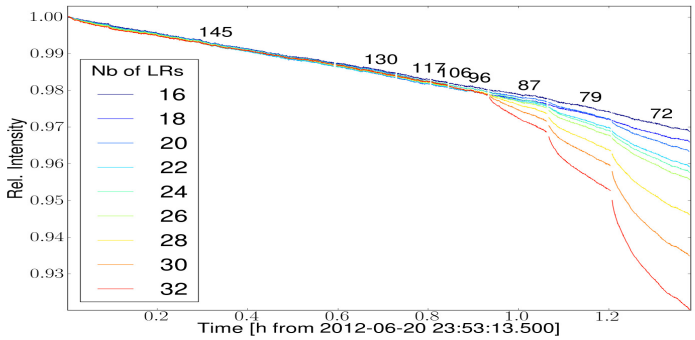


Fig. 13: Beam losses due to long-range (LR) beam–beam effects as a function of crossing angle. The relative intensity is shown as a function of time during the experiment. The corresponding crossing angles decending from 145 μrad to 72 μrad are indicated in the figure. The lines correspond to the bunches with a diffeent number of long range (LR) interactions.

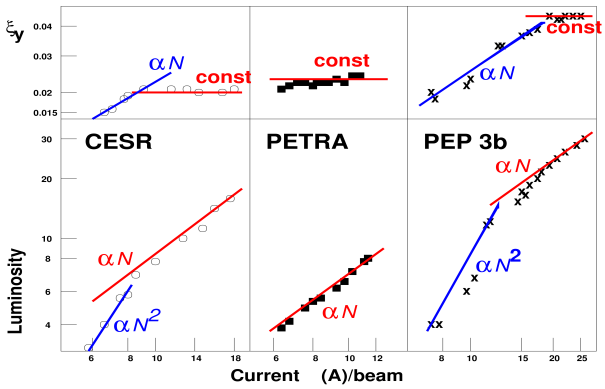


Fig. 14: Beam–beam limit as function of intensity for different lepton collides

8.3 Coherent beam–beam effects

Under the right conditions, coherent beam–beam effects can build up and lead to a very fast beam loss [20]. With a large number of bunches, such as those used at the Large Hadron Collider, many different coherent modes must be expected, some of which are potentially unstable. A detailed analysis of the mechanisms of the excitation of coherent effects and possible mitigation techniques should form part of the design of accelerators with high-intensity beams.

8.4 Linear colliders

Although linear colliders are single-pass, beam–beam effects play an important role in linear colliders [21]. Beam–beam-induced pinching effects are used to enhance the luminosity and control of the disruption is necessary. Furthermore, bremsstrahlung from the collision has a strong impact on machine performance in terms of the useful luminosity, i.e., beam-induced background.

9 Mitigation

To suppress collective instabilities, different mitigation mechanisms are available, passive as well as active.

9.1 Passive mitigation

The most prominent and most efficient mitigation effect is an effect called Landau damping. Strictly speaking, it is not dissipative damping, which would lead to emittance growth and should be distinguished from decoherence (a very common confusion). The computation of the beam stability is rather reliable and is used in most hadron machines [32, 33].

9.2 Active mitigation

Active mitigation can be achieved with different types of feedback systems [34].

10 Electron cloud

10.1 Origin and consequences

Synchrotron radiation from high-energy beams hits the wall of the vacuum chamber and creates photoelectrons [35]. These are affected, i.e., accelerated, by the following bunches and can hit the wall with higher energies, liberating more electrons, as a consequence leading to a build-up of a localized cloud of electrons. The intensity as well as the distance between bunches is an important parameter. Most important is the number of liberated secondary electrons. Consequences of the electron cloud are the coupling of successive bunches, potentially leading to coherent (coupled bunch) instabilities. Another consequence is the deterioration of the vacuum pressure.

10.2 Mitigation

Possible mitigation techniques are a proper design of the vacuum chamber to reduce the effect of the synchrotron radiation. Choice of the proper bunch distance and reduction of the secondary electron yield are very efficient techniques.

11 Ions

Similar considerations hold for ions, degrading the vacuum or ions trapped in the beams [36]. Beam instabilities for high-intensity and low-emittance beams must be assessed.

12 Numerical and simulation tools

Numerical models are essential for the design and operation of an accelerator [37]. Conditions that cannot be studied in existing machines must be simulated. Furthermore, simulation enables one to disentangle the different processes leading to observed effects and provides an excellent analysis tool.

Other typical applications are the computation of impedances and wakefields using, e.g., particle-in-cell codes and particle tracking. The implementation of collective effects via macroparticle and multi-particle simulations is now possible with the available computing resources and relatively easy to implement.

13 Observations and diagnostics

Measurement and diagnosis of beam parameters and the beam quality are essential for the operation and performance of the machine. At high brightness and intensity, it is often not possible to use intercepting devices for the measurements.

13.1 High-intensity beams

Observation and diagnosis of high-intensity beams are vital for the control and protection of the machines [38]. High-intensity beams deliver a stronger signal and better signal-to-noise ratios but signals may be

distorted. A typical example is the measurement of the beam profile. In that case, additional effort is needed to obtain a reliable measurement. Care must be taken during the analysis of the measurements; collective effects, such as coherent and incoherent tune shifts, self-fields, and wakes, must be taken into account. Measurement at low intensities is an option, but not applicable in all cases. Some effects are visible only when the intensity is high.

13.2 High-brightness beams

Apart from the intensity, the most important parameter for a high-brightness beam is the emittance [39]. At lower energies, measurement may be done in a regime dominated by space charge. A second challenge is the measurement of the very short bunch length. New ideas are being studied and tested.

14 Sources and injectors

High-brightness beams require a high brightness at the source and its conservation in 6D [40]. The optics of the beam transport and the acceleration must ensure this conservation. An essential part of the chain is the section for beam conditioning, typically before the accelerating section. A limit to the intensity is imposed by the space charge forces leading to emittance growth, even for electrons. For an overview of ion sources see Ref. [41].

15 Practical considerations

The beam intensity may also be limited because of other problems, e.g., engineering or operational difficulties, such as:

- machine protection;
- vacuum;
- cryogenics.

15.1 Machine protection

High-intensity and high-power beams require a sophisticated system to protect the machine against damage. This has become increasingly important for the high-energy machines being studied nowadays. Possible failure scenarios must be considered during machine design at a very early stage [42].

The safety of personnel is of vital importance and is included in the considerations.

15.2 Vacuum

Beyond other problems, beam-induced absorption and multipacting can limit the maximum allowed intensity. A proper design of the vacuum chamber and mitigation techniques, such as conditioning, are required. A particular case mentioned already is the build-up of an electron cloud and the consequences for the vacuum and the beam dynamics [43].

15.3 Beam loss consequences

Beam losses not only affect the performance of the machine, but are significant for such events as [44]:

- failure of electronics, both single events and cumulative effects;
- activation of material and air;
- chemical reactions under the influence of ionizing particles.

References

[1] L. Rivkin, Needs for high intensity and high brightness, these proceedings.

[2] A. Chao, *Physics of Collective Beam Instabilities in high Energy Accelerators* (Wiley, New York, 1993).

[3] A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1998).

[4] A. Chao, Beam dynamics with high intensity, these proceedings.

[5] A. Chao, Instabilities in circular accelerators, these proceedings.

[6] M. Dienes *et al.* (eds), *Frontiers of Particle Beams: Intensity Limitations* (Springer, Berlin, 1992), http://dx.doi.org/10.1007/3-540-55250-2

[7] Proceedings of the CERN Accelerator School: High Power Hadron Machines, 24 May–2 June 2011, Bilbao, Spain, edited by R. Bailey, CERN-2013-001 (CERN, Geneva, 2013), http://dx.doi.org/10.5170/CERN-2013-001

[8] Proceedings of the CERN Accelerator School: Advanced Accelerator Physics, 18–29 August 2013, Trondheim, Norway, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), http://dx.doi.org/10.5170/CERN-2014-009

[9] B. Zotter and S. Kheifets, *Impedances and Wakefields in High Energy Particle Accelerators* (World Scientific, Singapore, 1997).

[10] R. Wanzenberg, Impedances and wakefields I, these proceedings.

[11] M. Dohlus, Impedances and wakefields II, these proceedings.

[12] O. Boine-Frankenheim, Space charge and impedances, these proceedings.

[13] U. Niedermayer, Simulation of beam coupling impedance, these proceedings.

[14] E. Shaposhnikova, Beam based impedance measurements, these proceedings.

[15] I. Hofmann, Space charge in linear machines, these proceedings.

[16] G. Franchetti, Space charge in circular machines, these proceedings.

[17] Proceedings of the ICFA Mini-Workshop on Beam-beam Effects in Hadron Colliders, CERN, Geneva, Switzerland, 18–22 April 2013, edited by W. Herr and G. Papotti, CERN-2014-004 (CERN, Geneva, 2014), http://dx.doi.org/10.5170/CERN-2014-004.

[18] T. Pieloni, Beam–beam effects in hadron colliders, these proceedings.

[19] C. Milardi, Beam–beam effects in lepton colliders, these proceedings.

[20] X. Buffat, Coherent beam–beam effects, these proceedings.

[21] D. Schulte, Beam–beam effects in linear colliders, these proceedings.

[22] M. Martini, Intrabeam scattering, these proceedings.

[23] S.D. Holmes, in *Frontiers of Particle Beams: Intensity Limitations*, Eds. M. Dienes *et al.* (Springer, Berlin, 1992), p. 189, http://dx.doi.org/10.1007/3-540-55250-2_31

[24] E. Metral, Effects at transition, these proceedings.

[25] Cornell University, Internal report, unpublished.

[26] M. Ferrario, in Proc. of the CERN Accelerator School: Advanced Accelerator Physics, 18–29 August 2013, Trondheim, Norway, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp.357-375, http://dx.doi.org/10.5170/CERN-2014-009.357

[27] E. Keil and W. Schnell, Concerning longitudinal instability in the ISR, Technical report CERN-ISR-TH-RF/69-48 (1969).

[28] D. Boussard, Observation of the longitudinal microwave instability in the CERN-PS, Technical report LABII/RF/Int./75-2 (1975).

[29] G. Rumolo, in Proc. of the CERN Accelerator School: Advanced Accelerator Physics, 18–29 August 2013, Trondheim, Norway, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014),

pp.199-219, http://dx.doi.org/10.5170/CERN-2014-009.199

[30] G. Papotti *et al.*, The SPS Beam Quality Monitor, from Design to Operation, Proc. International Particle Accelerator Conference (San Sebastian, Spain, 2011).

[31] J. Gareyte and F. Sacherer, Head-Tail type instabilities in the CERN PS and Booster, Proc. International conference on High Energy Accelerators, Stanford, 1977.

[32] W. Herr, in Proc. of the CERN Accelerator School: Advanced Accelerator Physics, 18–29 August 2013, Trondheim, Norway, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp.377-404, http://dx.doi.org/10.5170/CERN-2014-009.377

[33] V. Kornilov, Passive mitigation, these proceedings.

[34] H. Schmickler, in Proeedings of the CERN Accelerator School: Advanced Accelerator Physics, 18–29 August 2013, Trondheim, Norway, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp.503-546, http://dx.doi.org/10.5170/CERN-2014-009.503

[35] G. Rumolo, Electron cloud, these proceedings.

[36] R. Nagaoka, Ions, these proceedings.

[37] K. Li, Numerical methods, these proceedings.

[38] V. Kornilov, Observations and diagnostics in high intensity beams, these proceedings.

[39] A. Cianchi, Observations and diagnostics in high brightness beams, these proceedings.

[40] E. Chiadroni, High brightness photo injectors, these proceedings.

[41] Proceedings of the CERN Accelerator School: Ion Sources, Senec, Slovakia, 29 May–8 June 2012, edited by R. Bailey, CERN-2013-007 (CERN, Geneva, 2013), http://dx.doi.org/10.5170/CERN-2013-007

[42] R. Schmidt, in Proceedings of the CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 18–29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp.221-243, http://dx.doi.org/10.5170/CERN-2014-009.221

[43] P. Chiggiato, Vacuum issues, these proceedings.

[44] F. Cerutti, Beam loss consequences, these proceedings.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

An Introduction to Wake Fields and Impedances

M. Dohlus and R. Wanzenberg
Deutsches Elektronen-Synchrotron DESY, D-22603 Hamburg, Germany

Abstract
The concepts of wake fields and impedance are introduced to describe the electromagnetic interaction of a bunch of charged particles with its environment in an particle accelerator. The various components of the environment are the vacuum chamber, cavities, bellows, dielectric-coated pipes, and other kinds of obstacles that the beam has to pass on its way through the accelerator. The wake fields can act back on the beam and lead to instabilities, which may limit the achievable current per bunch, the total current, or even both. Some typical examples are used to illustrate the wake function and its basic properties. Then wake fields in cavities and resonant structures are studied in detail. Finally, the frequency-domain view of the wake field or impedance is explained, and basic properties of the impedance are derived.

Keywords
Wake field; impedance; modes in a cavity.

1 Introduction

A beam in an accelerator interacts with its vacuum chamber surroundings via electromagnetic fields. In this lecture the concept of wake fields is introduced to describe the electromagnetic interaction of a bunch of charged particles with its environment. The various components of the environment are the vacuum chamber, cavities, bellows, dielectric-coated pipes, and other kinds of obstacles the beam has to pass on its way through the accelerator. The wake fields can act back on the beam and lead to instabilities, which may limit the achievable current per bunch, the total current, or even both.

This lecture builds upon a previous lecture on wake fields and impedance given by T. Weiland about 25 years ago [1]. We recommend that the reader also consult the excellent textbooks [2–4] which cover the subject matter of this lecture.

We start with some typical examples from accelerator physics in which wake field effects are important.

Then, in Section 2, the concept of wake potential is formally introduced and multipole expansions are studied for structures with cylindrical symmetry. The Panofsky–Wenzel theorem, which links the longitudinal and transverse wake forces, is explained.

Section 3 is devoted to the analysis of wake fields due to resonant modes in a cavity. The fundamental theorem of beam loading is explained in detail. Finally, analytical results for a pillbox cavity are presented. The coupling of the beam to one mode of a cavity leads to the concept of the loss parameter.

In Section 4, the impedance is introduced as the Fourier transform of the wake potential. The properties of the wake functions (time-domain view) are translated to properties of the impedance (frequency-domain view).

1.1 Basic concepts

Consider a point charge q moving in free space at a velocity v close to the speed of light. The electromagnetic field is Lorentz-contracted into a thin disk perpendicular to the particle's direction of motion [5], which we choose to be the z-axis in a cylindrical coordinate system. The opening angle of the field

https://doi.org/10.23730/CYRSP-2017-003.15

distribution is of the order of $1/\gamma$, where $\gamma = \left(1 - (v/c)^2\right)^{-1/2}$. The field distribution is shown in Fig. 1. Even for an electron beam with an energy of $E = 10\,\mathrm{MeV}$, the opening angle ϕ is no greater than $50\,\mathrm{mrad}$ or $2.89°$:

$$\phi = \frac{1}{\gamma} = \frac{0.511\,\mathrm{MeV}}{E} = 2.89°$$

($m_0 c^2 = 0.511\,\mathrm{MeV}$ is the rest mass of the electron).

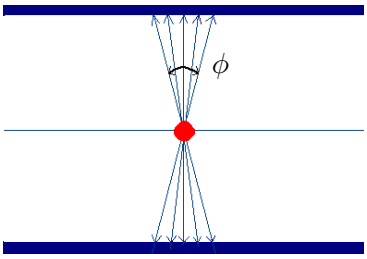


Fig. 1: Electromagnetic field carried by a relativistic point charge q

In the ultra-relativistic limit $v \to c$ (or $\gamma \to \infty$), the disk containing the field shrinks to a δ-function distribution. The non-vanishing field components are

$$E_r = \frac{q}{2\pi\,\varepsilon_0 r}\,\delta(z - c\,t), \qquad H_\varphi = \frac{E_r}{Z_0} \quad \text{with } Z_0 = 377\,\Omega.$$

Since the electric field $\boldsymbol{E}$ points strictly radially outward from the point charge, all field components are identically zero both ahead of and behind the point charge, and hence there are no forces on a test particle either preceding or following the charge q.

For v slightly less than c, this is not strictly true. However, if we look at some typical bunch charges and energies of high-energy accelerators and synchrotron light sources, as shown in Table 1, we will notice that the space charge force $V_\mathrm{s} = e/(4\,\pi\,\epsilon_0\,d^2\,\gamma^2)$ (where d is the rms distance between two electrons in the bunch) scales with $1/\gamma^2$. It is then obvious that as a good approximation, any space charge effects can be neglected for the accelerators under consideration. Nevertheless, space charge effects are important in heavy ion or low-energy proton accelerators.

Table 1: Typical bunch charges and energies of high-energy accelerators and synchrotron light sources [6–8]

Machine	Charge (nC)	Energy (GeV)	$\gamma = \left(1 - (v/c)^2\right)^{-1/2}$
LHC	20	7000	7 500
LEP	100	60	195 700
PETRA III	20	6	11 700

In the next section we will restrict ourselves to the ultra-relativistic case ($\gamma = \infty$, $v = c$), so space charge effects are neglected.

1.2 Some simple examples

Consider some typical settings where electromagnetic fields occur behind a bunch with charge q moving with velocity c through a structure. A bunch moves through a cylindrical pipe along the z-axis. All electric field lines terminate transversely on surface charges on the wall of the pipe, assuming a perfectly

conducting wall. There will be no wake fields behind the charge. The situation is different, however, if the cross-section of the beam pipe changes. A step-out transition is shown in Fig. 2. All fields have been calculated using a numerical wake field solver from MAFIA or the CST studio suite [9–12]. Here

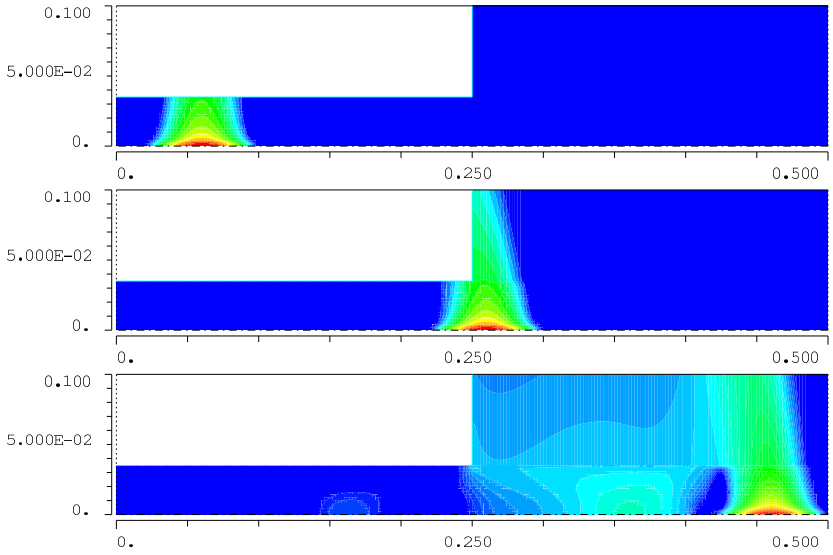


Fig. 2: Wake fields behind a bunch generated at a step-out transition from a small to a larger beam pipe

we have assumed that all pipe walls are perfect conductors. The wake field is generated because of the change in geometry. It should be noted that any beam pipe with finite conductivity, as well as flat beam pipes, can generate wake fields (resistive wall wake fields) [13]. Furthermore, a dielectric-coated pipe, which could be used as a travelling-wave acceleration section, will generate wake fields; see, for example, [14].

Another example is a cavity in a beam pipe; see Fig. 3. Again, a bunch is moving through a cylindrical pipe along the z-axis. Wake fields are generated because of geometrical changes in the pipe cross-section. In this respect the situation is similar to the previously considered case of a step-out transition. The main difference is that the bunch can excite modes in the cavity and therefore long-range wake fields, which can ring for a long time in the cavity (depending on the conductivity of the cavity wall).

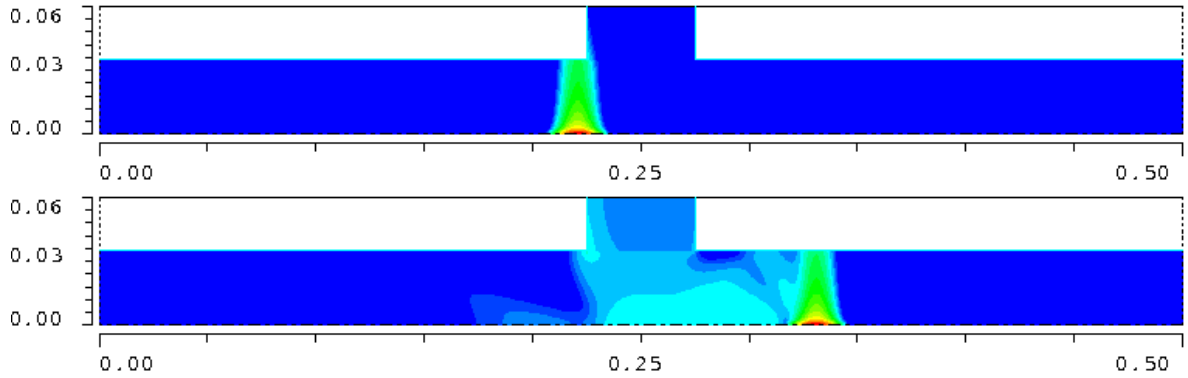


Fig. 3: Wake fields in a cavity

From these basic considerations we have learned that for electron accelerators the dominant wake forces are caused by geometrical changes along the beam pipe. Space charge effects are negligible for

ultra-relativistic particles. Wake fields due to the resistive wall or dielectric coatings should always be checked in detail according to the specific situation.

2 Wake fields

2.1 Wake fields in a resonant cavity with beam pipes

The examples above give us a qualitative understanding of wake fields and how they are generated. Before proceeding to mathematical descriptions in terms of wake potentials, let us take a closer look at the example considered in Section 1.

An ultra-relativistic point particle with charge q_1 traverses a small cavity parallel to the z-axis, with offset (x_1, y_1); see Fig. 4. The electromagnetic force on any test charge q_2 is given as a function of

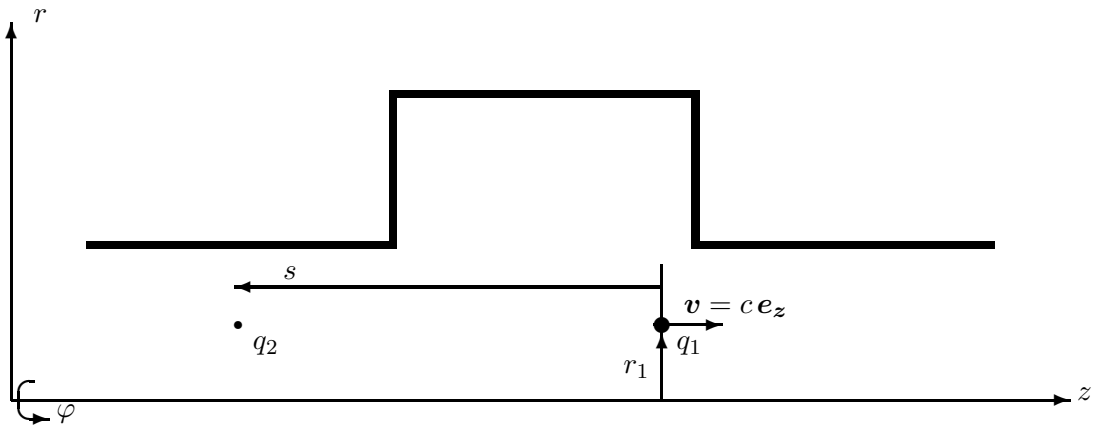


Fig. 4: An ultra-relativistic point particle with charge q_1 traverses a small cavity parallel to the z-axis, followed by a test charge q_2

space and time coordinates by the Lorentz equation

$$\boldsymbol{F}(\boldsymbol{r},t) = q_2\big(\boldsymbol{E}(\boldsymbol{r},t) + \boldsymbol{v}\times\boldsymbol{B}(\boldsymbol{r},t)\big),$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ are the fields generated by q_1; they are solutions of the Maxwell equations

$$\boldsymbol{\nabla}\times\boldsymbol{B} = \mu_0\boldsymbol{j} + \frac{1}{c^2}\frac{\partial}{\partial t}\boldsymbol{E}, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0,$$
$$\boldsymbol{\nabla}\times\boldsymbol{E} = -\frac{\partial}{\partial t}\boldsymbol{B}, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{E} = \frac{1}{\epsilon_0}\rho$$

and have to satisfy several boundary conditions.

In our case the charge and current distributions are

$$\rho(\boldsymbol{r},t) = q_1\,\delta(x-x_1)\delta(y-y_1)\delta(z-ct),$$
$$\boldsymbol{j}(\boldsymbol{r},t) = c\,\boldsymbol{e_z}\,\rho(\boldsymbol{r},t).$$

After interaction of q_1 with the cavity, there remain electromagnetic fields in the cavity. The source particle has lost energy to cavity modes and excited fields that propagate in the semi-infinite beam pipes.

Now consider a test charge q_2 following q_1 at a distance s with the same velocity $v \approx c$ and with offset (x_2, y_2). The Lorentz force is

$$\boldsymbol{F}(x_1,y_1,x_2,y_2,s,t) = q_2\big(\boldsymbol{E}(x_2,y_2,z=ct-s,t) + c\,\boldsymbol{e_z}\times\boldsymbol{B}(x_2,y_2,z=ct-s,t)\big).$$

The change in momentum of the test charge can be calculated as the time-integrated Lorentz force,

$$\Delta\boldsymbol{p}(x_1,y_1,x_2,y_2,s) = \int \boldsymbol{F}\,\mathrm{d}t.$$

This leads to the concept of wake functions.

The electromagnetic fields $\boldsymbol{E}_\mathrm{d}$ and $\boldsymbol{B}_\mathrm{d}$ and the Lorentz Force $\boldsymbol{F}_\mathrm{d}$ of a distributed source $\rho_\mathrm{d}(\boldsymbol{r}, t) = \eta(x_1 - \bar{x}_1, y_1 - \bar{y}_1)\lambda(z - ct)$ can be calculated either by integration over all source points,

$$\boldsymbol{F}_\mathrm{d}(\bar{x}_1, \bar{y}_1, x_2, y_2, s, t) = \int \boldsymbol{F}(x_1, y_1, x_2, y_2, s, t + z_1/c)\eta(x_1 - \bar{x}_1, y_1 - \bar{y}_1)\frac{\lambda(z_1)}{q_1}\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}z_1,$$

or by solving the electromagnetic problem for the distributed source; here λ is the line charge density, η is the transverse distribution normalized to 1, and $\bar{x}_1$ and $\bar{y}_1$ describe a transverse shift of the center of the distribution. A calculation of the electric fields of a distributed source is shown in Fig. 5.

A fundamental difference between fields of point particles (with time dependency $\delta(z - ct)$) and fields of distributed sources (with time dependency $\lambda(z - ct)$) is that the frequency spectrum of point particles is not limited. In particular, long Gaussian bunches may stimulate only a few resonances (modes) in a cavity structure, or even none.

We can distinguish between the long-range regime of the wake, where the interaction between particles is driven by resonances, and the short-range regime, where the superposition of time-harmonic cavity fields is not sufficient to describe the effects. For instance, the fields in Fig. 2 are not determined by oscillations, while the fields in Fig. 5 will ring harmonically on one or several frequencies after the (source) bunch has left the domain.

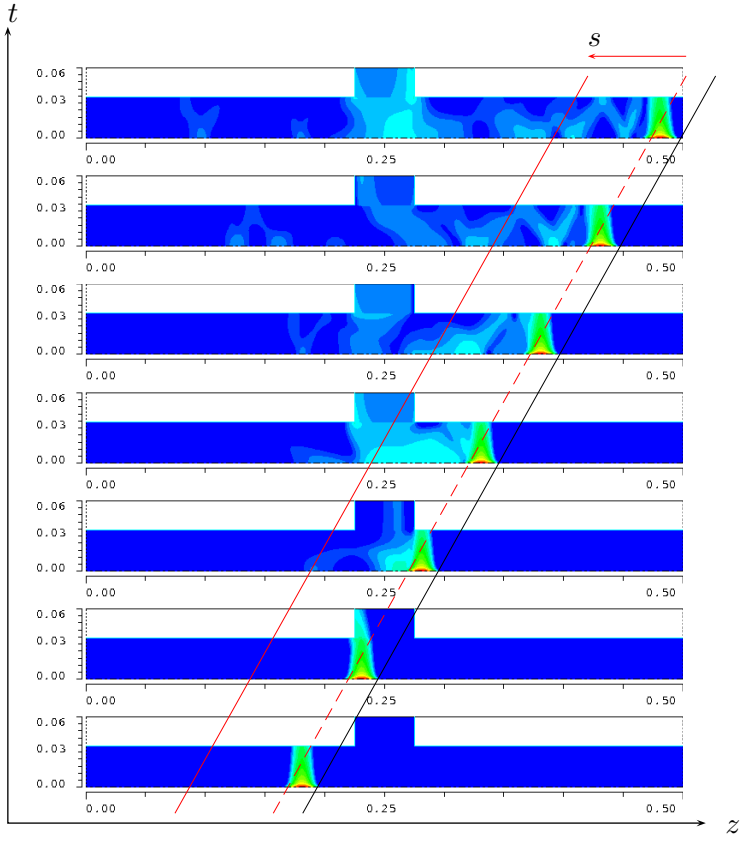


Fig. 5: Wake fields generated by a Gaussian bunch traversing a cavity

2.2 Basic definitions

Consider a point charge q_1 traversing a structure with offset (x_1, y_1) parallel to the z-axis at the speed of light (see Fig. 4). Then the *wake function* is defined as

$$\boldsymbol{w}(x_1, y_1, x_2, y_2, s) = \frac{1}{q_1} \int_{-\infty}^{\infty} \mathrm{d}z \left[\boldsymbol{E}(x_2, y_2, z, t) + c\, \boldsymbol{e}_z \times \boldsymbol{B}(x_2, y_2, z, t)\right]_{t=(s+z)/c}.$$

The distance s is measured from the source q_1 in the opposite direction to $\boldsymbol{v}$. The change in momentum of a test particle with charge q_2 following behind at a distance s with offset (x_2, y_2) is given by

$$\Delta \boldsymbol{p} = \frac{1}{c}\, q_1 q_2 \boldsymbol{w}(s).$$

Since $\boldsymbol{e_z} \cdot (\boldsymbol{e_z} \times \boldsymbol{B}) = 0$, the longitudinal component of the wake function is simply

$$w_{\parallel}(x_1, y_1, x_2, y_2, s) = \frac{1}{q_1} \int_{-\infty}^{\infty} \mathrm{d}z\, E_z(x_2, y_2, z, (s+z)/c).$$

Figure 6 shows the longitudinal component of the wake potential for the above example with the small cavity. The gray line represents the Gaussian charge distribution in the range from -5σ to 10σ. Owing to transient wake field effects, the head of the bunch (left-hand side of the figure) is decelerated, while a test charge at a certain position behind the bunch will be accelerated.

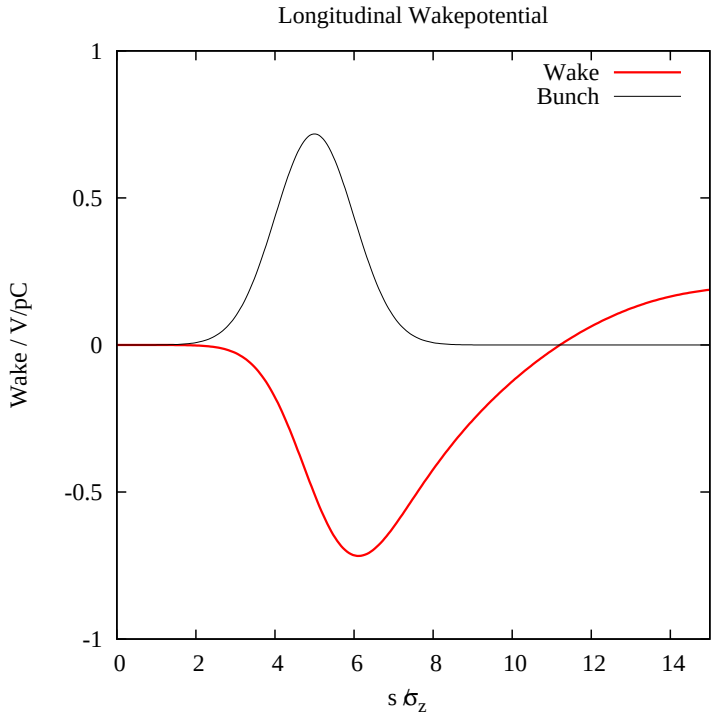


Fig. 6: Longitudinal wake potential

The notion of wakes, as presented above, is restricted to sources and test particles that travel at the velocity of light through a structure with semi-infinite input and output beam pipes. Therefore, for the integrals to converge, it is necessary that there be no length-independent forces in the pure beam pipes. This is the case for $v \to c$ and perfect conductivity of the pipes. The concept of a wake per length,

$$\boldsymbol{w}'(x_1, y_1, x_2, y_2, s) = \frac{1}{q_1}\left[\boldsymbol{E}_{\mathrm{p}}(x_2, y_2, -s, 0) + v\, \boldsymbol{e}_z \times \boldsymbol{B}_{\mathrm{p}}(x_2, y_2, -s, 0)\right],$$

is used to describe the effect in beam pipes 'p' of finite conductivity and/or velocity $v \leq c$. Suppose that the input and output beam pipes have the same cross-section; then a generalized wake function

$$\boldsymbol{w}_\mathrm{s}(x_1, y_1, x_2, y_2, s) = \frac{1}{q_1}\int_{-\infty}^{\infty} \mathrm{d}z \left[\boldsymbol{E}_\mathrm{s}(x_2, y_2, z, t) + v\,\boldsymbol{e}_z \times \boldsymbol{B}_\mathrm{s}(x_2, y_2, z, t)\right]_{t=(s+z)/v}$$

can be defined for the scattered fields $\boldsymbol{E}_\mathrm{s} = \boldsymbol{E} - \boldsymbol{E}_\mathrm{p}$ and $\boldsymbol{B}_\mathrm{s} = \boldsymbol{B} - \boldsymbol{B}_\mathrm{p}$. If the conditions for the wake function are fulfilled (i.e. convergence of the integral), then the wake function equals the generalized wake function.

The *wake potential* is defined similarly to the wake function, but for a distributed source:

$$\begin{aligned}\boldsymbol{W}(\bar{x}_1, \bar{y}_1, x_2, y_2, s) &= \frac{1}{q_1}\int_{-\infty}^{\infty} \mathrm{d}z \left[\boldsymbol{E}_\mathrm{d}(x_2, y_2, z, t) + c\,\boldsymbol{e}_z \times \boldsymbol{B}_\mathrm{d}(x_2, y_2, z, t)\right]_{t=(s+z)/c} \\ &= \frac{1}{q_1 q_2}\int_{-\infty}^{\infty} \mathrm{d}z \left[\boldsymbol{F}_\mathrm{d}(\bar{x}_1, \bar{y}_1, x_2, y_2, z, t)\right]_{t=(s+z)/c}.\end{aligned}$$

It can be calculated from the wake function by the convolution

$$\boldsymbol{W}_\mathrm{d}(\bar{x}_1, \bar{y}_1, x_2, y_2, s) = \int \boldsymbol{w}(x_1, y_1, x_2, y_2, s + z_1)\eta(x_1 - \bar{x}_1, y_1 - \bar{y}_1)\frac{\lambda(z_1)}{q_1}\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}z_1.$$

Note that the s-coordinate measures in the negative z-direction while λ depends on the positive longitudinal coordinate. Usually numerical codes for computing wakes, such as ECHO, calculate electromagnetic fields for distributed sources and therefore wake potentials.

2.3 Some theory

2.3.1 The Panofsky–Wenzel theorem

We follow the arguments of A. Chao [2, 15] to introduce the Panofsky–Wenzel theorem [16]. Therefore we use the following different notation for the generalized wake function:

$$\boldsymbol{w}_\mathrm{p}(x_1, y_1, x_2, y_2, s) = \boldsymbol{w}_\mathrm{p}(x_1, y_1, \boldsymbol{r}_2),$$

with the observer vector $\boldsymbol{r}_2 = x_2\boldsymbol{e}_x + y_2\boldsymbol{e}_y - s\boldsymbol{e}_z$. We calculate curl $\boldsymbol{w}_\mathrm{p}$ with respect to the observer or the test particle:

$$\nabla_2 \times \boldsymbol{w}_\mathrm{p}(x_1, y_1, \boldsymbol{r}_2) = \nabla_2 \times \frac{v}{q_1}\int_{-\infty}^{\infty} \mathrm{d}t \left[\boldsymbol{E}_\mathrm{s}(\boldsymbol{r}_2 + \boldsymbol{v}t, t) + \boldsymbol{v} \times \boldsymbol{B}_\mathrm{s}(\boldsymbol{r}_2 + \boldsymbol{v}t, t)\right].$$

Using $\operatorname{curl}\boldsymbol{E} = -\partial\boldsymbol{B}/\partial t$ gives

$$\begin{aligned}\nabla_2 \times \boldsymbol{w}_\mathrm{p}(x_1, y_1, \boldsymbol{r}_2) &= \frac{v}{q_1}\int_{-\infty}^{\infty} \mathrm{d}t \left[-\frac{\partial}{\partial t}\boldsymbol{B}_\mathrm{s}(\ldots, t) + \boldsymbol{v}(\nabla_2 \boldsymbol{B}_\mathrm{s}(\ldots, t)) - \boldsymbol{B}_\mathrm{s}(\ldots, t)(\nabla_2 \boldsymbol{v})\right] \\ &= \frac{v}{q_1}\int_{-\infty}^{\infty} \mathrm{d}t \left[-\frac{\partial}{\partial t} - v\frac{\partial}{\partial z}\right]\boldsymbol{B}_\mathrm{s}(\boldsymbol{r}_2 + \boldsymbol{v}t, t) \\ &= \frac{v}{q_1}\int_{-\infty}^{\infty} \mathrm{d}t \left[-\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}_\mathrm{s}(\boldsymbol{r}_2 + \boldsymbol{v}t, t)\right] \\ &= -\frac{v}{q_1}\boldsymbol{B}_\mathrm{s}(\boldsymbol{r}_2 + \boldsymbol{v}t, t)\Big|_{t=-\infty}^{t=\infty}.\end{aligned}$$

As the scattered field is zero for negative infinite time and vanishes for positive infinite time and infinite distance from the scattering object, the wake is curl-free. The Panofsky–Wenzel theorem is then reformulated in our original notation as the set of equations

$$\frac{\partial}{\partial s}w_{\mathrm{p}x} = -\frac{\partial}{\partial x_2}w_{\mathrm{p}\parallel},$$

$$\frac{\partial}{\partial s} w_{\mathrm{p}y} = -\frac{\partial}{\partial y_2} w_{\mathrm{p}\parallel},$$
$$\frac{\partial}{\partial x_2} w_{\mathrm{p}y} = \frac{\partial}{\partial y_2} w_{\mathrm{p}x}.$$

Note that the Panofsky–Wenzel theorem holds for the generalized wake function ($v \leq c$) and for the wake function ($v = c$).

Integration of the transverse gradient of the longitudinal wake function yields the transverse wake potential

$$\boldsymbol{w}_\perp(x_1, y_1, x_2, y_2, s) = -\boldsymbol{\nabla}_{2\perp} \int_{-\infty}^{s} \mathrm{d}s'\, w_\parallel(x_1, y_1, x_2, y_2, s').$$

2.3.2 Wake is harmonic with respect to observer offset

Now we calculate $\operatorname{div} \boldsymbol{w}$ with respect to the observer. First, note that

$$\nabla_2 \cdot \boldsymbol{w}(x_1, y_1, \boldsymbol{r}_2) = \nabla_2 \cdot \frac{c}{q_1} \int_{-\infty}^{\infty} \mathrm{d}t \left[\boldsymbol{E}(\boldsymbol{r}_2 + \boldsymbol{c}t, t) + \boldsymbol{c} \times \boldsymbol{B}(\boldsymbol{r}_2 + \boldsymbol{c}t, t)\right].$$

Using Maxwell's equations, $\operatorname{div} \boldsymbol{E} = \rho/\varepsilon$ and $\operatorname{curl} \boldsymbol{B} = \mu \boldsymbol{J} + c^{-2} \partial \boldsymbol{E}/\partial t$, together with $\boldsymbol{J} = \boldsymbol{c}\rho$ gives

$$\begin{aligned}
\nabla_2 \cdot \boldsymbol{w}(x_1, y_1, \boldsymbol{r}_2) &= \frac{c}{q_1} \int_{-\infty}^{\infty} \mathrm{d}t \left[\nabla_2 \cdot \boldsymbol{E} + \boldsymbol{c}(\nabla_2 \times \boldsymbol{B})\right] \\
&= \frac{1}{q_1} \int_{-\infty}^{\infty} \mathrm{d}t \left[-\frac{\partial}{\partial t} E_z(\boldsymbol{r}_2 + \boldsymbol{c}t, t)\right] \\
&= -\frac{1}{q_1} \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{\partial}{\partial s} E_z(\boldsymbol{r}_2 + z\boldsymbol{e}_z, (z+s)/c)\right] \\
&= -\frac{\partial}{\partial s}\, w_\parallel(x_1, y_1, x_2, y_2, s).
\end{aligned}$$

The term $\partial w_\parallel/\partial s$ appears on both sides of the equation, so we can write

$$\frac{\partial w_x}{\partial x_2} + \frac{\partial w_y}{\partial y_2} = 0.$$

With the Panofsky–Wenzel equations we find that the longitudinal wake is a harmonic function with respect to the transverse coordinates of the test particle:

$$\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2}\right) w_\parallel = -\frac{\partial}{\partial s} \left(\frac{\partial w_x}{\partial x_2} + \frac{\partial w_y}{\partial y_2}\right) = 0.$$

2.3.3 Wake is harmonic with respect to source offset

The longitudinal wake is also a harmonic function with respect to the transverse coordinates of the source particle [17], i.e. $L_1 w_\parallel = 0$ with $L_1 = \partial^2/\partial x_1^2 + \partial^2/\partial y_1^2$. To prove this, we have to calculate $\tilde{E}_z = L_1 E_z$, which is equivalent to the solution of the field problem for the source $\tilde{\rho} = L_1 \rho$. The source ρ is the point particle q_1 travelling at the speed of light along $(x_1, y_1, z = ct)$. It gives rise to the electromagnetic fields

$$\boldsymbol{E}_\mathrm{f} = q_1 \frac{\delta(z - ct)}{2\pi\varepsilon} \frac{(x - x_1)\boldsymbol{e}_x + (y - y_1)\boldsymbol{e}_y}{(x - x_1)^2 + (y - y_1)^2},$$
$$\boldsymbol{B}_\mathrm{f} = c^{-1} \boldsymbol{e}_z \times \boldsymbol{E}_\mathrm{f}$$

in free space. The fields $\tilde{\boldsymbol{E}} = L_1 \boldsymbol{E}_\mathrm{f}$ and $\tilde{\boldsymbol{B}} = L_1 \boldsymbol{B}_\mathrm{f}$ are caused by the source $\tilde{\rho} = L_1 \rho$. These fields are zero for all points with $(x, y) \neq (x_1, y_1)$, as

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} \right) \frac{(x - x_1)\boldsymbol{e}_x + (y - y_1)\boldsymbol{e}_y}{(x - x_1)^2 + (y - y_1)^2} = \boldsymbol{0}.$$

Obviously $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{B}}$ satisfy any linear boundary condition for any geometry, provided that the boundary does not intersect the trajectory $(x_1, y_1, z = ct)$. Therefore these fields are also solutions to the bounded wake problem, and all components of $\boldsymbol{w}$ are harmonic with respect to (x_1, y_1), since

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} \right) \boldsymbol{w} = \frac{c}{q_1} \int_{-\infty}^{\infty} \mathrm{d}t \left[\tilde{\boldsymbol{E}} + \boldsymbol{c} \times \tilde{\boldsymbol{B}} \right] = \boldsymbol{0}.$$

This information will help us to evaluate the r-dependence of the wake function in cylindrical symmetric structures in the next subsection, and it will enable us to efficiently calculate the wake function in fully 3D structures.

2.3.4 Restrictions

The Panofsky–Wenzel theorem is applicable if the input and output beam pipes have the same cross-section. The longitudinal wake is harmonic if the trajectories (x_1, y_1, ct) and (x_2, y_2, ct) do not intersect with the boundary.

2.4 Wake function in cylindrically symmetric structures

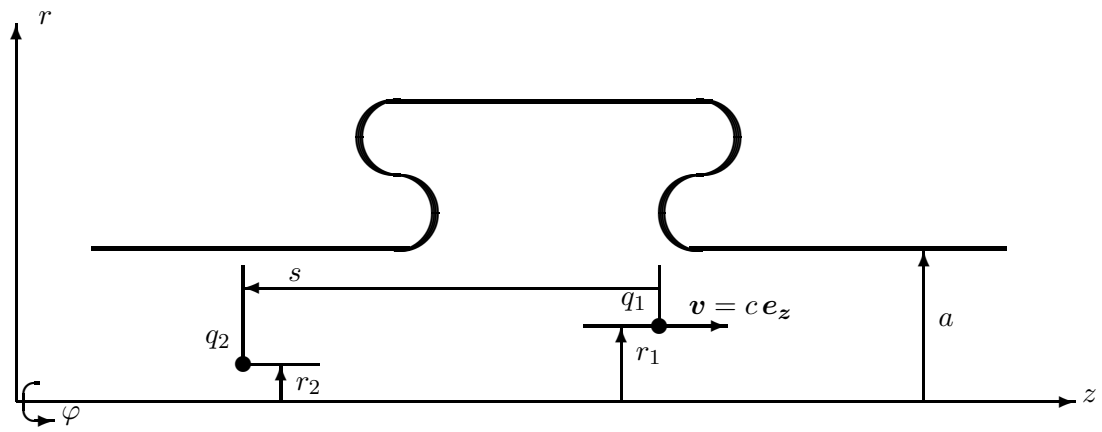


Fig. 7: A bunch with total charge q_1 traversing a cavity with offset r_1, followed by a test charge q_2 with offset r_2

Consider now a cylindrically symmetric acceleration cavity with side tubes of radius a (see Fig. 7). The particular shape in the region $r > a$ is of no importance for the following investigations. Two charges pass through the structure from left to right with the speed of light: q_1 at a radius of r_1 and q_2 at a radius of r_2. We wish to find an expression for the net change in momentum, $\Delta \boldsymbol{p}(r_1, \varphi_1, r_2, \varphi_2, s)$, experienced by q_2 due to the wake fields generated by q_1. In the following we write the wake function and potential in polar coordinates. Let us start with the case of $\varphi_1 = 0$:

$$\Delta p_z(r_1, 0, r_2, \varphi_2, s) = q_1 \, q_2 \, w_\parallel(r_1, 0, r_2, \varphi_2, s).$$

The wake function can be expanded in a multipole series

$$w_\parallel(r_1, 0, r_2, \varphi_2, s) = \mathrm{Re} \left\{ \sum_{m=-\infty}^{\infty} \exp(\mathrm{i}\, m\, \varphi_2) G_m(r_1, r_2, s) \right\}.$$

Since $w_\parallel$ is a harmonic function in (r_2, φ_2), we have

$$\begin{aligned} L_2\, w_\parallel(r_1, 0, r_2, \varphi_2, s) &= \left(\frac{1}{r_2}\frac{\partial}{\partial r_2} r_2 \frac{\partial}{\partial r_2} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi_2^2}\right) w_\parallel(r_1, 0, r_2, \varphi_2, s) \\ &= \mathrm{Re}\left\{ \sum_{m=-\infty}^{\infty} \exp(\mathrm{i}\, m\, \varphi_2) \left(\frac{1}{r_2}\frac{\partial}{\partial r_2} r_2 \frac{\partial}{\partial r_2} - \frac{m^2}{r_2^2}\right) G_m(r_1, r_2, s) \right\} \\ &= 0, \end{aligned}$$

where L_2 is the transverse Laplace operator with respect to the offset of the test particle. So, for all m, the expansion functions $G_m(r_1, r_2, s)$ have to satisfy the Poisson equation

$$\frac{1}{r_2}\frac{\partial}{\partial r_2}\left(r_2 \frac{\partial}{\partial r_2} G_m(r_1, r_2, s)\right) - \frac{m^2}{r_2^2} G_m(r_1, r_2, s) = 0.$$

The solutions are

$$\begin{aligned} G_0(r_1, r_2, s) &= U_0(r_1, s) + V_0(r_1, s)\, \ln r_2, \\ G_m(r_1, r_2, s) &= U_m(r_1, s)\, r_2^m + V_m(r_1, s)\, r_2^{-m} \quad \text{for } m > 0. \end{aligned}$$

Keeping only the solutions which are regular at the origin ($r_2 = 0$), the longitudinal wake potential can be written as

$$w_\parallel(r_1, 0, r_2, \varphi_2, s) = \sum_{m=0}^{\infty} r_2^m\, U_m(r_1, s)\, \cos m\varphi_2,$$

with expansion functions $U_m(r_1, s)$ that depend on the details of the given cavity geometry.

By azimuthal symmetry, the dependence on φ_1 is $w_\parallel(r_1, \varphi_1, r_2, \varphi_2, s) = w_\parallel(r_1, 0, r_2, \varphi_2 - \varphi_2, s)$, as longitudinal fields depend only on the relative azimuthal angle of the observer with respect to the source. Using the fact that $w_\parallel$ is also a harmonic function in (r_1, φ_1), we find with the same arguments as before that $U_m(r_1, s)$ can be factorized as $r_1^m w_m(s)$.

It follows that for the general case of a charge q_1 at (r_1, φ_1) generating fields that act on a second charge q_2 at (r_2, φ_2), the *longitudinal wake function* is given by

$$w_\parallel(r_1, \varphi_1, r_2, \varphi_2, s) = \sum_{m=0}^{\infty} r_1^m\, r_2^m\, w_m(s)\, \cos m(\varphi_2 - \varphi_1).$$

The transverse wake function is, by the Panofsky–Wenzel theorem,

$$\begin{aligned} \boldsymbol{w}_\perp(r_1, \varphi_1, r_2, \varphi_2, s) &= -\left(\boldsymbol{e_r}\, \frac{\partial}{\partial r_2} + \boldsymbol{e_\varphi}\, \frac{1}{r_2}\frac{\partial}{\partial \varphi_2}\right) \int_{-\infty}^{s} \mathrm{d}s'\, w_\parallel(r_1, \varphi_1, r_2, \varphi_2, s') \\ &= \sum_{m=0}^{\infty} \Bigg\{ -\boldsymbol{e_r}\, m\, {r_1}^m\, {r_2}^{m-1} \int_{-\infty}^{s} \mathrm{d}s'\, w_m(s')\, \cos m(\varphi_2 - \varphi_1) \\ &\qquad + \boldsymbol{e_\varphi}\, m\, {r_1}^m\, {r_2}^{m-1} \int_{-\infty}^{s} \mathrm{d}s'\, w_m(s')\, \sin m(\varphi_2 - \varphi_1) \Bigg\}. \end{aligned}$$

Each azimuthal order is fully characterized by a scalar function $w_m(s)$. This function can be calculated by solving Maxwell's equations for the given geometry and any choice of $(r_1, \varphi_1, r_2, \varphi_2)$, yielding

$$w_m(s) = \frac{\int_{-\infty}^{\infty} \mathrm{d}z\, E_{zm}(r_2, \varphi_2, z, (z+s)/c))}{r_1^m\, r_2^m\, \cos m(\varphi_2 - \varphi_1)}.$$

A particular choice of r_2 can be used to avoid the infinite integration range: since E_z vanishes at the metallic tube boundary, only the cavity gap contributes to the integral. The integration range is reduced to the cavity gap by setting r_2 to the radius of the beam tube. This trick is possible if no obstacle intersects with the infinite cylindrical beam pipe.

This type of wake integration is utilized by computer codes such as ECHO [18, 19] for bunches of finite length. Wake potentials can be calculated by such programs in the time domain, but wake functions (of point sources) need asymptotic considerations; see [20].

It should be mentioned that in many practical cases, due to the $(r/a)^m$ dependence, the longitudinal wake is dominated by the monopole term and the transverse wakes by the dipole term:

$$w_\parallel(r_1, \varphi_1, r_2, \varphi_2, s) = w_0(s),$$
$$\boldsymbol{w}_\perp(r_1, \varphi_1, r_2, \varphi_2, s) = r_1 \int_{-\infty}^{s} \mathrm{d}s'\, w_1(s') \left[-\boldsymbol{e_r}\, \cos(\varphi_2 - \varphi_1) + \boldsymbol{e_\varphi}\, \sin(\varphi_2 - \varphi_1)\right].$$

2.5 Fully 3D structures

While for cylindrical symmetric structures the dependence of the wake on transverse coordinates is explicitly known and can be used to reduce the integration range and domain of the field calculation, more general structures require us to use the harmonic property of the wake for a beam tube of arbitrary shape. The simple 3D cavity in Fig. 8, with a beam tube of square cross-section, is used to demonstrate this. We suppress the dependence of the wake function (or potential) on the offset of the source and write simply $\tilde{W}_\parallel(x, y, s) = W_\parallel(x_1, y_1, x, y, s)$. This function is harmonic in the observer offset,

$$\boldsymbol{\nabla}_\perp^2\, \tilde{W}_\parallel(x, y, s) = 0.$$

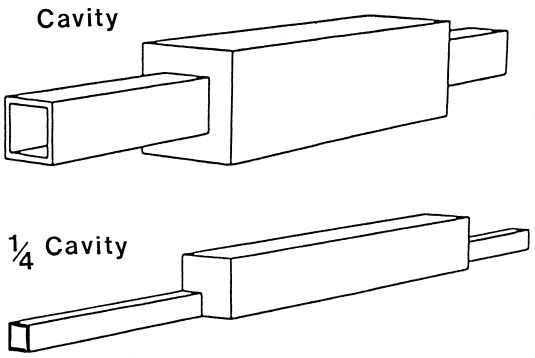


Fig. 8: A 3D cavity structure with two symmetry planes (top) and a quarter of the structure (bottom)

For points x and y on the surface of the beam tube, we can calculate the wake by a finite-range integration through the cavity gap, as shown in Fig. 9. If we know $\tilde{W}$ for all surface points, we can calculate the wake for any point inside the tube by numerical solution of the boundary value problem. Therefore a 2D Poisson problem has to be solved. In our example, with two transverse symmetries, only a quarter of the structure needs to be considered to calculate the wake of a source in the center.

The transverse wake potential can be calculated from the longitudinal one using the Panofsky–Wenzel theorem. The transverse gradient of the longitudinal wake potential in a beam tube is also indicated in Fig. 9.

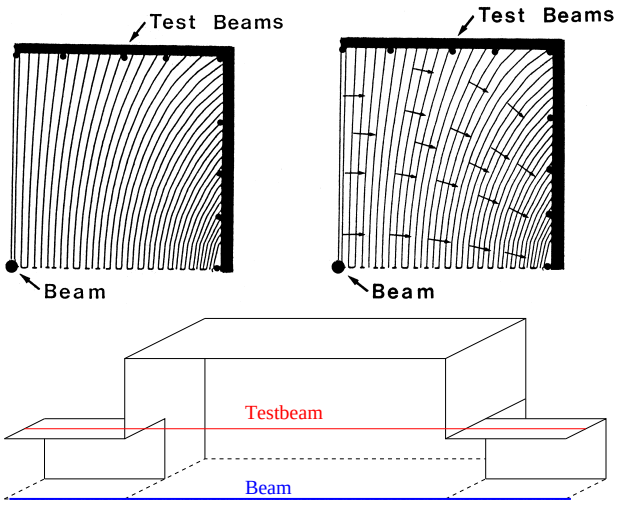


Fig. 9: Illustration of the indirect test beam method. The upper pictures show lines of constant longitudinal wake potential and the gradient of the longitudinal wake potential; an integration gives the transverse wake potential according to the Panofsky–Wenzel theorem. The lower diagram depicts the paths of the beam and the test beam.

3 Cavities, resonant structures and eigenmodes

3.1 Eigenmodes

Many structures in an accelerator environment can be considered as a hollow space with semi-infinite beam pipes on both sides. Usually this vacuum volume is bounded by metal surfaces with high conductivity. As a good approximation, the cavity walls can be treated as perfect electric conducting (PEC) boundaries, and sometimes the beam pipes are even neglected so that the volume is closed.

Electromagnetic fields with frequencies below the lowest cutoff frequency of the beam pipes are trapped in the volume, and the fields oscillate at discrete frequencies:

$$\boldsymbol{E}(\boldsymbol{r},t) = \sum_\nu \hat{A}_\nu \hat{\boldsymbol{E}}_\nu(\boldsymbol{r}) \cos(\hat{\omega}_\nu t + \hat{\varphi}_\nu),$$
$$\boldsymbol{B}(\boldsymbol{r},t) = \sum_\nu \hat{A}_\nu \hat{\boldsymbol{B}}_\nu(\boldsymbol{r}) \sin(\hat{\omega}_\nu t + \hat{\varphi}_\nu).$$

These oscillations are called eigenmodes or cavity modes. They are characterized by their field patterns $\hat{\boldsymbol{E}}_\nu(\boldsymbol{r})$ and $\hat{\boldsymbol{B}}_\nu(\boldsymbol{r})$ and their eigenfrequencies $\hat{\omega}_\nu$. The modes may ring with any amplitude $\hat{A}_\nu$ and phase $\hat{\varphi}_\nu$, and the amplitude normalization of the eigenfields is arbitrary. Such modes are called standing-wave modes, as the electric and magnetic fields ring at all spatial points with the same phase, but the electric field is phase-shifted by 90° relative to the magnetic field. For simplicity, in the following we omit the mode index ν but indicate all indexable (mode-specific) quantities with a hat. We will introduce further mode-specific quantities, such as the quality $\hat{Q}$, the modal longitudinal loss parameter $\hat{k}$, and the mode energy

$$\hat{\mathcal{W}} = \frac{1}{2}\int \varepsilon \hat{E}^2 \,\mathrm{d}V = \frac{1}{2}\int \mu^{-1} \hat{B}^2 \,\mathrm{d}V,$$

which depends on the arbitrary amplitude normalization. The total electromagnetic field energy of all the modes is[1]

$$\mathcal{W} = \frac{1}{2}\int \varepsilon E(\boldsymbol{r},\boldsymbol{t})^2 \,\mathrm{d}V + \frac{1}{2}\int \mu^{-1} B(\boldsymbol{r},\boldsymbol{t})^2 \,\mathrm{d}V = \sum |\hat{A}|^2 \,\hat{\mathcal{W}}.$$

[1]The field energy of a particular mode does not depend on the stimulation of other modes, as the mode fields are orthogonal to each other; see Appendix A.

Eigenmodes can be computed with electromagnetic field solvers such as those in [9, 10]; see also Fig. 10. Usually closed volumes are considered, which are completely surrounded by PEC or perfect magnetic conducting (PMC) surfaces. As the mode field in beam pipes decays exponentially, even open problems (involving infinitely long pipes) can be handled with such programs, by using a perfectly conducting boundary after a sufficiently long piece of pipe.

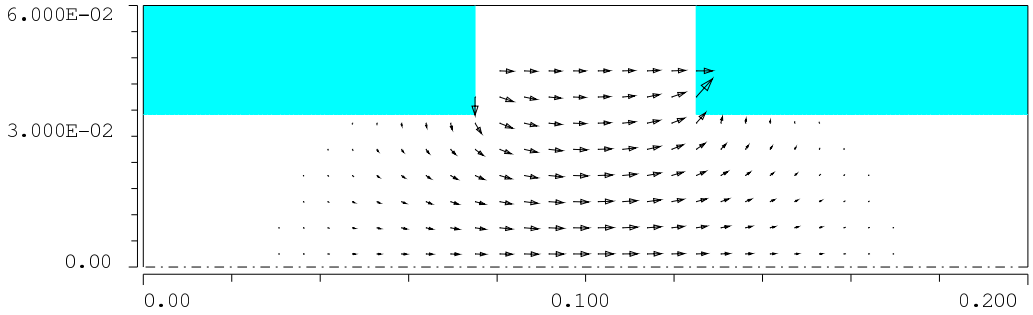


Fig. 10: Electric field of a mode in a rotationally symmetric cavity with beam pipes

In structures with symmetries (e.g. rotational symmetry), eigenmodes and beam-pipe modes of the same symmetry condition are coupled. Therefore the lowest cutoff frequency for a particular symmetry defines the highest possible eigenfrequency for the corresponding eigenmodes. For instance, monopole modes may have resonance frequencies that are above the lowest dipole mode cutoff frequency, which is lower than the lowest monopole mode cutoff frequency. Beyond that, there can exist quasi-trapped modes above the lowest cutoff frequency that have very weak coupling to the pipes. The energy flow (per period) of such fields into the beam pipes may be comparable to the energy loss (per period) of non-trapped modes to non-perfectly conducting metallic boundaries.

3.2 Excitation of eigenmodes and the per-mode loss parameter

We consider a cavity of length[2] L and a bunch with charge q_1, offset (x_1, y_1) and velocity c, which enters the cavity at time $t = 0$. The electromagnetic fields after the charge has left the cavity, namely

$$\boldsymbol{E}(\boldsymbol{r}, t > L/c) = \sum \mathrm{Re}\big\{\hat{A}\hat{\boldsymbol{E}}(\boldsymbol{r})\exp(\mathrm{i}\,\hat{\omega}t)\big\} + \boldsymbol{E}_\mathrm{r}(\boldsymbol{r}, t),$$
$$\boldsymbol{B}(\boldsymbol{r}, t > L/c) = \sum \mathrm{Im}\big\{\hat{A}\hat{\boldsymbol{B}}(\boldsymbol{r})\exp(\mathrm{i}\,\hat{\omega}t)\big\} + \boldsymbol{B}_\mathrm{r}(\boldsymbol{r}, t),$$

can be split into eigenfields and a residual part, $\boldsymbol{E}_\mathrm{r}$ or $\boldsymbol{B}_\mathrm{r}$. The long-range interaction between bunches or particles is essentially driven by the modal part, as the residual fields decay or are not stimulated resonantly. The complex mode amplitudes are proportional to the source charge and depend on the source offset. Hence they can be expressed as $\hat{A} = q_1\hat{f}(x_1, y_1)$.

Suppose that a small test charge δq follows the source particle on the same trajectory at a distance of $s > 0$. It induces the additional amplitude $\delta\hat{A} = \delta q\exp(-\mathrm{i}\,\hat{\omega}s/c)\hat{f}(x_1, y_1)$. Therefore the energy of the modes is increased by

$$\begin{aligned}\delta\mathcal{W}_\mathrm{modes} &= \sum\big(|\hat{A} + \delta\hat{A}|^2 - |\hat{A}|^2\big)\hat{\mathcal{W}}\\ &\approx \sum 2\,\mathrm{Re}\big\{\hat{A}\delta\hat{A}^*\big\}\hat{\mathcal{W}}\\ &\approx 2q_1\delta q\sum|\hat{f}(x_1, y_1)|^2\,\mathrm{Re}\big\{\exp(\mathrm{i}\,\hat{\omega}s/c)\big\}\hat{\mathcal{W}}.\end{aligned}$$

[2]The relevant length is not exactly the length of the cavity, but rather the length with non-zero field of the modes. For open structures, with beam pipes, this length is in principle infinite, but for practical considerations the field has decayed sufficiently after a pipe length of a few times the widest dimension of the cross-section.

On the other hand, the test particle gains kinetic energy

$$\begin{aligned}\delta\mathcal{W}_\mathrm{k} &= \int_{-\infty}^{\infty} \delta q\, E_z(x_1, y_1, z-s, z/c)\,\mathrm{d}z \\ &= q_1 \delta q \sum \mathrm{Re}\left\{ \hat{f}(x_1,y_1) \int_{-\infty}^{\infty} \hat{E}_z(x_1,y_1,z) \exp(\mathrm{i}\,\hat{\omega}(z+s)/c)\,\mathrm{d}z \right\} + \cdots .\end{aligned}$$

The sum of the field energy and the kinetic energy is conserved, if terms with the same oscillation frequency $\exp(\mathrm{i}\,\hat{\omega}s/c)$ cancel:

$$2|\hat{f}(x_1,y_1)|^2\,\hat{\mathcal{W}} + \hat{f}(x_1,y_1) \int_{-\infty}^{\infty} \hat{E}_z(x_1,y_1,z) \exp(\mathrm{i}\,\hat{\omega}(z)/c)\,\mathrm{d}z = 0.$$

This is satisfied with $\hat{f}(x_1,y_1) = -\hat{v}^*(x_1,y_1)/\sqrt{\hat{\mathcal{W}}}$ and the normalized mode voltages

$$\hat{v}(x,y) = \frac{1}{2\sqrt{\hat{\mathcal{W}}}} \int_{-\infty}^{\infty} \hat{E}_z(x,y,z) \exp(\mathrm{i}\,\hat{\omega}z/c)\,\mathrm{d}z,$$

which do not depend on the arbitrary normalization mode fields.

The amplitude excited by the charge q_1 is

$$\hat{A} = q_1 \hat{f}(x_1,y_1) = -q_1 \hat{v}^*(x_1,y_1)/\sqrt{\hat{\mathcal{W}}}\,,$$

and the energy of all modes is

$$\mathcal{W}_\mathrm{EM,modes} = \sum |\hat{A}|^2 \hat{W} = q_1^2 \sum \hat{k},$$

with the per-mode loss parameter

$$\hat{k} = |\hat{v}(x_1,y_1)|^2 = \frac{1}{4\hat{\mathcal{W}}} \left| \int_{-\infty}^{\infty} \hat{E}_z(x_1,y_1,z) \exp(\mathrm{i}\,\hat{\omega}(z)/c)\,\mathrm{d}z \right|^2 .$$

The excitation of mode amplitudes depends linearly on the source distribution: another particle with charge q_2 and offset (x_2, y_2) at a distance s gives rise to the additional amplitude

$$\hat{A} = -q_2 \hat{v}^*(x_2,y_2) \exp(-\mathrm{i}\,\hat{\omega}s/c)/\sqrt{\hat{\mathcal{W}}}\,,$$

with phase shift $-\hat{\omega}s/c$ due to the time shift s/c. Therefore it is possible to calculate the mode excitation for arbitrary charge distributions; for example, for a one-dimensional bunch with offset (x_1, y_1) and line charge density $\lambda(z,t) = \lambda(z-ct)$,

$$\hat{A} = \hat{v}^*(x_1,y_1) \int \lambda(-s) \exp(-\mathrm{i}\,\hat{\omega}s/c)\,\mathrm{d}s.$$

In particular, a Gaussian bunch with charge q_1 and rms length σ excites the amplitudes $\hat{A} = q_1 \hat{v}^*(x_1,y_1) \exp(-(\hat{\omega}\sigma/c)^2/2)$. We introduce the shape-dependent per-mode loss parameter

$$\hat{k}_\sigma = \hat{k} \exp(-(\hat{\omega}\sigma/c)^2).$$

The electromagnetic field energy of all modes, after such a bunch has traversed the cavity, is

$$\mathcal{W}_{\mathrm{EM,modes},\sigma} = \sum |\hat{A}|^2 \hat{\mathcal{W}} = q_1^2 \sum \hat{k}_\sigma.$$

3.3 Contribution of eigenmodes to the wake function

After the source particle has traversed the cavity, the electromagnetic fields are

$$\boldsymbol{E}(\boldsymbol{r}, t > L/c) = q_1 \sum \mathrm{Re}\left\{-\hat{v}^*(x_1, y_1)\hat{\mathcal{W}}^{-1/2}\hat{\boldsymbol{E}}(\boldsymbol{r})\exp(\mathrm{i}\,\hat{\omega}t)\right\} + \boldsymbol{E}_\mathrm{r}(\boldsymbol{r}, t),$$
$$\boldsymbol{B}(\boldsymbol{r}, t > L/c) = q_1 \sum \mathrm{Im}\left\{-\hat{v}^*(x_1, y_1)\hat{\mathcal{W}}^{-1/2}\hat{\boldsymbol{B}}(\boldsymbol{r})\exp(\mathrm{i}\,\hat{\omega}t)\right\} + \boldsymbol{B}_\mathrm{r}(\boldsymbol{r}, t),$$

Therefore the momentum of a test charge q_2 at a distance $s > L$ behind the source, with offset (x_2, y_2) and velocity c, is changed by

$$\Delta\boldsymbol{p} = \frac{q_1 q_2}{c}\sum \mathrm{Re}\left\{-\hat{v}^*(x_1, y_1)\hat{\mathcal{W}}^{-1/2}\int_{-\infty}^{\infty}\mathrm{d}z\left[\left(\hat{\boldsymbol{E}} - \mathrm{i}\,\boldsymbol{c}\times\hat{\boldsymbol{B}}\right)\exp(\mathrm{i}\,\hat{\omega}(z+s)/c)\right]\right\} + \Delta\boldsymbol{p}_\mathrm{r},$$

where the term $\Delta\boldsymbol{p}_\mathrm{r}$ stands for the contribution of the residual fields. Likewise, we can split the wake into modal and residual parts:

$$\boldsymbol{w}(x_1, y_1, x_2, y_2, s > L) = \frac{c\Delta\boldsymbol{p}}{q_1 q_2} = \sum \hat{\boldsymbol{w}}(x_1, y_1, x_2, y_2, s) + \boldsymbol{w}_\mathrm{r}(x_1, y_1, x_2, y_2, s),$$

where

$$\hat{\boldsymbol{w}}(x_1, y_1, x_2, y_2, s > L) = -2\,\mathrm{Re}\big\{\hat{v}^*(x_1, y_1)\hat{\boldsymbol{v}}(x_2, y_2)\exp(\mathrm{i}\,\hat{\omega}s/c)\big\},$$

with the normalized vectorial voltages

$$\hat{\boldsymbol{v}}(x, y) = \frac{1}{2\sqrt{\hat{\mathcal{W}}}}\int_{-\infty}^{\infty}\mathrm{d}z\left[\left(\hat{\boldsymbol{E}}(x, y, z) - \mathrm{i}\,\boldsymbol{c}\times\hat{\boldsymbol{B}}(x, y, z)\right)\exp(\mathrm{i}\,\hat{\omega}z/c)\right].$$

3.4 Causality and the fundamental theorem of beam loading

Consider two point particles with charges and coordinates q_1, x_1, y_1, $z_1 = ct$ and q_2, x_2, y_2, $z_2 = ct - s$. The distance parameter s may be positive or negative. The electromagnetic fields are caused by both particles together. Therefore the integrated longitudinal fields observed by the particles are

$$V_1 = \int E_z(x_1, y_1, z_1, t)\,\mathrm{d}z = q_1 w_\parallel(x_1, y_1, x_1, y_1, 0) + q_2 w_\parallel(x_2, y_2, x_1, y_1, -s),$$
$$V_2 = \int E_z(x_2, y_2, z_2, t)\,\mathrm{d}z = q_1 w_\parallel(x_1, y_1, x_2, y_2, s) + q_2 w_\parallel(x_2, y_2, x_2, y_2, 0).$$

The gain of electromagnetic field energy, $\Delta\mathcal{W}_\mathrm{EM} = -q_1 V_1 - q_2 V_2$, is always positive, as the interaction volume is initially field-free; it is

$$\begin{aligned}\Delta\mathcal{W}_\mathrm{EM,total} = &-q_1^2 w_\parallel(x_1, y_1, x_1, y_1, 0)\\ &- q_2^2 w_\parallel(x_2, y_2, x_2, y_2, 0)\\ &- q_1 q_2\big(w_\parallel(x_1, y_1, x_2, y_2, s) + w_\parallel(x_2, y_2, x_1, y_1, -s)\big).\end{aligned}$$

It is natural to write the electromagnetic field energy in terms of the per-mode and residual wake functions,

$$\begin{aligned}\Delta\mathcal{W}_\mathrm{EM,total} = &-q_1^2\sum \hat{w}_\parallel(x_1, y_1, x_1, y_1, 0)\\ &- q_2^2\sum \hat{w}_\parallel(x_2, y_2, x_2, y_2, 0)\\ &- q_1 q_2\sum\big(\hat{w}_\parallel(x_1, y_1, x_2, y_2, s) + \hat{w}_\parallel(x_2, y_2, x_1, y_1, -s)\big)\end{aligned}$$

$$+ \Delta w_{\parallel \mathrm{r}},$$

and to compare this expression with the field energy calculated from the amplitudes $\hat{A}$ of the modes. The mode amplitudes excited by both particles are just given by the superposition $\hat{A} = \hat{A}^{(1)} + \hat{A}^{(2)}$ of $\hat{A}^{(1)} = -q_1 \hat{v}^*(x_1, y_1)/\sqrt{\hat{\mathcal{W}}}$, excited by q_1, and $\hat{A}^{(2)} = -q_2 \hat{v}^*(x_2, y_2)/\sqrt{\hat{\mathcal{W}}} \exp(-\mathrm{i}\,\hat{\omega} s)$, excited by q_2. Therefore the gain of field energy stored in the oscillating modes is

$$\begin{aligned}
\Delta \mathcal{W}_{\mathrm{EM,modes}} &= \sum \left| q_1 \hat{v}^*(x_1, y_1) + q_2 \hat{v}^*(x_2, y_2) \exp(-\mathrm{i}\,\hat{\omega} s/c) \right|^2 \\
&= q_1^2 \sum |v(x_1, y_1)|^2 \\
&\quad + q_2^2 \sum |v(x_2, y_2)|^2 \\
&\quad + 2 q_1 q_2 \sum \mathrm{Re}\big\{ \hat{v}^*(x_1, y_1) \hat{v}(x_2, y_2) \exp(\mathrm{i}\,\hat{\omega} s) \big\}.
\end{aligned}$$

Comparing terms of the same mode with the same charges, we get

$$\begin{aligned}
\hat{w}_{\parallel}(x_1, y_1, x_1, y_1, 0) &= -|\hat{v}(x_1, y_1)|^2 = -\hat{k}(x_1, y_1), \\
\hat{w}_{\parallel}(x_2, y_2, x_2, y_2, 0) &= -|\hat{v}(x_2, y_2)|^2 = -\hat{k}(x_2, y_2), \\
\hat{w}_{\parallel}(x_1, y_1, x_2, y_2, s) + \hat{w}_{\parallel}(x_2, y_2, x_1, y_1, -s) &= -2\,\mathrm{Re}\big\{ \hat{v}^*(x_1, y_1) \hat{v}(x_2, y_2) \exp(\mathrm{i}\,\hat{\omega} s) \big\}.
\end{aligned}$$

These equations provide information about the per-mode wake functions for any s, *without* the $s > L$ restriction imposed in Section 3.2; they are derived in Appendix B by another method. In particular, we find that for $x_1 = x_2$ and $y_1 = y_2$,

$$\hat{w}_{\parallel}(x_1, y_1, x_1, y_1, s) + \hat{w}_{\parallel}(x_1, y_1, x_1, y_1, -s) = -2\hat{k}(x_1, y_1) \cos(\hat{\omega} s/c).$$

It seems natural to claim causality for the individual mode functions, so that we get the *fundamental theorem of beam loading*:

$$\hat{w}_{\parallel}(x_1, y_1, x_1, y_1, s) = -\hat{k}(x_1, y_1) \begin{cases} 0 & \text{for } s < 0, \\ 1 & \text{for } s = 0, \\ 2\cos(\hat{\omega} s/c) & \text{otherwise.} \end{cases}$$

Indeed, in Appendix B it is shown that for closed-cavity volumes,

$$w_{\parallel}(x_1, y_1, x_1, y_1, s > 0) = -2 \sum \hat{k}(x_1, y_1) \cos(\hat{\omega} s/c);$$

however, it is also found that the individual mode functions differ from the fundamental beam loading equation for $|s| < L$. Usually this discrepancy is no problem, as the *short-range* regime of interactions in the same bunch is quite distinct from the *long-range* regime of bunch-to-bunch interactions. The short-range wake is mostly calculated by time-domain methods without any distinction between $\hat{\omega} \neq 0$ and $\hat{\omega} = 0$ eigenmodes, whereas the long-range wake is calculated for $s > L$ from oscillating eigenmodes.

Formally the discrepancy can be solved by *definition*: the per-mode wake functions are causal, and the residual wake function is just the difference between the wake function and the summation over the so-defined per-mode functions. In the case of closed cavities, as considered in Appendix B, this leads to a residual wake that is zero.

Thus we come to an interesting consequence of the beam loading theorem: a source particle q_1 loses energy $q_1 V$ with $V = q_1 \hat{k}$, but a test particle that follows at a very short distance (and with the same offset) observes the voltage $-2V$.

3.5 Loss parameters

Loss parameters describe the loss of energy of a source particle or source distribution to electromagnetic field energy.

We have seen that the total energy loss of a point particle is given by the wake function for $x_1 = x_2$, $y_1 = y_2$ and $s = 0$, so that

$$k_{\mathrm{tot}} = -w(x_1, y_1, x_2, y_2, 0) = \mathcal{W}_{\mathrm{EM,total}}/q_1,$$

and we know that the loss to eigenmodes is given by the per-mode loss parameters

$$\hat{k} = |v(x_1, y_1)|^2.$$

The sum of all the per-mode loss parameters converges for cavities *with* beam pipes to a value below the total loss parameter k_{tot}, as not only are there modes excited, but also field energy is scattered and propagates along the beam pipes. (The wake of a closed cavity is completely determined by oscillating modes, but the sum is divergent.)

The wake potential (of distributed sources) and the shape-dependent loss parameter are usually calculated directly using electromagnetic time-domain solvers. The shape-dependent total loss parameter is the convolution of the longitudinal wake potential with the charge density function; for instance, for bunches with longitudinal profile $\lambda(z,t) = \lambda(z - ct)$ and negligible transverse dimensions,

$$k_{\mathrm{tot},\sigma} = -\int_{-\infty}^{\infty} W(x_1, y_1, x_1, y_1, z)\lambda(z)\,\mathrm{d}z.$$

The excitation of eigenmodes by distributed sources was discussed in Section 3.2; the shape-dependent per-mode loss parameter for a thin Gaussian bunch is

$$\hat{k}_\sigma = \hat{k}\exp\bigl(-(\hat{\omega}\sigma/c)^2\bigr),$$

and for a general longitudinal profile λ it is

$$\hat{k}_\sigma = \hat{k}\left|\int_{-\infty}^{\infty}\lambda(z)\cos(\hat{\omega}z/c)\,\mathrm{d}z\right|^2.$$

For closed cavities, the sum of the per-mode loss parameters $\hat{k}_\sigma$ converges to the total loss parameter $k_{\mathrm{tot},\sigma}$. This is also true for long bunches with $\sigma \gg c/\omega_{\mathrm{cutoff}}$, which cannot excite frequencies above the lowest cutoff frequency $\omega_{\mathrm{cutoff}} \propto \pi c/a$ of the beam pipes, where a is the characteristic transverse dimension of the pipes.

For the extreme case of ultra-short bunches it is difficult to calculate the wake potential, as a very high spatial resolution is required. In this case, only a small fraction of energy is lost to resonant modes and only a small part of the wakes is caused by resonances.

3.6 Analytical calculation for a pillbox

For a closed pillbox cavity, all modes can be calculated analytically [21]. Consider a pillbox with radius R and gap length g.

The normalized electromagnetic field of the monopole modes indexed by (n,p) is given by

$$\hat{E}_z^{(n,p)} = \frac{j_n}{R}\,J_0\left(j_n\,\frac{r}{R}\right)\cos\left(\frac{\pi\,p\,z}{g}\right)\exp(\mathrm{i}\,\hat{\omega}_{n,p}\,t),$$
$$\hat{E}_r^{(n,p)} = \frac{\pi p}{g}\,J_1\left(j_n\,\frac{r}{R}\right)\sin\left(\frac{\pi\,p\,z}{g}\right)\exp(\mathrm{i}\,\hat{\omega}_{n,p}\,t),$$

$$\hat{H}_{\phi}^{(n,p)} = \mathrm{i}\,\hat{\omega}_{n,p}\,\epsilon_0\,J_1\left(j_n\,\frac{r}{R}\right)\,\cos\left(\frac{\pi\,p\,z}{g}\right)\,\exp(\mathrm{i}\,\hat{\omega}_{n,p}\,t),$$

where j_n is the nth zero of the Bessel function $J_0(x)$ and

$$\left(\frac{\hat{\omega}_{n,p}}{c}\right)^2 = \left(\frac{j_n}{R}\right)^2 + \left(\frac{\pi p}{g}\right)^2.$$

Note that we now write the mode index explicitly as a double index (n,p). The stored energy is given by

$$\begin{aligned}\hat{\mathcal{W}}_{n,p} &= \frac{\mu_0}{2}\int_0^R \mathrm{d}r\,r\int_0^{2\pi}\mathrm{d}\varphi\int_0^g \mathrm{d}z\,\hat{H}_{\varphi}^{(n,p)}\left(\hat{H}_{\varphi}^{(n,p)}\right)^* \\ &= \frac{\pi\epsilon_0}{4}\left(\frac{\hat{\omega}_{n,p}}{c}\right)^2 g\,R^2\,J_1^2(j_n).\end{aligned}$$

Hence the normalized voltage on the axis becomes

$$\begin{aligned}\hat{v}_{n,p} &= \frac{1}{2\sqrt{\hat{\mathcal{W}}_{n,p}}}\int_z^g \mathrm{d}z\,E_z(r=0,z,t=z/c) \\ &= \frac{\mathrm{i}}{\sqrt{\pi\epsilon_0 g}}\frac{1-(-1)^p\,\exp(\mathrm{i}\,\hat{\omega}_{n,p}\,g/c)}{j_n\,J_1(j_n)},\end{aligned}$$

and the loss parameters can now be calculated as

$$\hat{k}_{n,p} = |\hat{v}_{n,p}|^2 = \frac{2}{\pi\,\epsilon_0\,g}\,\frac{1-(-1)^p\,\cos(\hat{\omega}_{n,p}\,g/c)}{j_n^2\,J_1^2(j_n)},$$

so that the expression for the monopole wake function becomes

$$w_{\parallel}(s) = -2\sum_{n=1}^{\infty}\sum_{p=0}^{\infty}\hat{k}_{n,p}\cos(\hat{\omega}_{n,p}\,s/c).$$

The wake function is the sum of all the voltages induced in all the modes.

For a Gaussian bunch with charge density

$$\rho(\boldsymbol{r},t) = q_1\,\delta(x)\delta(y)\lambda(z-ct), \qquad \lambda(s) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{s^2}{2\,\sigma^2}\right),$$

the wake potential is given by

$$W_{\parallel}(s) = \int_{-\infty}^{\infty}\lambda(s-s')\,w_{\parallel}(s')\,\mathrm{d}s'.$$

But even for such a simple example as a pillbox cavity, it is very hard to compute the wake potential by a modal analysis since many modes are needed. The reason is that for s inside the bunch, the charge distribution contributing to the wake potential is cut off because of causality. Hence the Fourier spectrum of the charge distribution contains many (say $>$1000) modes. Since an accelerator is not made up of closed boxes, the modal analysis is not sufficient for calculating the wake potential. The continuous spectrum of waveguide modes in the beam pipes contributes also to the impedance, especially to the short-range wake.

4 Impedances

4.1 Definitions

The Fourier transform of the negative[3] wake function is called the impedance or coupling-impedance:

$$Z_{\parallel}(x_1, y_1, x_2, y_2, \omega) = -\frac{1}{c}\int_{-\infty}^{\infty} \mathrm{d}s\, w_{\parallel}(x_1, y_1, x_2, y_2, s)\exp(-\mathrm{i}\omega s/c).$$

The wake function and impedance are two descriptions of the same thing, namely the coupling between the beam and its environment. The wake function is the time-domain description, while the impedance is the frequency-domain description:

$$w_{\parallel}(x_1, y_1, x_2, y_2, s) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, Z_{\parallel}(x_1, y_1, x_2, y_2, \omega)\exp(\mathrm{i}\omega s/c).$$

The reason for the usefulness of the impedance is that it often contains a number of sharply defined frequencies corresponding to the modes of the cavity or the long-range part of the wake. Figure 11 shows the real part of the impedance for a cavity. Below the cutoff frequency of the beam pipe, there is a sharp peak for each cavity mode. The spectrum above the cutoff frequency is continuous, caused by residual fields (not related to eigenmodes) and by the 'turn-on' of the harmonic eigen-oscillations. The continuous part of the spectrum is important for short-range wakes, especially for very short bunches.

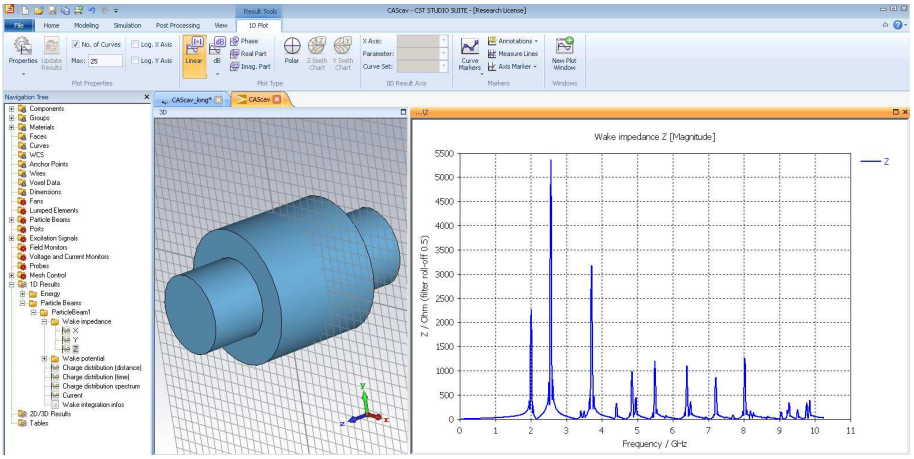


Fig. 11: Real part of the impedance for a cavity with side pipes; the peaks correspond to cavity modes. The results were obtained with the CST wakefield solver.

For the transverse impedance, it is often convenient to use a definition containing an extra factor i:

$$\boldsymbol{Z}_{\perp}(x_1, y_1, x_2, y_2, \omega) = \frac{\mathrm{i}}{c}\int_{-\infty}^{\infty} \mathrm{d}s\, \boldsymbol{w}_{\perp}(x_1, y_1, x_2, y_2, s)\exp(-\mathrm{i}\omega s/c).$$

The reason is that the transverse–longitudinal relations due to the Panofsky–Wenzel theorem then read as follows in the frequency domain:

$$\frac{\omega}{c}\,\boldsymbol{Z}_{\perp}(x_1, y_1, x_2, y_2, \omega) = \left(\boldsymbol{e}_x\frac{\partial}{\partial x_2} + \boldsymbol{e}_y\frac{\partial}{\partial y_2}\right) Z_{\parallel}(x_1, y_1, x_2, y_2, \omega).$$

[3]The sign is chosen so as to obtain a non-negative real part for $x_1 = x_2$ and $y_1 = y_2$.

4.2 Some properties of impedances and wakes

In the spatial s-domain, the relationship between the wake potential of a line charge density $\lambda(z-ct)$ and the wake functions of a point particle is described by the convolution

$$W(x_1,y_1,x_2,y_2,s)=\int_{-\infty}^{\infty}\mathrm{d}z\; w(x_1,y_1,x_2,y_2,s+z)\lambda(z).$$

The corresponding equation in the frequency domain for the Fourier transform of the negative wake potential is $V(x_1,y_1,x_2,y_2,\omega)=Z_\parallel(x_1,y_1,x_2,y_2,\omega)I(\omega)$, where

$$I(\omega)=\int_{-\infty}^{\infty}i(t)\exp(-\mathrm{i}\,\omega t)\,\mathrm{d}t=\int_{-\infty}^{\infty}c\lambda(-ct)\exp(-\mathrm{i}\,\omega t)\,\mathrm{d}t$$

is the beam current in the frequency domain. The energy loss of the bunch to electromagnetic fields,

$$\begin{aligned}\mathcal{W}_{\text{loss}}&=\int_{-\infty}^{\infty}W(x_1,y_1,x_1,y_1,s)\lambda(-s)\,\mathrm{d}s\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}V(x_1,y_1,x_1,y_1,\omega)I(\omega)^*\,\mathrm{d}\omega\\&=\frac{1}{\pi}\int_{0}^{\infty}\mathrm{Re}\{Z(x_1,y_1,x_1,y_1,\omega)\}|I(\omega)|^2\,\mathrm{d}\omega,\end{aligned}$$

has to be non-negative for any bunch shape λ. Therefore the real part of the longitudinal impedance must be non-negative for all offsets with $x_1=x_2$ and $y_1=y_2$. The real part can be negative for, say, $x_1=-x_2$ and $y_1=-y_2$, for a structure with azimuthal symmetry and frequency close to a dipole resonance.

As the wake potential is a real function, the real part of the impedance is an even function of the frequency while the imaginary part is an odd function of it:

$$\mathrm{Re}\{Z_\parallel(\ldots,\omega)\}=\mathrm{Re}\{Z_\parallel(\ldots,-\omega)\},\qquad \mathrm{Im}\{Z_\parallel(\ldots,\omega)\}=-\mathrm{Im}\{Z_\parallel(\ldots,-\omega)\}.$$

Hence, the wake function is given in terms of the impedance as

$$\begin{aligned}w_\parallel(\ldots,s)&=-\frac{1}{2\,\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\; Z_\parallel(\ldots,\omega)\exp(\mathrm{i}\,\omega s/c)\\&=-\frac{1}{2\,\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\big(\mathrm{Re}\{Z_\parallel(\ldots,\omega)\}\cos(\omega s/c)-\mathrm{Im}\{Z_\parallel(\ldots,\omega)\}\sin(\omega s/c)\big).\end{aligned}$$

Furthermore, the electromagnetic field ahead of the source particle is zero for $v=c$, as electromagnetic waves cannot overtake the source. Therefore, the wake function is *causal*, and the real and imaginary parts of the impedance are dependent on each other. From $w_\parallel(\ldots,s<0)=0$ it follows that for $u=-s>0$,

$$\int_{-\infty}^{\infty}\mathrm{d}\omega\,\mathrm{Re}\{Z_\parallel(\ldots,\omega)\}\cos(\omega u/c)=-\int_{-\infty}^{\infty}\mathrm{d}\omega\,\mathrm{Im}\{Z_\parallel(\ldots,\omega)\}\sin(\omega u/c),$$

so only the real (or imaginary) part of the impedance is really needed:

$$w_\parallel(\ldots,s>0)=\frac{1}{\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\mathrm{Re}\{Z_\parallel(\ldots,\omega)\}\cos(\omega s/c).$$

4.3 Shunt impedance and quality factor

The modal part of the wake function is

$$\hat{\boldsymbol{w}}(x_1, y_1, x_2, y_2, s) = -2\,\mathrm{Re}\big\{\hat{v}^*(x_1, y_1)\hat{\boldsymbol{v}}(x_2, y_2)\exp(\mathrm{i}\,\hat{\omega}s/c)\big\}\begin{cases}0 & \text{for } s < 0,\\ 1 & \text{for } s = 0,\\ 2 & \text{otherwise.}\end{cases}$$

We are interested in the longitudinal component on the axis ($x_1 = y_1 = x_2 = y_2 = 0$), and for simplicity we omit the transverse coordinates, so

$$\hat{w}_\parallel(s) = -2\hat{k}\cos(\hat{\omega}s/c)\begin{cases}0 & \text{for } s < 0,\\ 1 & \text{for } s = 0,\\ 2 & \text{otherwise.}\end{cases}$$

Then the impedance per mode,

$$\hat{Z}_\parallel(\omega) = -\frac{1}{c}\int_{-\infty}^{\infty}\hat{w}_\parallel(s)\exp(-\mathrm{i}\,\omega s/c)\,\mathrm{d}s,$$

is calculated as

$$\hat{Z}_\parallel(\omega) = 2\hat{k}\left\{\pi\delta(\omega+\hat{\omega}) + \pi\delta(\omega-\hat{\omega}) + \frac{\mathrm{i}\,\omega}{\hat{\omega}^2-\omega^2}\right\}.$$

This is equivalent to the impedance of a parallel resonant circuit (see Fig. 12),

$$\hat{Z}_\parallel(\omega) = \lim_{\hat{R}\to\infty}\left(\mathrm{i}\,\omega\hat{C} + \frac{1}{\mathrm{i}\,\omega\hat{L}} + \frac{1}{\hat{R}}\right)^{-1}$$

with $\hat{C} = 1/(2\hat{k})$, $\hat{L} = 2\hat{k}/\hat{\omega}^2$ and $\hat{R}\to\infty$.

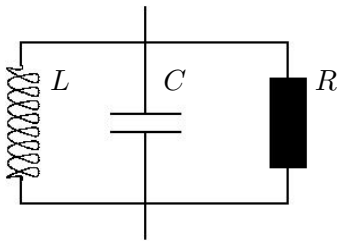


Fig. 12: Equivalent circuit model of the impedance of one mode

Although the resistor $\hat{R}$ was introduced for obvious formal reasons, it is helpful to consider weak losses of a resonator with a high quality factor $\hat{Q} = \hat{R}/(\hat{\omega}\hat{L})$. The impedance per mode of a resonator with weak losses is

$$\hat{Z}_\parallel(\omega) = 2\hat{k}\,\frac{\mathrm{i}\,\omega}{\hat{\omega}^2-\omega^2+\mathrm{i}\,\omega\hat{\omega}/\hat{Q}}.$$

The resistor $R = \hat{Z}_\parallel(\hat{\omega}) = 2\hat{k}\hat{Q}/\hat{\omega}$ is called the *shunt impedance*. The *quality factor* $\hat{Q}$ describes the decay time

$$\hat{\tau} = 2\hat{Q}/\hat{\omega},$$

the resonance bandwidth

$$\Delta\hat{\omega} = \omega/\hat{Q},$$

with $|\hat{Z}_\parallel(\hat{\omega})/\hat{Z}_\parallel(\hat{\omega} \pm \Delta\hat{\omega}/2)|^2 \approx 2$, and the energy loss per unit time

$$\hat{P} = \hat{\omega}\hat{\mathcal{W}}/\hat{Q}.$$

The last relation is used to determine the quality factor by perturbation theory: the energy loss (without beam) is caused by wall losses; as a good approximation these can be calculated from the fields obtained for the mode with infinite conductivity. If $\hat{H}_\mathrm{t}$ is the magnetic field tangential to the surface, the total power dissipated into the wall is given by a surface integral

$$\hat{P} = \frac{1}{2}\int_{\partial V} \mathrm{Re}\{\hat{\boldsymbol{E}}_\mathrm{s} \times \hat{\boldsymbol{H}}^*\} \cdot \mathrm{d}\boldsymbol{A} = \frac{1}{2}\int_{\partial V} \sqrt{\frac{\hat{\omega}\mu}{2\kappa}}\, |\hat{\boldsymbol{H}}|^2 \,\mathrm{d}A$$

where $\hat{\boldsymbol{E}}_\mathrm{s} = Z_\mathrm{s}\boldsymbol{n} \times \hat{H}$ is the tangential component of the electric field on the surface, with surface impedance $Z_\mathrm{s} = \sqrt{\mathrm{i}\,\hat{\omega}\mu/\kappa}$ for conductivity κ.

Appendix A: Eigenmodes of a closed cavity

We consider a (simply connected) cavity volume V_c, with perfectly conducting walls (boundary ∂V_c) and without current density. We search for time-harmonic eigensolutions, which can be written as

$$\boldsymbol{E}(\boldsymbol{r}, t) = \hat{\boldsymbol{E}}(\boldsymbol{r})\cos(\hat{\omega}t),$$
$$\boldsymbol{B}(\boldsymbol{r}, t) = \hat{\boldsymbol{B}}(\boldsymbol{r})\sin(\hat{\omega}t),$$

where $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{B}}$ are the eigenfields and $\hat{\omega}$ the (angular) eigenfrequencies. Substituting these into Maxwell's equations gives

$$\nabla\varepsilon\hat{\boldsymbol{E}} = \hat{\rho},$$
$$\nabla \times \hat{\boldsymbol{E}} = -\hat{\omega}\hat{\boldsymbol{B}},$$
$$\nabla\hat{\boldsymbol{B}} = 0,$$
$$\nabla \times \mu^{-1}\hat{\boldsymbol{B}} = -\hat{\omega}\varepsilon\hat{\boldsymbol{E}}.$$

We apply the operator $\nabla \times \mu^{-1}$ to the first curl equation and use the second curl equation to eliminate the magnetic flux density, thus obtaining the eigenproblem

$$\varepsilon^{-1}\nabla \times \mu^{-1}\nabla \times \hat{\boldsymbol{E}} = \hat{\lambda}\hat{\boldsymbol{E}},$$

with the eigenvalues $\hat{\lambda} = \hat{\omega}^2$ and the boundary condition $\boldsymbol{n} \times \hat{\boldsymbol{E}} = \boldsymbol{0}$. The operator $\varepsilon^{-1}\nabla\times\mu^{-1}\nabla\times$ is self-adjoint[4] with scalar product

$$\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \frac{1}{2}\int_{V_\mathrm{c}} \varepsilon\boldsymbol{A} \cdot \boldsymbol{B}\,\mathrm{d}V.$$

Therefore the problem has an infinite number of discrete real eigenvalues and a complete orthogonal system of eigenvectors,

$$\langle \hat{\boldsymbol{E}}_\xi, \hat{\boldsymbol{E}}_\tau\rangle = \hat{\mathcal{W}}_\xi\delta_{\xi\tau},$$

where $\hat{\mathcal{W}}_\xi$ is the electromagnetic field energy of mode ξ. The eigenvalues $\hat{\lambda}$ are non-negative so that all eigenfrequencies $\hat{\omega}$ are real.[5]

[4] The property $\langle\varepsilon^{-1}\nabla\times\mu^{-1}\nabla\times\boldsymbol{A}, \boldsymbol{B}\rangle = \langle\boldsymbol{A}, \varepsilon^{-1}\nabla\times\mu^{-1}\nabla\times\boldsymbol{B}\rangle$ can be shown with help of the identity $\nabla[\boldsymbol{A}\times\mu^{-1}\nabla\times\boldsymbol{B} - \boldsymbol{B}\times\mu^{-1}\nabla\times\boldsymbol{A}] = \boldsymbol{B}\times\nabla\times\mu^{-1}\nabla\times\boldsymbol{A} - \boldsymbol{A}\times\nabla\times\mu^{-1}\nabla\times\boldsymbol{B}$ and the divergence theorem. The left-hand side gives a surface integral that is zero because of the boundary conditions. The right-hand side corresponds to the assertion.

[5] This property can be shown by using the identity $\nabla[\hat{\boldsymbol{E}}\times\mu^{-1}\nabla\times\hat{\boldsymbol{E}}] = \mu^{-1}(\nabla\times\hat{\boldsymbol{E}})^2 - \hat{\boldsymbol{E}}\nabla\times\mu^{-1}\nabla\times\hat{\boldsymbol{E}}$ and the divergence theorem. The left-hand side gives a surface integral that is zero because of the boundary conditions. The volume integral of the first term on the right-hand side is non-negative; the integral of the second term gives $-2\hat{\lambda}\hat{\mathcal{W}}$. As $\hat{\mathcal{W}}$ is positive, $\hat{\lambda}$ cannot be negative.

There are obviously two types of eigensolutions:

$$\begin{aligned} \hat{\omega} &= 0, & \hat{\omega} &\neq 0, \\ \nabla\varepsilon\hat{\boldsymbol{E}} &\not\equiv 0, & \nabla\varepsilon\hat{\boldsymbol{E}} &= 0, \\ \nabla\times\hat{\boldsymbol{E}} &= 0, & \nabla\times\hat{\boldsymbol{E}} &= -\hat{\omega}\hat{\boldsymbol{B}}, \\ \hat{\boldsymbol{B}} &\equiv \boldsymbol{0}, & \hat{\boldsymbol{B}} &\not\equiv \boldsymbol{0}. \end{aligned}$$

Eigenfields for $\hat{\omega} = 0$ are curl-free and are just solutions to the electrostatic problem for any source distribution $\hat{\rho}$ and the boundary condition $\boldsymbol{n}\times\hat{\boldsymbol{E}} = 0$. Oscillating eigenfields are free of divergence; this is a consequence of Maxwell's second curl equation.

In Appendix B we use the property that any linear combination of eigensolutions with $\hat{\omega} = 0$ is orthogonal to any linear combination of oscillating eigenfields.

Appendix B: Wake of a closed cavity

We consider a (simply connected) cavity volume V_c of arbitrary shape, with perfectly conducting walls, that is located between the planes $z = 0$ and $z = L$. It is traversed by a point particle with charge q_1, offset (x_1, y_1) and velocity $v = c$. The stimulating charge and current density are

$$\begin{aligned} \rho(\boldsymbol{r}, t) &= q_1\delta(x - x_1)\delta(y - y_1)\delta(z - ct), \\ \boldsymbol{j}(\boldsymbol{r}, t) &= c\boldsymbol{e}_z\rho(\boldsymbol{r}, t). \end{aligned}$$

We use the complete orthogonal system of eigensolutions to describe the time-dependent electric field:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{\nu\in C}\hat{a}_\nu(t)\hat{\boldsymbol{E}}_\nu(\boldsymbol{r}),$$

where ν is the mode index, C is the set of all indexes and $\hat{a}_\nu(t)$ are the time-dependent coefficients. As in the main text, we shall write all mode-specific quantities with a hat and omit the index ν. We solve Maxwell's equations

$$\begin{aligned} \nabla\varepsilon\boldsymbol{E} &= \rho, \\ \nabla\times\boldsymbol{E} &= -\frac{\partial}{\partial t}\boldsymbol{B}, \\ \nabla\boldsymbol{B} &= 0, \\ \nabla\times\mu^{-1}\boldsymbol{B} &= \boldsymbol{J} + \varepsilon\frac{\partial}{\partial t}\boldsymbol{E} \end{aligned}$$

by applying the operator $\varepsilon^{-1}\nabla\times\mu^{-1}$ to the first curl equation and eliminating the magnetic flux density with the help of the second curl equation:

$$\varepsilon^{-1}\nabla\times\mu^{-1}\nabla\times\boldsymbol{E} = -\varepsilon^{-1}\frac{\partial}{\partial t}\boldsymbol{J} - \frac{\partial^2}{\partial t^2}\boldsymbol{E}.$$

By using the modal expansion and the eigenmode equation, we obtain

$$\sum_{\nu\in C}\hat{a}(t)\hat{\omega}^2\hat{\boldsymbol{E}} = -\varepsilon^{-1}\frac{\partial}{\partial t}\boldsymbol{J} - \frac{\partial^2}{\partial t^2}\sum_{\nu\in C}\hat{a}(t)\hat{\boldsymbol{E}}.$$

This set of scalar equations can be decoupled by applying the operator $\langle\hat{\boldsymbol{E}}_\xi, \cdots\rangle$ to both sides and using the orthogonality condition:

$$\hat{a}_\xi(t)\hat{\omega}_\xi^2\hat{\mathcal{W}}_\xi = -\varepsilon^{-1}\frac{\partial}{\partial t}\langle\hat{\boldsymbol{E}}_\xi, \boldsymbol{J}\rangle - \frac{\partial^2}{\partial t^2}\hat{a}_\xi(t)\hat{\mathcal{W}}_\xi.$$

Finally, we substitute the dirac current density and suppress the index, to arrive at

$$\left(\hat{\omega}^2 + \frac{\partial^2}{\partial t^2}\right) \hat{a}(t) = \frac{-1}{\hat{\mathcal{W}}_\mathcal{E}} \frac{\partial}{\partial t} \langle \hat{\boldsymbol{E}}, \boldsymbol{J} \rangle = -\frac{c q_1}{2\hat{\mathcal{W}}} \frac{\partial}{\partial t} \hat{E}(x_1, y_1, ct).$$

This ordinary differential equation can be solved[6] to give

$$\hat{a}(t) = \frac{-q}{\sqrt{\hat{\mathcal{W}}}} \operatorname{Re}\{\hat{v}^*(x_1, y_1, ct) \exp(\mathrm{i}\,\hat{\omega} t)\}$$

with

$$v(x, y, z) = \frac{1}{2\sqrt{\hat{\mathcal{W}}}} \int_{-\infty}^{z} \hat{E}_z(x, y, s) \exp(\mathrm{i}\,\hat{\omega} s/c)\,\mathrm{d}s$$

and

$$\frac{\partial}{\partial z} v(x, y, z) = \frac{1}{2\sqrt{\hat{\mathcal{W}}}} \hat{E}_z(x, y, z) \exp(\mathrm{i}\,\hat{\omega} z/c).$$

The longitudinal wake function is the sum over all modes,

$$w_\parallel(x_1, y_1, x_2, y_2, s) = \sum_{\nu \in C} \hat{w}_\parallel(x_1, y_1, x_2, y_2, s),$$

with the 'per-mode' contributions

$$\begin{aligned}
\hat{w}_\parallel(x_1, y_1, x_2, y_2, s) &= \frac{1}{q_1} \int_{-\infty}^{\infty} \hat{a}\left(\frac{z+s}{c}\right) \hat{E}_z(x_2, y_2, z)\,\mathrm{d}z \\
&= \frac{-1}{\sqrt{\hat{\mathcal{W}}}} \int_{-\infty}^{\infty} \operatorname{Re}\left\{\hat{v}^*(x_1, y_1, z+s) \exp\left(\mathrm{i}\,\hat{\omega}\frac{z+s}{c}\right)\right\} \hat{E}_z(x_2, y_2, z)\,\mathrm{d}z \\
&= -2 \operatorname{Re}\left\{\exp(\mathrm{i}\,\hat{\omega} s/c) \int_{-\infty}^{\infty} \hat{v}^*(x_1, y_1, z+s) \frac{\partial}{\partial z} v(x_2, y_2, z)\,\mathrm{d}z\right\}.
\end{aligned}$$

None of these terms is causal, i.e. $\hat{w}_\parallel(x_1, y_1, x_2, y_2, s < 0) \not\equiv 0$, but the sum has to be! In the following we use causality to find the simplified representation of the longitudinal wake function

$$w_\parallel(x_1, y_1, x_1, y_1, s > 0) = -2 \sum_{\hat{\omega} \neq 0} \hat{k}(x_1, y_1) \cos(\hat{\omega} s/c)$$

for $x_1 = x_2$ and $y_1 = y_2$, where $\hat{k}(x_1, y_1)$ is the longitudinal per-mode loss parameter, as defined in the main text. We therefore split the summation over all modes into the components

$$\begin{aligned}
w_{\parallel \mathrm{d}}(x_1, y_1, x_2, y_2, s) &= \sum_{\hat{\omega} = 0} \hat{w}_\parallel(x_1, y_1, x_2, y_2, s), \\
w_{\parallel \mathrm{c}}(x_1, y_1, x_2, y_2, s) &= \sum_{\hat{\omega} \neq 0} \hat{w}_\parallel(x_1, y_1, x_2, y_2, s)
\end{aligned}$$

and use the causality relation

$$w_{\parallel \mathrm{d}}(x_1, y_1, x_2, y_2, s < 0) + w_{\parallel \mathrm{c}}(x_1, y_1, x_2, y_2, s < 0) = 0$$

together with the anti-symmetry of the non-resonant part,

$$w_{\parallel \mathrm{d}}(x_1, y_1, x_2, y_2, s) = -w_{\parallel \mathrm{d}}(x_2, y_2, x_1, y_1, -s)$$

[6] The causal solution of $\ddot{a} + \omega^2 a = \dot{b}$ is $a(t) = \operatorname{Re}\{\int_{-\infty}^{t} b(\tau) \exp(\mathrm{i}\,\omega(t-\tau))\,\mathrm{d}\tau\}$.

proved in (A) below, to eliminate $w_{\parallel \mathrm{d}}$, yielding

$$w_\parallel(x_1, y_1, x_2, y_2, s > 0) = w_{\parallel \mathrm{c}}(x_1, y_1, x_2, y_2, s) + w_{\parallel \mathrm{c}}(x_2, y_2, x_1, y_1, -s).$$

To get the simplified representation for $x_1 = x_2$ and $y_1 = y_2$, we have to show that the condition

$$\text{(B)} \qquad \hat{w}_\parallel(x_1, y_1, x_1, y_1, s) + \hat{w}_\parallel(x_1, y_1, x_1, y_1, -s) = -2\hat{k}(x_1, y_1)\cos(\hat{\omega}s/c)$$

is fulfilled for eigenmodes with $\hat{\omega} \neq 0$.

We will now prove (A) and (B).

(A) For *non-oscillating modes* ($\hat{\omega} = 0$), the normalized voltage integrals $\hat{v}(x, y, z)$ are real and the contribution per mode is

$$\hat{w}_\parallel(x_1, y_1, x_2, y_2, s) = -2\int_{-\infty}^{\infty} \hat{v}(x_1, y_1, z + s)\frac{\partial}{\partial z}\hat{v}(x_2, y_2, z)\,\mathrm{d}z.$$

Therefore the required symmetry is fulfilled:

$$\begin{aligned}
\hat{w}_\parallel(x_1, y_1, x_2, y_2, s) &= -2\int_{-\infty}^{\infty} \hat{v}(x_1, y_1, z + s)\frac{\partial}{\partial z}\hat{v}(x_2, y_2, z)\,\mathrm{d}z \\
&= 2\int_{-\infty}^{\infty} \hat{v}(x_2, y_2, z)\frac{\partial}{\partial z}\hat{v}(x_1, y_1, z + s)\,\mathrm{d}z \\
&= 2\int_{-\infty}^{\infty} \hat{v}(x_2, y_2, z - s)\frac{\partial}{\partial z}\hat{v}(x_1, y_1, z)\,\mathrm{d}z = -\hat{w}_\parallel(x_2, y_2, x_1, y_1, -s).
\end{aligned}$$

The physical meaning of this symmetry is that the energy transfer from particle 1 to particle 2 (by $\hat{w}_\parallel(x_1, y_1, x_2, y_2, s)$) plus the reverse energy transfer (by $\hat{w}_\parallel(x_2, y_2, x_1, y_1, -s)$) is zero. This is obvious as no energy is left to the non-resonant mode after both particles have departed the volume. The voltage $\hat{v}(x, y, z)$ is zero for $z < 0$ before the source q_1 entered the cavity and, as the eigensolution is curl-free, it is zero for $z \geq L$. Therefore $\hat{w}_\parallel(\cdots, s) = 0$ for $|s| > L$. Two particles can interact only through non-oscillating modes if they are simultaneously in the cavity at any time.

(B) The normalized voltage integral for *oscillating modes* ($\omega \neq 0$) does not depend on z after q_1 has left the cavity:

$$v(x, y, z > L) = \frac{1}{2\sqrt{\hat{\mathcal{W}}}}\int_{-\infty}^{L} \hat{E}_z(x, y, s)\exp(\mathrm{i}\,\hat{\omega}s/c)\,\mathrm{d}s = v(x, y).$$

Therefore the following integral relation can be derived:

$$\begin{aligned}
\hat{v}^*(x_1, y_1)\hat{v}(x_2, y_2) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial z}\left\{\hat{v}^*(x_1, y_1, z + s)\hat{v}(x_2, y_2, z)\right\}\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \hat{v}(\cdots_2, z)\frac{\partial}{\partial z}\hat{v}^*(\cdots_1, z + s)\,\mathrm{d}z + \int_{-\infty}^{\infty} \hat{v}^*(\cdots_1, z + s)\frac{\partial}{\partial z}\hat{v}(\cdots_2, z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \hat{v}(\cdots_2, z - s)\frac{\partial}{\partial z}\hat{v}^*(\cdots_1, z)\,\mathrm{d}z + \int_{-\infty}^{\infty} \hat{v}^*(\cdots_1, z + s)\frac{\partial}{\partial z}\hat{v}(\cdots_2, z)\,\mathrm{d}z.
\end{aligned}$$

This relation is needed to prove the symmetry:

$$\begin{aligned}
&\hat{w}(x_1, y_1, x_2, y_2, s) \\
&\quad = -2\,\mathrm{Re}\left\{\exp(\mathrm{i}\,\hat{\omega}s/c)\int_{-\infty}^{\infty} \hat{v}^*(x_1, y_1, z + s)\frac{\partial}{\partial z}\hat{v}(x_2, y_2, z)\,\mathrm{d}z\right\}
\end{aligned}$$

$$= -2\,\mathrm{Re}\left\{\exp(\mathrm{i}\,\hat{\omega}s/c)\left[\hat{v}^*(x_1,y_1)\hat{v}(x_2,y_2) - \int_{-\infty}^{\infty}\hat{v}^*(\cdots_1, z+s)\frac{\partial}{\partial z}\hat{v}(\cdots_2, z)\,\mathrm{d}z\right]\right\}$$
$$= -2\,\mathrm{Re}\big\{\exp(\mathrm{i}\,\hat{\omega}s/c)\hat{v}^*(x_1,y_1)\hat{v}(x_2,y_2)\big\} - \hat{w}(x_2,y_2,x_1,y_1,-s).$$

With $x_1 = x_2$ and $y_1 = y_2$, we find that

$$\hat{w}(x_1,y_1,x_1,y_1,s) + \hat{w}(x_1,y_1,x_1,y_1,-s) = -2\hat{k}(x_1,y_1)\cos(\hat{\omega}s/c),$$

where $\hat{k}(x_1,y_1) = |\hat{v}(x_1,y_1)|^2$; in particular, for the origin,

$$\hat{w}(x_1,y_1,x_1,y_1,0) = -\hat{k}(x_1,y_1).$$

Acknowledgements

First and foremost we would like to thank the audience of our lecture for their interest. We are also grateful to the CAS team for their organizational efforts and patience.

References

[1] T. Weiland and R. Wanzenberg, Wakefields and impedances, Proceedings of US–CERN School (Hilton Head, 1990), edited by M. Dienes, M. Month and S. Turner (Springer, Berlin, 1992). http://dx.doi.org/10.1007/3-540-55250-2_26

[2] A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).

[3] B.W. Zotter and S.A. Kheifets, *Impedances and Wakes in High-Energy Particle Accelerators* (World Scientific, Singapore, 1998). http://dx.doi.org/10.1142/3068

[4] K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, Hackensack, 2005). http://dx.doi.org/10.1142/5835

[5] J.D. Jackson, *Classical Electrodynamics*, 2nd edition (Wiley, New York, 1975).

[6] O. Brüning, H. Burkhardt and S. Myers, *Prog. Part. Nucl. Phys.* **67** (2012) 705. http://dx.doi.org/10.1016/j.ppnp.2012.03.001

[7] *LEP Design Report. Vol. 2: The LEP Main Ring*, CERN-LEP-84-01 (1984).

[8] K. Balewski *et al.* (eds.), PETRA III: A Low Emittance Synchrotron Radiation Source, DESY 2004-035 (February 2004).

[9] *MAFIA User Guide*, CST – Computer Simulation Technology AG, Bad Nauheimer Str. 19, 64289 Darmstadt, Germany.

[10] *CST Studio Suite*, CST – Computer Simulation Technology AG, Bad Nauheimer Str. 19, 64289 Darmstadt, Germany.

[11] T. Weiland, Transient electromagnetic fields excited by bunches of charged particles in cavities of arbitrary shape, Proceedings of the 11th International Conference on High-Energy Accelerators, Geneva, 1980 (Birkhäuser, Basel, 1980), pp. 570–575. http://dx.doi.org/10.1007/978-3-0348-5540-2_75

[12] T. Weiland, *Part. Accel.* **15** (1984) 245.

[13] A. Piwinski, Longitudinal and transverse wake fields in flat vacuum chambers, DESY 84/097 (October 1984).

[14] K. Steinigke, *Frequenz* **44** (1990), 4–8. http://dx.doi.org/10.1515/freq.1990.44.1.2

[15] A. Chao, Beam dynamics with high intensity, these proceedings.

[16] W.K.H. Panofsky and W.A. Wenzel, *Rev. Sci. Instrum.* **27** (1956), 967. http://dx.doi.org/10.1063/1.1715427

[17] I. Zagorodnov, K. Bane and G. Stupakov, *Phys. Rev. ST Accel. Beams* **18** (2015), 104401. http://dx.doi.org/10.1103/physrevstab.18.104401

[18] I. Zagorodnov and T. Weiland, *Phys. Rev. ST Accel. Beams* **8** (2005), 042001. http://dx.doi.org/10.1103/physrevstab.8.042001

[19] I. Zagorodnov, ECHO 2D, http://www.desy.de/~zagor/WakefieldCode_ECHOz/.

[20] T. Weiland and I. Zagorodnov, *The Short-Range Transverse Wake Function for TESLA Accelerating Structure*, TESLA Report 2003-19, DESY (2003).

[21] T. Weiland and B. Zotter, *Part. Accel.* **11** (1981), 143.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam Dynamics of Collective Instabilities in High-Energy Accelerators

A. Chao
SLAC National Accelerator Laboratory, Menlo Park, United States

Abstract
This lecture provides an abbreviated overview of the basic physics of collective instabilities of intense charged particle beams in high-energy accelerators.

Keywords
Collective instabilities, Gauss's law, impedance, Landau damping, Vlasov equation, wakefield

1 Gauss's law

Collective instabilities in accelerators mostly come from an intense charged particle beam electromagnetically interacting with its vacuum chamber environment. As the beam interacts with its environment, it generates an electromagnetic field called the *wakefield*, and the wakefield acts back on the beam, disturbing its motion; if the perturbation is strong enough, the beam becomes unstable.

To discuss the wakefield, we must start with its ultimate origin, *Gauss's law*, which states that each charged particle always has a definite amount of electric field lines attached to it. We can distort these field lines but we can never cut them loose from the charge under any circumstances. Furthermore, the amount of field lines attached to each charge can never be changed, neither increased nor decreased.

Gauss's law is amazing. Mathematically, it reads

$$\nabla \cdot \vec{E} = 4\pi\rho \, .$$

Physically it reads: electric field lines are absolutely attached to the charges.

The integral form of Gauss's law is

$$\oint_S \vec{E} \cdot \mathrm{d}\vec{S} = 4\pi Q \, ,$$

where Q is the total charge inside the volume enclosed by the surface S. It is amazing that this law holds no matter how the charges are moving – non-relativistically, relativistically, or under acceleration, or whether they are embedded in any type of material. It also does not matter how close the charges might be immediately next to the surface S. The field integral will make a sudden change when a charge crosses the surface even infinitesimally.

2 A moving charge

If the charge is stationary and if it is in a free space, its field lines are as shown in Fig. 1(a). For a moving charge, we see Fig. 1(b). When v approaches c, we have Fig. 1(c), when all the electric fields stay in an infinitely thin sheet as result of the theory of relativity. For most accelerators, case (c) is closest to the case under consideration.

When the charge is moving, it also generates a magnetic field. This magnetic field also contracts to a thin pancake when $v = c$. The direction of the electric field is radial; the direction of the magnetic field is azimuthal (right-hand rule):

$$E_r \quad = \quad \frac{2q}{r}\delta(z - ct) \, ,$$

https://doi.org/10.23730/CYRSP-2017-003.43

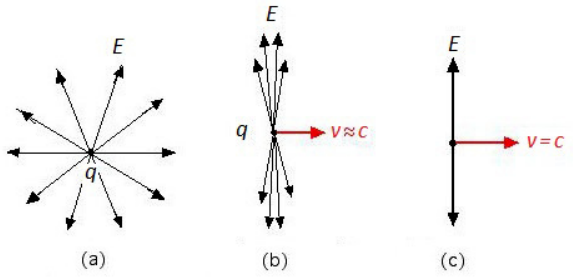


Fig. 1: Electric field lines of a charge: (a) stationary; (b) moving relativistically; (c) when $v = c$

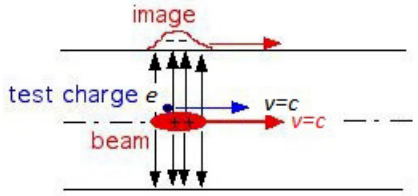


Fig. 2: Ultrarelativistic beam going down perfectly conducting smooth vacuum pipe

$$B_\theta = \frac{2q}{r}\delta(z - ct) \, .$$

One observes that

$$B_\theta = E_r \quad \text{when} \quad v = c \, .$$

However, when $v = 0$, there is no magnetic field. When v increases, B_θ increases, but is still weaker than E_r. Only when $v = c$ do we have $B_\theta = E_r$. The fact that $B_\theta = E_r$ when $v = c$ has important consequences, as explained next.

3 The vacuum chamber

We now add the vacuum chamber. Consider a very smooth vacuum chamber beam pipe. (How smooth does the chamber have to be? A 1 mm discontinuity on the pipe is considered a potential problem. In some circumstances, a 1 μm roughness on the wall surface can have a significant effect.) Consider the smooth pipe wall to be perfectly conducting.

The ultrarelativistic beam going down the axis of the pipe, together with its electromagnetic field and the smooth vacuum chamber, is as shown in Fig. 2.

The electromagnetic fields are perfectly and cleanly terminated on the pipe wall. No fields penetrate into the wall because it is a perfect conductor. The image charge on the wall is exactly equal and opposite to that of the beam, and it also moves with $v = c$ (except that this is phase velocity, not group velocity). The entire field pattern moves with the beam. There is no wakefield.

Is this beam stable? Consider a particular particle in the beam, the 'test particle' e in Fig. 2. This test particle will see an electric force $e\vec{E}$ due to the electric field carried by the beam. This force is easily seen to push e towards the vacuum chamber wall because the test charge e has the same sign as the charges of the beam.

But there is also a magnetic force. The magnetic field is in the azimuthal direction (right-hand rule). The magnetic force is $(e/c)\,\vec{v} \times \vec{B}$. It is easily seen that this magnetic force is pointing towards the pipe axis.

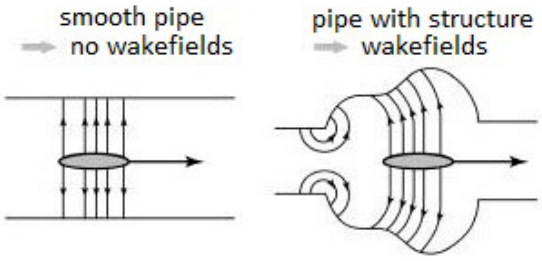


Fig. 3: Discontinuities generate wakefields

We mentioned that when $v = c$, we have $E_r = B_\theta$. In the ultrarelativistic limit, therefore, the electric and the magnetic forces exactly cancel. The particles in the ultrarelativistic beam see the electric force and magnetic force, but they do not see a net force. The collective electromagnetic fields carried by the beam do not influence particle motion. There are no collective instabilities.

This cancellation between the electric and magnetic forces due to the beam's self fields is very fortunate and very important. Without this cancellation, no modern accelerators would have worked.

We conclude that there are three possible ways for a collective instability to occur:

1. the beam is not relativistic enough;
2. the vacuum chamber is too resistive;
3. the vacuum chamber is not smooth enough.

If none of these apply, the beam is stable as just illustrated. If any one of these conditions occurs, the exact cancellation of the electric and magnetic forces is disrupted, and the beam can encounter an instability.

We construct accelerators to be as close to the cancellation condition as possible. The electric and magnetic forces generally cancel to high accuracy by design. However, the cancellation is never perfect. Vacuum chambers made of copper or aluminium are not perfectly conducting. There will be many small necessary discontinuities along the vacuum chamber pipe; beam position monitors, vacuum pumping ports, etc. There are also big discontinuities known as RF cavities. As to the condition of $v = c$, it is never satisfied completely. So the cancellation of electric and magnetic forces is not perfect. And that residual non-cancellation leads to collective instabilities.

4 Wakefields due to discontinuities

When a beam traverses a discontinuity, an electromagnetic wakefield is generated. An intense beam will generate a strong wakefield (Fig. 3). When the wakefield becomes too strong, the beam becomes unstable.

A wakefield is generated because the beam's image charges now have to move around a corner when encountering a discontinuity. Wakefields are the radiation fields of the image charges when their apparent trajectories are bent. (These are apparent trajectories. Image charges do not physically move along the wall surface.)

Once we accept that wakefields are a result of radiation, then just as with any other radiation, it is natural to ask about the frequency content of these wakefields. The answer is that it depends on the details of the beam and the detailed geometry of the discontinuity. In general, it covers a wide range, with wavelengths varying from micrometres to metres. To describe the frequency content of the wakefields, we introduce a quantity called *impedance*. Impedance is essentially the Fourier transform of wakefield.

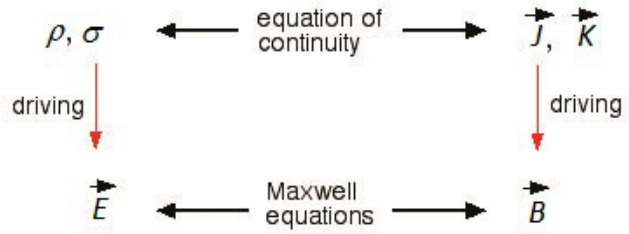


Fig. 4: Definition of metal: $\rho = 0$, $\vec{J} = \sigma\vec{E}$; definition of insulator: $\vec{J} = \vec{0}$, $\rho = \epsilon\nabla \cdot \vec{E}$

5 Wakefield due to a resistive wall

To discuss the resistive wall wakefield, let us first review the structure of electromagnetism by consulting Fig. 4. We note a clear symmetry between the electric family and the magnetic family in this chart. This symmetry, however, only holds in vacuum. It is lost when we consider a metal or an insulator. Metals break the symmetry by making a preference to the magnetic family $(\vec{B}, \vec{J})$, while insulators make a preference in favour of the electric family $(\vec{E}, \rho)$. No charges are allowed inside a metal while currents are allowed to penetrate. Inside a metal, therefore, there is more magnetic field than electric field. Conversely, currents are not allowed inside an insulator, and there is more electric field than magnetic field.

In the case of a wall of resistive metal, the wakefield is generated by the following physical process.

1. When the beam's image charges flow on the vacuum chamber wall, the electric field carried by the point charge will be terminated immediately by the image charges on the wall surface, while the magnetic field carried by the point charge is mostly cancelled by the image current on the wall surface, but this cancellation is not exact because the current has penetrated the wall by a skin depth.
2. As the image current slowly resurfaces after the point charge has passed by, this resurfacing image current drives new magnetic fields. These new magnetic fields occur after the point charge has left.
3. The resurfacing current and magnetic field will execute some transient behaviour, and quickly oscillate a few times. After the initial transient, the resurfacing current and magnetic field decays away but at a very slow rate.
4. The resurfacing changing magnetic field now drives an electric field (Maxwell's equation). This yields some electric field inside the resistive wall after all, but this electric field is very weak.

For the case of a resistive wall pipe with circular cross-section, and an ultrarelativistic point charge q going down its axis, Fig. 5 shows the electric component of the wakefield inside the vacuum chamber. Note that there is also a matching magnetic field pattern, and that both the electric and magnetic field patterns follow the leading point charge as a frozen pattern, indicating a phase velocity of c, but it is important to know that the wakefield energy flows not purely in the forward direction with the speed of light. Underneath this apparent frozen energy flow is an important flow of energy from the point charge q towards to wall surface to be deposited as wall heating.

The quantity χ that enters the horizontal axis is a small dimensionless parameter defined by

$$\chi = \frac{c}{4\pi\sigma b} ,$$

with b the vacuum pipe radius and σ the conductivity of the pipe material. For example, if $b = 5$ cm and the wall is made of aluminium, $\chi = 1.5 \times 10^{-9}$. Note that we have used cgs units, in which conductivity can be converted by applying $1\ \Omega^{-1}\mathrm{m}^{-1} = 0.9 \times 10^{10}\mathrm{s}^{-1}$.

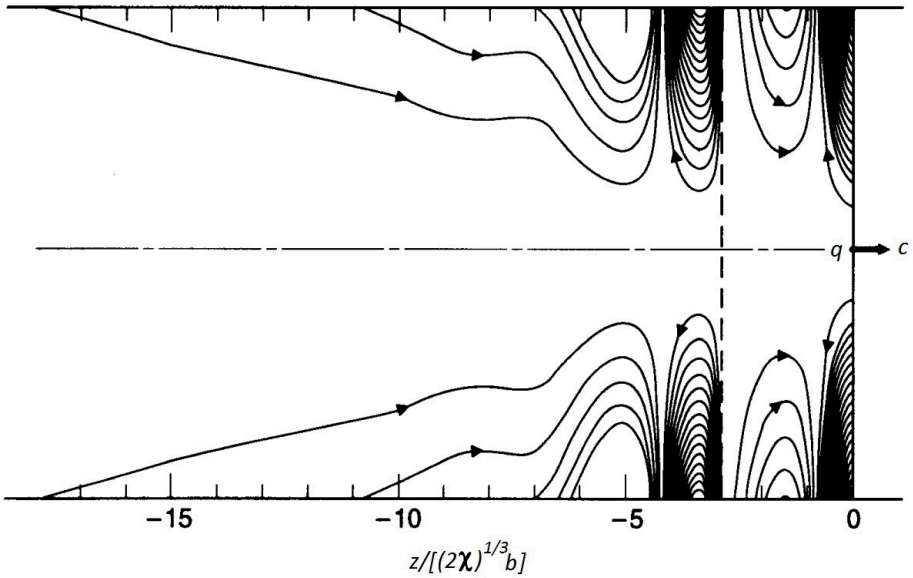


Fig. 5: Electric component of the wakefield inside the vacuum chamber for a resistive wall pipe with circular cross-section. The field line density is increased by a factor of 40 to the left of the dotted line.

As shown in Fig. 5, there is no wakefield ahead of the point charge, as causality would dictate. The wakefield pattern following the point charge is measured along distance z in units of $b(2\chi)^{1/3}$. Since $\chi \ll 1$, the resistive wall wakefield decays very quickly following the passage of the point charge.

However, after the quick initial decay, at long distances, the remaining resistive wall wake starts to decay very slowly. This means that the resistive wall wakefield has a long tail. An intense beam bunch, for example, can leave a wakefield that lasts long enough to affect its motion when the bunch returns after making one complete circuit around a circular accelerator.

As will be shown later, the fact that the resistive wall generates both short- and long-range wakefields is reflected by the fact that its corresponding impedance has an exceptionally wide spectrum, ranging from very short to very long wavelengths.

6 What happens to particle motion when there are wakefields?

Consider a beam with distribution ψ in phase space $(\vec{q}, \vec{p})$. The dynamics of the evolution of ψ are described by the Vlasov equation (see later),

$$\frac{\partial \psi}{\partial t} + \frac{\vec{p}}{m} \cdot \frac{\partial \psi}{\partial \vec{q}} + \vec{f} \cdot \frac{\partial \psi}{\partial \vec{p}} = 0 \,,$$

where

$$\vec{f} = e\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right) \,,$$
$$\vec{E} = \vec{E}_{\rm ext} + \vec{E}_{\rm wake} \,,$$
$$\vec{B} = \vec{B}_{\rm ext} + \vec{B}_{\rm wake} \,.$$

The wakefields are determined by Maxwell's equations, where the source terms ρ and $\vec{j}$ are determined by the beam distribution ψ,

$$\rho = \int \mathrm{d}^3 p \, \psi \,, \qquad \vec{j} = \int \mathrm{d}^3 p \, \vec{v} \psi \,.$$

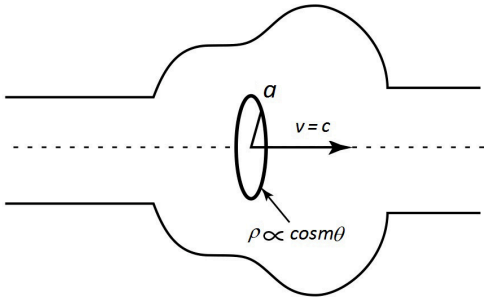


Fig. 6: We only need to calculate the wakefields generated by a rigid $\cos m\theta$ ring beam

We therefore have the situation when the beam distribution is described by the Vlasov equation whose force terms are given by the electromagnetic fields, while the electromagnetic fields are described by Maxwell's equations, whose source terms are given by the beam distribution. It is clear that a full treatment of the beam–wakefield system involves solving a coupled 'Vlasov–Maxwell equation'.

Beam–structure interaction is a difficult problem in general. It often requires numerical solution using particle-in-cell codes with demanding boundary conditions. Applying particle-in-cell codes is reasonable for small devices such as electron guns and klystrons, but becomes impractical for large accelerators.

So, can we simplify the problem for our purpose while maintaining sufficiently accurate results? Yes, we can. For *high-energy* accelerators, this complication can be avoided by using two simplifying approximations. These simplifications lead to the concepts of 'wake function' and 'impedance'.

1. Rigid-beam approximation:
   The first simplification is the rigid-beam approximation. At high energies, beam motion is affected little during the passage of a structure. This means that one can calculate the wakefields assuming the beam shape is rigid and its motion is ultrarelativistic with $v = c$. In fact, we only need to calculate the wakefields generated by a 'rigid $\cos m\theta$ ring beam' (Fig. 6). The wakefield of a general beam can be obtained by superposition.
2. Impulse approximation:
   The second simplification is the impulse approximation. We don't need to know $\vec{E}$ or $\vec{B}$ separately; we only need to know $\vec{f}$. For high energies, we don't even need the instantaneous $\vec{f}$. We only need the integrated impulse:
   $$\Delta\vec{p} = \int_{-\infty}^{\infty} \mathrm{d}t\ \vec{f}\,.$$
   Figure 7 shows the configuration of a ring beam and a test charge that follows it. The ring beam generates a wakefield. The test charge receives a wake-induced impulse in the impulse approximation.

As we will see, these two approximations drastically simplify the problem at hand, thus allowing us to treat large accelerators with complicated boundary conditions without invoking particle-in-cell codes.

7 The Panofsky–Wenzel theorem

The instantaneous wakefields are complicated; fortunately, $\Delta\vec{p}$ is much simpler and, at high energies, it is $\Delta\vec{p}$ that we need. The Panofsky–Wenzel theorem applies to $\Delta\vec{p}$. It is the basis of all beam instability analyses in high energy accelerators. In comparison, the particle-in-cell codes aim to calculate the instantaneous wakefields in all their gory details, so are inefficient for our purpose.

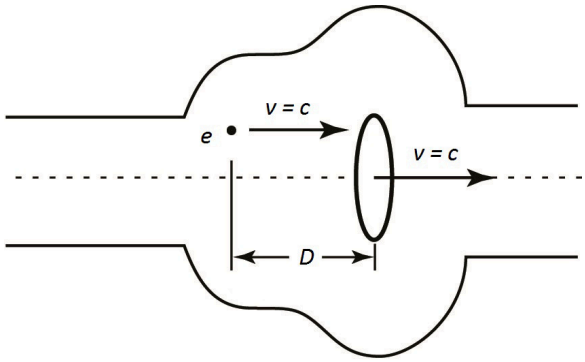


Fig. 7: A ring beam and a test charge that follows it

Maxwell's equations read

$$\begin{aligned} \nabla \cdot \vec{E} &= 4\pi\rho \,, \\ \nabla \times \vec{B} - \frac{1}{c}\frac{\partial \vec{E}}{\partial t} &= 4\pi\beta\rho\hat{z} \,, \\ \nabla \cdot \vec{B} &= 0 \,, \\ \nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} &= 0 \,, \end{aligned}$$

where we have made the important rigid-beam approximations, $\vec{j} = \rho\vec{v}$ and $\vec{v} = \beta c\hat{z}$.

The Lorentz force as seen by the rigid test charge e is given by

$$\vec{f} = e(\vec{E} + \beta\hat{z} \times \vec{B}) \,.$$

Both the beam and the test charge move with $\vec{v} = \beta c\hat{z}$. The impulse is

$$\Delta\vec{p}(x, y, D) = \int_{-\infty}^{\infty} \mathrm{d}t \; \vec{f}(x, y, D + \beta ct, t) \,.$$

Several important conditions can be found using Maxwell's equations. One of these is the Panofsky–Wenzel theorem. Without giving the derivation, it reads

$$\nabla \times \Delta\vec{p} = \vec{0} \,.$$

One can decompose the Panofsky–Wenzel theorem into a component parallel to $\hat{z}$ and a component perpendicular to $\hat{z}$, to obtain

$$\nabla \cdot (\hat{z} \times \Delta\vec{p}) = 0 \,, \tag{1}$$

$$\frac{\partial}{\partial D}\Delta\vec{p}_\perp = \nabla_\perp \Delta p_z \,. \tag{2}$$

Equation (1) says something about the transverse components of $\Delta\vec{p}$. Equation (2) says that the transverse gradient of the longitudinal wake impulse is equal to the longitudinal gradient of the transverse wake impulse.

Another important condition valid when $\beta = 1$ is

$$\nabla_\perp \cdot \Delta\vec{p}_\perp = 0 \,. \tag{3}$$

It is clear that the Panofsky–Wenzel theorem imposes strong constraints on the impulse received by a test charge from a relativistic beam.

8 Cylindrically symmetric pipe

In cylindrical coordinates, Eq. (1) gives

$$\begin{aligned} & \nabla \cdot [\hat{z} \times (\Delta p_r \hat{r} + \Delta p_\theta \hat{\theta})] = 0 \\ \Longrightarrow \qquad & \frac{\partial}{\partial r}(r \Delta p_\theta) = \frac{\partial}{\partial \theta} \Delta p_r \ . \end{aligned}$$

Equation (2) gives

$$\begin{aligned} & \frac{\partial}{\partial D}(\Delta p_r \hat{r} + \Delta p_\theta \hat{\theta}) = \left(\hat{r} \frac{\partial}{\partial r} + \frac{\hat{\theta}}{r} \frac{\partial}{\partial \theta} \right) \Delta p_z \\ \Longrightarrow \qquad & \begin{cases} \frac{\partial}{\partial D} \Delta p_r = \frac{\partial}{\partial r} \Delta p_z \\ \frac{\partial}{\partial D} \Delta p_\theta = \frac{1}{r} \frac{\partial}{\partial \theta} \Delta p_z \ . \end{cases} \end{aligned}$$

Equation (3) gives

$$\begin{aligned} & \frac{1}{r} \frac{\partial}{\partial r}(r \Delta p_r) + \frac{1}{r} \frac{\partial}{\partial \theta} \Delta p_\theta = 0 \\ \Longrightarrow \qquad & \frac{\partial}{\partial r}(r \Delta p_r) = -\frac{\partial}{\partial \theta} \Delta p_\theta \qquad (\beta = 1) \ . \end{aligned}$$

These results are surprisingly simple and general. They do not contain any beam source terms. The exact shape or distribution of the beam does not matter. Neither do these results depend on the boundary conditions. The boundary can be perfectly conducting or resistive metal, or it can be a dielectric. It does not have to be a sharply defined surface; it can, for example, be a gradually fading plasma surface. The only inputs needed are Maxwell's equations and the rigid-beam and impulse approximations.

We are now ready to consider a $\cos m\theta$ ring beam with $\vec{v} = c\hat{z}$ as we set out to solve Eqs. (1)–(3). The solution can be expressed in terms of a function $W_m(D)$, such that

$$\begin{aligned} c\Delta \vec{p}_\perp &= -eI_m W_m(D) m r^{m-1} (\hat{r} \cos m\theta - \hat{\theta} \sin m\theta) \ , \\ c\Delta p_z &= -eI_m W'_m(D) r^m \cos m\theta \ , \end{aligned} \tag{4}$$

where I_m is the mth multipole moment of the ring beam,$W_m(D)$ is the transverse wake function and $W'_m(D)$ is the longitudinal wake function. The longitudinal wake function is simply the derivative of the transverse wake function.

Equation (4) contains explicit dependences on r and θ. The fact that we can go so far without any specific details is surprising and shows the power of this line of analysis. The dependence on D is through $W_m(D)$, which can be obtained only if boundary conditions are invoked.

When the beam pipe is cylindrically symmetrical, each m-multipole component of the beam excites a wake pattern according to Eq. (4). Different m components do not mix.

9 Decomposing wakefields into modes

Armed with the Panofsky–Wenzel theorem, to analyse the instability problem, we proceed as follows. We first consider the beam to be a δ-function in z. If the beam has a finite length, the result can be obtained by superposition.

We next decompose the transverse distribution into 'modes' and consider a single transverse mode m. A general transverse distribution can be obtained by superposition with a summation over m.

So the problem is now reduced to finding the impulse integrated by a test charge that is trailing behind a beam slice with a transverse mth moment I_m moving along the pipe axis. In this configuration,

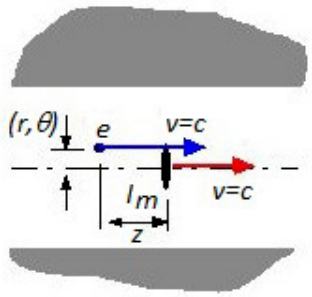


Fig. 8: A test charge trailing behind a beam slice with a transverse mth moment I_m moving along the pipe axis

Table 1: mth multipole wakefields

m	Distribution moments of beam	Longitudinal wake impulse	Transverse wake impulse
0	q	$-eq\,W_0'(z)$	0
1	$q\langle x\rangle$	$-eq\langle x\rangle x W_1'(z)$	$-eq\langle x\rangle W_1(z)\hat{x}$
	$q\langle y\rangle$	$-eq\langle y\rangle y W_1'(z)$	$-eq\langle y\rangle W_1(z)\hat{y}$
2	$q\langle x^2-y^2\rangle$	$-eq\langle x^2-y^2\rangle(x^2-y^2)W_2'(z)$	$-2eq\langle x^2-y^2\rangle W_2(z)(x\hat{x}-y\hat{y})$
	$q\langle 2xy\rangle$	$-eq\langle 2xy\rangle 2xy\,W_2'(z)$	$-2eq\langle 2xy\rangle W_2(z)(y\hat{x}+x\hat{y})$
3	$q\langle x^3-3xy^2\rangle$	$-eq\langle x^3-3xy^2\rangle \times(x^3-3xy^2)W_3'(z)$	$-3eq\langle x^3-3xy^2\rangle W_3(z) \times[(x^2-y^2)\hat{x}-2xy\hat{y}]$
	$q\langle 3x^2y-y^3\rangle$	$-eq\langle 3x^2y-y^3\rangle \times(3x^2y-y^3)W_3'(z)$	$-3eq\langle 3x^2y-y^3\rangle W_3(z) \times[2xy\hat{x}+(x^2-y^2)\hat{y}]$

as shown in Fig. 8, I_m is the driving beam, e is the test charge, z is the longitudinal distance by which e is trailing behind I_m, and (r,θ) is the transverse displacement of the test charge relative to the pipe axis.

For a cylindrical pipe, the mth multipole wakefield is driven when and only when the driving beam has an mth moment (Table 1).

In most applications, we care mostly about the $m=0$ monopole mode when discussing longitudinal collective instabilities and about the $m=1$ dipole mode when discussing transverse collective instabilities. Therefore, we mostly ask for $W_0(z)$ and $W_1(z)$. The reason $W_0(z)$ is not relevant for transverse instabilities is that the transverse impulse vanishes when $m=0$ for cylindrically symmetrical pipes.

The wakefield impulses have simple patterns—instantaneous wakefields do not share this simplicity. The $m=0$ and $m=1$ patterns are illustrated in Fig. 9.

10 Impedances

We mentioned that the wakefield wavelengths cover a wide range, from ∼1 μm to ∼1 m. What characterize the frequency content of the wakefields are the impedances, the Fourier transforms of the wake functions,

$$Z_m^{\parallel}(\omega)=\int_{-\infty}^{\infty}\frac{\mathrm{d}z}{c}\mathrm{e}^{-\mathrm{i}\omega z/c}\,W_m'(z)\,, \tag{5}$$

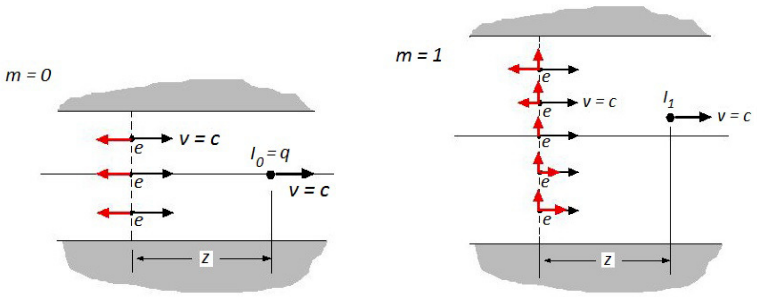


Fig. 9: $m = 0$ and $m = 1$ wakefield impulse patterns

$$Z_m^{\perp}(\omega) = \mathrm{i} \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{c} \mathrm{e}^{-\mathrm{i}\omega z/c}\, W_m(z) \ .$$

Since we have already discussed the wake functions, we consider Eq. (5) the definition of impedances.

Instead of wake functions, an accelerator designer therefore could ask about the impedance of the accelerator. The impedance is the quantity most directly related to the maximum beam current allowed by the accelerator. Inverting the Fourier transforms, we have

$$W'_m(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega z/c}\ Z_m^{\parallel}(\omega) \ ,$$

$$W_m(z) = \frac{-\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega z/c}\ Z_m^{\perp}(\omega) \ .$$

The Panofsky–Wenzel theorem, which relates the longitudinal wake function to the derivative of the transverse wake function, also gives a relationship between the longitudinal and transverse impedances for a given m;

$$Z_m^{\parallel}(\omega) = \frac{\omega}{c} Z_m^{\perp}(\omega) \ .$$

11 Some analytical examples of impedances and wake functions

We mentioned earlier that there are three ways that wakefields are generated:

1. the beam is not relativistic;
2. the vacuum chamber is resistive;
3. the vacuum chamber is not smooth.

Three cases, each representing one of these three ways, that permit analytical expressions are given below.

11.1 Direct space charge

This wakefield and impedance come about when the beam is not sufficiently relativistic. Figure 10 shows the space charge wakefields in the x–y plane driven by an annular, infinitely thin, $\cos m\theta$ ring beam.

With a beam of radius a in a perfectly conducting round pipe of radius b and length L, we have the results of Table 2, where $Z_0 = \sqrt{\mu_0/\epsilon_0} \approx 377\ \Omega$. ($Z_0$ is the impedance of the vacuum. Yes, the vacuum has an impedance. An oscillating electromagnetic source will readily radiate into the vacuum,

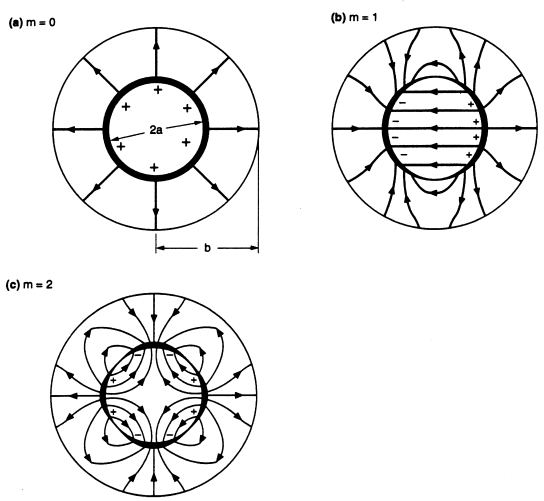


Fig. 10: Space charge wakefields in the x–y plane driven by an annular, infinitely thin, $\cos m\theta$ beam

Table 2: Results for direct space charge with a perfectly conducting vacuum chamber pipe

Impedances	Wake functions
$Z_0^{\parallel} = \mathrm{i}\dfrac{Z_0 L\omega}{4\pi c\gamma^2}\left(1+2\ln\dfrac{b}{a}\right)$	$W_0' = \dfrac{Z_0 cL}{4\pi\gamma^2}\left(1+2\ln\dfrac{b}{a}\right)\delta'(z)$
$Z_{m\neq 0}^{\perp} = \mathrm{i}\dfrac{Z_0 L}{2\pi\gamma^2 m}\left(\dfrac{1}{a^{2m}}-\dfrac{1}{b^{2m}}\right)$	$W_{m\neq 0} = \dfrac{Z_0 cL}{2\pi\gamma^2 m}\left(\dfrac{1}{a^{2m}}-\dfrac{1}{b^{2m}}\right)\delta(z)$

Table 3: Results for resistive wall

Impedances	Wake functions
$Z_m^{\parallel} = \frac{\omega}{c} Z_m^{\perp}$	$W_m = -\left(\dfrac{c}{\pi b^{m+1}(1+\delta_{m0})}\right)\left(\sqrt{\dfrac{Z_0}{\pi\sigma_c}}\right)\left(\dfrac{L}{\lvert z\rvert^{1/2}}\right)$
$Z_m^{\parallel} = \left(\frac{1-\mathrm{sgn}(\omega)\mathrm{i}}{1+\delta_{0m}}\right)\left(\frac{L}{\pi\sigma_c\delta_{\mathrm{skin}}b^{2m+1}}\right)$	$W_m' = -\left(\frac{c}{2\pi b^{m+1}(1+\delta_{m0})}\right)\left(\sqrt{\frac{Z_0}{\pi\sigma_c}}\right)\left(\frac{L}{\lvert z\rvert^{3/2}}\right)$

thus losing energy. In fact, the vacuum impedance is very large. A well designed accelerator will have an impedance only a fraction of the vacuum impedance. Accelerators are very poor radiators compared with the vacuum, and this is done purposedly.)

Owing to the factor $1/\gamma^2$, space charge effects are most significant for low-to-medium energy proton or heavy-ion accelerators.

The space charge impedance is purely imaginary, and is $\propto \mathrm{i}\omega$, as if it is a pure inductance. However, its sign is as if it is a capacitance. By convention, we call it 'capacitive'.

11.2 Resistive wall

Another case solvable analytically is for a round resistive pipe with radius b, conductivity σ_{c}, and length L. Defining the skin depth

$$\delta_{\mathrm{skin}} = \sqrt{\frac{2c}{|\omega| Z_0 \sigma_{\mathrm{c}}}}\,,$$

(For example, $\delta_{\mathrm{skin}}\,[\mathrm{mm}] = \frac{0.066}{\sqrt{f\,[\mathrm{MHz}]}}$ for copper.) one finds the results of Table 3.

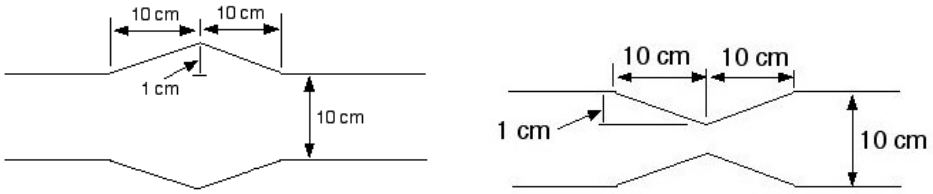


Fig. 11: Homework: find the impedances of these vacuum chamber geometries

The impedance is proportional to $(1 - \mathrm{i})$, i.e., it is half resistive and half inductive.

The $|z|^{-1/2}$ dependence of $W_m(z)$ indicates that the resistive wall wakefield (particularly, its transverse component) decays slowly and typically lasts long after the beam passage, sometimes long enough for the beam to see its own wakefield at its next revolution. The initial quick transient wakefields have been dropped in these expressions.

11.3 Slowly varying wall boundaries

The third way to generate impedances is by discontinuities. Consider a case when the vacuum chamber (perfectly conducting) wall varies along the accelerator slowly; a perturbation technique can be applied. Specify the wall variation by $h(z)$ (cylindrically symmetric bump). At low frequencies, $k = \omega/c <$ 1/(bump length or depth). The impedance is purely inductive—opposite in sign to the space charge impedance,

$$Z_0^{\parallel} = -\frac{2\mathrm{i}kZ_0}{b}\int_0^{\infty} \kappa|\tilde{h}(\kappa)|^2 \mathrm{d}\kappa \ ,$$

where

$$\tilde{h}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(z)\mathrm{e}^{-\mathrm{i}kz}\mathrm{d}z \ .$$

When the boundary varies rapidly, this formula breaks down. Numerical calculation then has to be applied.

Homework

Find the impedances of the vacuum chamber geometries in Fig. 11.

12 Beam energy spread in a linac

Consider a beam bunch travelling down an accelerator along the axis of the vacuum chamber pipe. The $m = 0$ wakefield excited by the beam produces a longitudinal force on particles in the beam. The main effect of this longitudinal force is a retarding voltage, causing energy changes of individual particles. As a result, there is a net energy loss of the beam to the wakefields. Furthermore, since not all particles in the bunch lose the same amount of energy, the wakefield also causes the beam to acquire an energy *spread*.

12.1 One-particle model

Consider first a one-particle model in which the beam bunch is a macroparticle of charge Ne. Travelling down the linac, it experiences the self-generated retarding longitudinal field and loses energy

$$\Delta E = -\frac{1}{2}Ne^2 W_0'(0^-) \ ,$$

where the factor $\frac{1}{2}$ is due to the fundamental theorem of beam loading.

Take the SLAC linac for example: $W_0'(0^-) = 7\ \mathrm{cm}^{-1} \times (L_0/L)$, where L_0 is the total linac length, 3 km and L is the length of an RF structure cell, 3.5 cm. We find $\Delta E = 2.2$ GeV for $N = 5 \times 10^{10}$. (To convert cgs units to other units, one may apply $\frac{e^2}{mc^2} = r_0$, the classical radius of the particle under consideration.)

12.2 Two-particle model

This estimate can be improved by a two-particle model. The beam bunch is represented by two macroparticles, one leading and another trailing at a distance $|z|$. The parasitic loss per particle in the leading macroparticle due to its self-field is 1.1 GeV. The trailing macroparticle loses, in addition to the 1.1 GeV due to the self-field,

$$\Delta E = -\frac{1}{2} N e^2 W_0'(z)\ ,$$

due to the wakefield left behind by the leading macroparticle.

Take $z = -\sigma_z = -1$ mm, $N = 5 \times 10^{10}$, and $W_0'(-1\ \mathrm{mm}) = 4.5\ \mathrm{cm}^{-1} \times (L_0/L)$, each particle in the trailing macroparticle loses an additional 1.4 GeV. The net loss of a trailing particle is 2.5 GeV.

The one-particle model estimates a parasitic loss per particle of 2.2 GeV. The two-particle model estimates an average loss of $(1.1 + 2.5)/2 = 1.8$ GeV. The two-particle model has introduced an energy split of 1.4 GeV, or a 2.8% energy spread if the beam energy at the end of the linac is 50 GeV.

12.3 An issue with linear colliders

For linear colliders, this energy spread makes it difficult to focus the beam to a small spot at the collision point in a final focus system, and is to be avoided. Most of this spread can be removed by properly phasing the accelerating RF voltage relative to the beam.

One concern for a high-intensity linear collider can be described as follows. The energy spread at the end of the linac scales is

$$\frac{\Delta E}{E} \approx \frac{\frac{1}{2} N e^2 W_0'}{G L_0} \approx \frac{\frac{1}{2} N e^2}{G b^2}\ ,$$

where G is the acceleration gradient, and $W_0' \approx L_0/b^2$ is the longitudinal wake function, where b is the vacuum chamber radius characterizing the size of the accelerating cavities.

On the other hand, the efficiency of energy extraction by the beam from the field energy U stored in the accelerating cavities

$$U \approx \frac{1}{8\pi} \left(\frac{G}{e}\right)^2 \pi b^2 L_0$$

is given by

$$\text{extraction efficiency} \approx \frac{NE}{U} \approx \frac{8 N e^2}{G b^2}\ ,$$

which is found to be equal to 16 times the energy spread.

In other words, to improve the energy spread of the beam at the end of the linac necessitates sacrificing the energy extraction efficiency. One way to ameliorate this problem is to compensate $(\Delta E/E)$ by phasing the RF voltage. Another way is to send a *train* of M bunches per filling of the RF cavities. This will increase the energy extraction efficiency by a factor of M, although at the cost of having to deal with multibunch interactions due to the long-range wakefields.

12.4 General bunch distribution

We now depart from the simplified models and consider a bunch with a general longitudinal distribution $\rho(z)$. The energy change for a test charge e at longitudinal position z can be written as $eV(z)$, where

$$V(z) = -\int_z^\infty \mathrm{d}z' \rho(z') W_0'(z - z')$$

or, equivalently,

$$V(z) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{e}^{\mathrm{i}\omega z/c} Z_0^{\parallel}(\omega) \tilde{\rho}(\omega) \ .$$

A negative value of $V(z)$ means that the test charge loses energy from the wakefield. An additional integration of $V(z)$ over the bunch then gives the total parasitic loss,

$$\Delta\mathcal{E} = \int_{-\infty}^{\infty} \rho(z) V(z) \mathrm{d}z \ .$$

For a bunch with Gaussian longitudinal distribution and uniform disc transverse distribution, for example, the energy spread due to space charge effects is

$$\frac{V(z)}{L} = \sqrt{\frac{2}{\pi}} \frac{q}{\gamma^2 \sigma_z^2} \left(\ln \frac{b}{a} + \frac{1}{2} \right) f\left(\frac{z}{\sigma_z} \right) \ ,$$
$$f(u) = u \mathrm{e}^{-u^2/2} \ .$$

Generally, particles in the front of the bunch ($z > 0$) lose energy due to wakefields, while particles in the back of the bunch ($z < 0$) may gain or lose energy, depending on the length of the bunch. This is not true for the special case of the space charge effect, for which particles in the front of the bunch gain energy, and particles in the back of the bunch lose energy. For the space charge effect, the energy gained by the bunch head is necessarily given up by the bunch tail so that the net energy of the bunch is unchanged.

Consider a numerical example of a 50 MeV proton transport line. If we take $q = 10^{10}e$, $\sigma_z = 3$ cm, $a = 2$ cm, and $b = 5$ cm, we obtain a longitudinal space charge force of ± 6 V/m for particles located at $z = \pm\sigma_z$. The net energy change of these particles after travelling 100 m of this transport line is $eV/\beta = \pm 2$ keV. The space charge induced beam energy spread is therefore $\pm 4 \times 10^{-5}$.

For a resistive wall, we have

$$\frac{V(z)}{L} = \frac{q}{4b\sigma_z^{3/2}} \sqrt{\frac{c}{2\pi\sigma}} f\left(\frac{z}{\sigma_z} \right) \ ,$$
$$f(u) = -|u|^{3/2} \mathrm{e}^{-u^2/4} \left[(I_{-1/4} - I_{3/4}) \, \mathrm{sgn}(u) - I_{1/4} + I_{-3/4} \right] \ ,$$

with the Bessel functions $I_{\pm 1/4}$ and $I_{\pm 3/4}$ evaluated at $u^2/4$. Continuing this numerical example, assuming an aluminium pipe, a particle located at $0.5\sigma_z$ ahead of the bunch centre loses an energy of 0.1 eV after travelling 100 m, and a particle located at $1.8\sigma_z$ behind the bunch centre gains 0.04 eV.

13 Beam break-up in a linac

In the previous section, the beam was centred in a cylindrically symmetric pipe. There were no transverse wake forces ($m = 0$). If the beam is executing a betatron oscillation, an $m = 1$ dipole wakefield is excited by the bunch head, which causes transverse deflection of the bunch tail. For a high-intensity beam, this leads to a transverse break-up of the beam. The first observation of beam break-up was made on the SLAC linac.

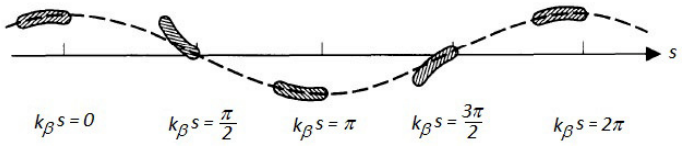


Fig. 12: Propagation of beam along linac

13.1 Two-particle model

To proceed with a simplified macroparticle model, we first note that a one-particle model is not useful because a point charge does not exert a transverse wake force on itself. In the two-particle model, the leading macroparticle, unperturbed by its own transverse wakefield, executes a free betatron oscillation

$$y_1(s) = \hat{y} \cos k_\beta s \ .$$

The trailing macroparticle, at a distance $|z|$, sees a deflecting wakefield left behind by its leading partner,

$$\begin{aligned} y_2'' + k_\beta^2 y_2 &= -\frac{Ne^2 W_1(z)}{2EL}\ y_1 \\ &= -\frac{Nr_0 W_1(z)}{2\gamma L}\ \hat{y} \cos k_\beta s \ , \end{aligned}$$

where $W_1(z)$ is the transverse wake function per cavity period L. We have ignored acceleration of the beam energy. For the SLAC linac, $k_\beta \approx 0.06\ \mathrm{m}^{-1}$ and $k_\beta L \approx 0.002$.

The solution is

$$y_2(s) = \hat{y}\Big[\cos k_\beta s - \frac{Nr_0 W_1(z)}{4k_\beta \gamma L}\ s \sin k_\beta s\Big] \ .$$

The first term describes the free oscillation and the second term is the resonant response to the driving wake force. The amplitude of the second term grows linearly with s. The mechanism of the beam break-up is that particles in the tail of the beam are driven exactly on resonance by the oscillating wake left by the head of the beam.

At the end of the linac, the oscillation amplitude of the bunch tail relative to the bunch head is characterized by the dimensionless growth parameter

$$\Upsilon = -\frac{Nr_0 W_1(z) L_0}{4k_\beta \gamma L} \ ,$$

where L_0 is the total linac length.

For a beam bunch with realistic distribution, the bunch is distorted into a banana shape. The motion of the bunch head is $\cos k_\beta s$, while the deviation of the bunch tail relative to the bunch head is $s \sin k_\beta s$. When the bunch head is at a maximum displacement, the tail lines up with the bunch head, but when the bunch head displacement is zero, the tail swing is maximum (Fig. 12). As the beam propagates down the linac, the swing amplitude of the flapping tail increases with s until the tail breaks up and particles are lost. Note that the sign of the tail swing shown is not arbitrary, because $\Upsilon > 0$.

Figure 13 shows four transverse beam profiles observed at the end of the SLAC linac with $N = 2 \times 10^{10}$. The leftmost profile is for a carefully steered beam. When the beam was injected off centre by 0.2, 0.5, and 1 mm, the beam profiles were as shown in the corresponding right-hand panels. The beam sizes σ_x and σ_y were $\sim$120 μm.

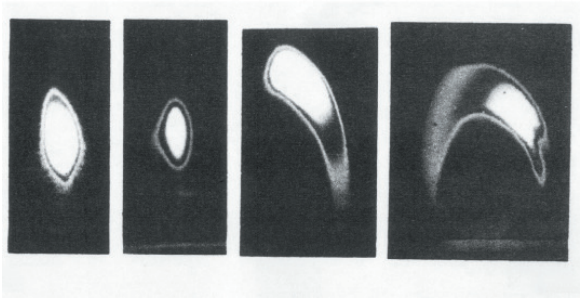


Fig. 13: Four transverse beam profiles observed at the end of the SLAC linac with $N = 2 \times 10^{10}$. From left to right: beam centred, beam offset by 0.2, 0.5, and 1 mm, respectively. (Courtesy John Seeman)

13.2 With acceleration

So far we have ignored beam acceleration, which has an important stabilizing effect because, as its energy increases, the beam becomes more rigid and less vulnerable to the wakefields. Furthermore, the driving beam's displacement also decreases with adiabadic damping. Repeating a similar analysis but taking acceleration into account yields the growth parameter

$$\Upsilon = -\frac{Nr_0W_1(z)L_0}{4k_\beta\gamma_f L}\ln\frac{\gamma_f}{\gamma_i}\ ,$$

which is basically simply replacing the factor L_0/γ with its integral counterpart $\int_0^{L_0} \mathrm{d}s/\gamma(s)$. Owing to acceleration, the tail amplitude thus grows logarithmically rather than linearly with s, and the growth parameter is much reduced. If the beam is accelerated in the SLAC linac from 1 to 50 GeV, the factor Υ becomes 14, instead of 180 if the beam coasts at 1 GeV.

The beam break-up instability described so far is quite severe, even with acceleration. To control it, the beam has to be tightly focused, rapidly accelerated, and carefully injected, and its trajectory must be carefully steered down the linac. Interestingly, the contribution from trajectory mis-steering can, in principle, be largely compensated for by an intentional mis-injection.

13.3 BNS damping

It turns out, however, that there is another interesting and effective method to ameliorate the situation. This method, known as *BNS damping*, after Balakin, Novokhatsky, and Smirnov, is described next.

Consider first the case without acceleration, where the leading macroparticle executes a free betatron oscillation. The idea of BNS damping requires the introduction of a slightly stronger betatron focusing of the bunch tail than the bunch head. The equation of motion of the tail particles can be written as

$$y_2'' + (k_\beta + \Delta k_\beta)^2 y_2 = -\frac{Nr_0W_1(z)}{2\gamma L}\hat{y}\cos k_\beta s\ .$$

The solution, assuming $|\Delta k_\beta/k_\beta| \ll 1$, is

$$y_2(s) = \hat{y}\cos(k_\beta + \Delta k_\beta)s + \frac{Nr_0W_1(z)}{4k_\beta\Delta k_\beta\gamma L}\,\hat{y}\big[\cos(k_\beta s + \Delta k_\beta s) - \cos k_\beta s\big]\ .$$

Compared with the result without Δk_β, one observes that, by introducing a slightly different focusing strength for the bunch tail, the beam break-up mechanism of the bunch head resonantly driving the bunch tail is removed. A further inspection shows that there exists a magical condition for the bunch tail to follow the bunch head exactly for all s, namely

$$\frac{Nr_0W_1(z)}{4k_\beta\Delta k_\beta\gamma L} = -1\ ,$$

or, equivalently,

$$\frac{\Delta k_\beta}{k_\beta} = -\frac{N r_0 W_1(z)}{4 k_\beta^2 \gamma L} = \frac{\Upsilon}{k_\beta L_0} \,, \tag{6}$$

where Υ is as defined before, and $k_\beta L_0$ is the total betatron phase advance of the linac. For short bunches, Υ and Δk_β are positive; the betatron focusing required to fulfil the BNS condition is therefore stronger at the bunch tail than at the bunch head.

Under the BNS condition, $y_2(s) = y_1(s) = \hat{y} \cos k_\beta s$, and the beam no longer breaks up. (The mechanism of BNS damping is not to be confused with that of Landau damping, to be discussed later. They have little in common other than the fact that both involve a frequency spread in the bunch population.) Physically, this happens because the additional external focusing force introduced for the bunch tail has compensated for the defocusing dipole deflection force due to the wakefield left behind by the bunch head. Note that the BNS focusing has to be adjusted according to the beam intensity.

There are different ways to provide the BNS focusing. One is to introduce a radio-frequency quadrupole whose strength changes as the bunch passes by, so that the head and tail of the bunch see different quadrupole strengths. Another is to choose the location of the bunch relative to the acceleration RF voltage in such a way that the bunch tail acquires a lower energy than the bunch head. The energy spread across the bunch then causes a spread in betatron focusing according to

$$\frac{\Delta k_\beta}{k_\beta} = \xi \frac{\Delta E}{E} \,,$$

where ξ is the chromaticity determined by the linac design. For a FoDo lattice design, for example,

$$\xi = -\frac{2}{\mu} \tan \frac{\mu}{2} \,,$$

where μ is the betatron phase advance per FoDo cell. By properly choosing the phase of the RF voltage relative to the beam bunch, the betatron focusing required by the BNS condition can be obtained, provided the required $\Delta k_\beta / k_\beta$ is not excessive.

For an accelerated beam, the BNS condition is still given by Eq. (6), except that the parameter Υ is now that given by the case with acceleration. Take the SLAC linac, for example, and assume $\mu = 90^\circ$; then the energy deviation of the bunch tail from the bunch head required by the BNS condition is about -5.5%. BNS damping has been routinely employed to control the beam break-up instability in the SLAC linac operations.

14 Parasitic heating

When a beam bunch of charge q and line density $\lambda(t)$ traverses an impedance $Z_0^\parallel(\omega)$, it loses energy to the impedance. This *parasitic energy loss* (or higher-order mode heating) is

$$\Delta \mathcal{E} = -\kappa^\parallel q^2 \,,$$

where $\kappa^\parallel$ is the *loss factor*, in units of V/pC,

$$\kappa^\parallel = \frac{1}{\pi} \int_0^\infty \mathrm{d}\omega \, \mathrm{Re} Z_0^\parallel(\omega) \, |\tilde{\lambda}(\omega)|^2 \,. \tag{7}$$

For a Gaussian bunch, $\lambda(t) = \mathrm{e}^{-t^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$, $\tilde{\lambda}(\omega) = \mathrm{e}^{-\omega^2\sigma^2/2}$.

Only the real part of the impedance contributes to the parasitic loss. The space charge or the slowly varying wall impedances do not cause net energy loss to the beam. However, this does not mean that individual particles do not change their energies. It only means that there is energy transfer among different parts of the bunch, while the total energy of the whole bunch remains unchanged.

For a resistive wall,

$$\frac{\kappa^{\|}(\sigma)}{L} = \frac{\Gamma(\frac{3}{4})c}{4\pi^2 b\sigma_z^{3/2}}\left(\frac{Z_0}{2\sigma_c}\right)^{1/2}, \qquad \Gamma\left(\frac{3}{4}\right) = 1.225\ .$$

where b is the pipe radius (assumed cylindrically symmetrical). It shows explicitly that parasitic loss is more pronounced for short bunches.

Parasitic loss gives rise to heating of the vacuum chamber wall where there are impedances. In high-intensity electron storage rings, the beam position monitors or bellows can heat up. This is especially serious for short bunches.

Most of the parasitic loss occurs as the beam traverses a discontinuity structure. Part of the wakefield gets trapped if the structure is cavity-like and if the wakefield frequency is below the cut-off frequency of the pipe. The trapped field energy is eventually deposited as heat on the cavity walls. The rest of the wakefield, with frequency higher than the cut-off frequency, propagates up and down the pipe and eventually dissipates on lossy material elsewhere in the vacuum chamber. For a cavity structure, $\kappa^{\|}$ is given by a sum over cavity modes below the cut-off frequency, plus a contribution above the cut-off frequency. Each cavity mode below the cut-off frequency contributes a resonator impedance, with

$$\kappa^{\|} \approx \begin{cases} \dfrac{\omega_r R_{\rm s}}{2Q_r}{\rm e}^{-\omega_r^2\sigma^2} & \text{high-}Q\text{ resonator}\ , \\ \dfrac{\omega_r R_{\rm s}}{2Q_r} & \text{low-}Q\text{ resonator, short bunch }\omega_r\sigma \ll 1\ , \\ \dfrac{R_{\rm s}}{4\sqrt{\pi}Q_r^2\omega_r^2\sigma^3} & \text{low-}Q\text{ resonator, long bunch }\omega_r\sigma \gg 1\ . \end{cases}$$

Above the cut-off frequency, the impedance per cavity can be represented by the diffraction model,

$$Z_0^{\|}(\omega) = [1 + {\rm sgn}(\omega){\rm i}]\frac{Z_0}{2\pi^{3/2}}\frac{1}{b}\sqrt{\frac{cg}{|\omega|}}\ ,$$

where g is the gap size of the cavity. This impedance has both real and imaginary parts.

For a single bunch in a circular accelerator, the integral in Eq. (7) is replaced by an infinite sum,

$$\kappa^{\|}(\sigma) = \frac{\omega_0}{2\pi}\sum_{p=-\infty}^{\infty} Z_0^{\|}(p\omega_0)\,|\tilde{\lambda}(p\omega_0)|^2\ .$$

For short bunches in large accelerators ($\omega_0 \ll 1/\sigma$), the sum can be replaced by an integral, and the difference between single passes and multiple passes disappears.

The parasitic loss by the beam goes into wakefields. Typically, only a small fraction of the particle energy becomes wakefields; most of the energy stored in the wakefields ends up as heat on the vacuum chamber walls. However, under unfavourable conditions, a small portion of the wakefield energy can be transferred systematically back to beam motion, causing beam instabilities. Parasitic loss, therefore, is the ultimate culprit for the various collective instabilities.

15 The Vlasov equation

The Vlasov equation describes the collective behaviour of a multiparticle system under the influence of electromagnetic forces. To construct the Vlasov equation, one starts with the single-particle equations of motion (assume 1D):

$$\begin{aligned} \dot{q} &= f(q,p,t) \\ \dot{p} &= g(q,p,t)\ . \end{aligned}$$

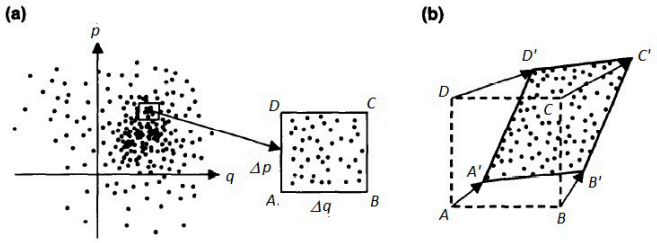


Fig. 14: Derivation of the Vlasov equation

The state of a particle at a given time t is represented by a point in the phase space (q, p). The motion of a particle is described by the motion of its representative point in phase space.

In a conservative deterministic system, the particle trajectory in phase space is completely determined by the initial conditions (q_0, p_0) at time $t = t_0$. Two particles having the same initial conditions must have exactly the same trajectory in phase space. It follows that the only way for two trajectories to meet at a given time is for them to coincide at all times. In other words, trajectories either completely coincide or never intersect.

Consider a distribution of particles occupying an area in the phase space. Because they cannot intersect with particles on the boundary of the distribution as the distribution evolves in time, particles inside cannot leak out of and particles outside cannot enter the distribution.

If the system is conservative,

$$f = \frac{\partial H}{\partial p} \quad \text{and} \quad g = -\frac{\partial H}{\partial q} \,,$$

where H is the Hamiltonian. It follows that

$$\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p} = 0 \,.$$

As will be seen later, this condition leads to an area preservation property: as the particle distribution evolves in the phase space, its shape may be distorted but its area remains constant. In fact, in a non-conservative system, $\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p}$ has the physical meaning of the rate of area shrinkage.

Consider a distribution of a group of particles in phase space at time t. A rectangular $\Delta q\, \Delta p$ box is drawn (Fig. 14(a)):

$$\begin{aligned} &A(q,\ p) \,, \\ &B(q + \Delta q,\ p) \,, \\ &C(q + \Delta q,\ p + \Delta p) \,, \\ &D(q,\ p + \Delta p) \,. \end{aligned}$$

At a later time, $t + \mathrm{d}t$, the box moves and deforms into a parallelogram $A'B'C'D'$ (Fig. 14(b)) with the same area as $ABCD$. All particles inside the box move with the box. Let the number of particles enclosed by the box be

$$\psi(q, p, t)\, \Delta q\, \Delta p \,,$$

where ψ is the phase space distribution density normalized by

$$\int_{-\infty}^{\infty} \mathrm{d}q \int_{-\infty}^{\infty} \mathrm{d}p\, \psi(q, p, t) = N \,.$$

The vertices of the parallelogram are

$$
\begin{aligned}
&A'[q + f(q,p,t)\mathrm{d}t,\ p + g(q,p,t)\mathrm{d}t]\ ,\\
&B'[q + \Delta q + f(q+\Delta q,p,t)\mathrm{d}t,\ p + g(q+\Delta q,p,t)\mathrm{d}t]\ ,\\
&C'[q + \Delta q + f(q+\Delta q,p+\Delta p,t)\mathrm{d}t,\ p + \Delta p + g(q+\Delta q,p+\Delta p,t)\mathrm{d}t]\ ,\\
&D'[q + f(q,p+\Delta p,t)\mathrm{d}t,\ p + \Delta p + g(q,p+\Delta p,t)\mathrm{d}t]\ .
\end{aligned}
$$

The condition that no particles leak into or out of the box gives

$$\psi(q,\ p,\ t)\ \text{area}(ABCD) = \psi(q + f\mathrm{d}t,\ p + g dt,\ t + \mathrm{d}t)\ \text{area}(A'B'C'D')\ .$$

For a Hamiltonian system, the area of the box is conserved:

$$
\begin{aligned}
\text{area}(A'B'C'D') &= |\overrightarrow{A'B'} \times \overrightarrow{A'D'}| \\
&= \Delta q\,\Delta p\left[1 + \left(\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p}\right)\mathrm{d}t\right] \\
&= \Delta q\,\Delta p = \text{area}(ABCD)\ .
\end{aligned}
$$

We then have

$$
\begin{aligned}
\psi(q,p,t) &= \psi(q + f\mathrm{d}t,\ p + g\mathrm{d}t,\ t + \mathrm{d}t) \\
&= \psi + \frac{\partial\psi}{\partial q} f\mathrm{d}t + \frac{\partial\psi}{\partial p} g\mathrm{d}t + \frac{\partial\psi}{\partial t}\mathrm{d}t\ ,
\end{aligned}
$$

or, after cancelling out ψ on both sides, we obtain the *Vlasov equation*

$$\frac{\partial\psi}{\partial t} + f\frac{\partial\psi}{\partial q} + g\frac{\partial\psi}{\partial p} = 0\ .$$

The Vlasov equation can also be put in a much more vague form:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = 0, \quad \text{or} \quad \psi = \text{constant in time.}$$

Sometimes loosely referred to as the *Liouville theorem*, this form states that the local particle density does not change if (an important if) the observer moves with the flow of boxes, but it does not tell how the boxes flow. The Vlasov form, in contrast, does not have this ambiguity, since it explicitly contains the single-particle information f and g.

Strictly speaking, f and g are given by external forces. Collisions among discrete particles in the system, for example, are excluded. However, if a particle interacts more strongly with the *collective* fields of the other particles than with its nearest neighbours, the Vlasov equation still applies if one treats the collective fields on the same footing as the external fields. In fact, this forms the basis of treating the collective instabilities using the Vlasov technique.

One special case where the Vlasov equation can be solved exactly is when the system is described by a Hamiltonian $H(q,p)$ that does not have an explicit time dependence. A stationary solution is found to be

$$\psi(q,p) = \text{any function of } H(q,p)\ .$$

In this system, individual particles stream along constant-H contours in phase space in such a way that the overall distribution is stationary.

In the derivation of the Vlasov equation, we have assumed that there are no diffusion or external damping effects. This is usually a good approximation for proton beams. For electron beams, synchrotron radiation contributes to both damping and diffusion, and one needs to apply the *Fokker–Planck equation*, a generalization of the Vlasov equation. However, when the instability occurs faster than the damping or diffusion times, the Vlasov treatment also at least approximately applies to electrons.

16 Potential-well distortion

As a first application of the Vlasov technique, we study the effect of a longitudinal wakefield on a distortion of the equilibrium shape of a beam bunch. The mechanism is a static one; no part of the beam bunch is executing collective oscillation. The extent of distortion depends on the beam intensity; higher beam intensities cause larger distortions.

Consider a bunched beam that travels along the axis of the vacuum chamber pipe in a circular accelerator. We assume that the beam does not have any transverse dimension, i.e., the beam is an infinitesimally thin thread. Such a beam does not generate transverse wakefields; only the $m = 0$ wake is excited.

Consider a particle in the beam executing longitudinal synchrotron oscillation. The phase space coordinates q and p are

$$q = z \quad \text{and} \quad p = -\frac{\eta c}{\omega_{\rm s}}\delta \,,$$

where η is the slippage factor defined by the accelerator lattice and $\omega_{\rm s}$ is the synchrotron oscillation frequency.

The single-particle equations of motion are

$$z' = -\eta\delta \quad \text{and} \quad \delta' = K(z) \,.$$

We leave $K(z)$ open for now, except that we do know it cannot depend on δ, because the system is conservative.

The Vlasov equation reads

$$\frac{\partial\psi}{\partial s} - \eta\delta\frac{\partial\psi}{\partial z} + K(z)\frac{\partial\psi}{\partial\delta} = 0 \,,$$

where we will set $\partial\psi/\partial s = 0$, since we are looking for a stationary distribution. The general stationary solution is

$$\psi(z,\delta) = \text{any function of the Hamiltonian } H \,,$$
$$H = \frac{\eta^2 c^2}{\omega_{\rm s}}\left[\frac{\delta^2}{2} + \frac{1}{\eta}\int_0^z K(z'){\rm d}z'\right] \,.$$

The second integral term in H is the potential-well term. A simple harmonic system would have a parabolic potential well.

If the potential well is provided by an external RF voltage $V_{\rm RF}(z)$, we have

$$K(z) = \frac{eV_{\rm RF}(z)}{CE} = \frac{\omega_{\rm s}^2}{c^2\eta V'_{\rm RF}(0)}V_{\rm RF}(z) \,.$$

A practical case is given by $V_{\rm RF} = \hat{V}\sin(\omega_{\rm RF}z/c)$. The deviation of $V_{\rm RF}(z)$ from a linear dependence on z is a cause of potential-well distortion. The general stationary distribution is given by any function of the Hamiltonian

$$H = \frac{\eta^2 c^2}{2\omega_{\rm s}}\delta^2 + \frac{\omega_{\rm s}c^2}{\omega_{\rm RF}^2}\left[1 - \cos\left(\frac{\omega_{\rm RF}z}{c}\right)\right] \,.$$

This Hamiltonian also describes the form of the RF bucket. A stationary distribution must conform to the contours of constant Hamiltonian inside the bucket. For small oscillation amplitudes, we have $K = \omega_{\rm s}^2 z/\eta c^2$, the case of simple harmonic motion.

One noteworthy special case of the stationary beam distribution is that given by $\exp(-\text{constant} \times H)$. This distribution is always Gaussian in δ. If the bunch length is much shorter than the RF wavelength, ($z \ll c/\omega_{\rm RF}$) the familiar quadratic form of the Hamiltonian is re-established, and the distribution

is also Gaussian in z. As the bunch length increases, the bunch shape deviates from Gaussian; the potential well is distorted by the RF bucket, although the distribution remains Gaussian in δ.

There is another reason for the Hamiltonian to deviate from the quadratic form, and thus to cause potential-well distortion; namely, the wakefields. Consider a bunch that is short compared with the RF wavelength. Let the wake function be $W_0'(z)$ integrated over the accelerator circumference, and assume that the wake has dissipated before the beam completes one revolution,

$$K(z) = \frac{\omega_s^2}{\eta c^2} z - \frac{r_0}{\gamma C} \int_z^\infty \mathrm{d}z' \, \rho(z') W_0'(z - z') \; .$$

The corresponding Hamiltonian is

$$H = \frac{\eta^2 c^2}{2\omega_s} \delta^2 + \frac{\omega_s}{2} z^2 - \frac{\eta c^2 r_0}{\omega_s \gamma C} \int_0^z \mathrm{d}z'' \int_{z''}^\infty \mathrm{d}z' \, \rho(z') W_0'(z'' - z') \; .$$

The stationary solution to the Vlasov equation must be a function of H. The complication here is that the complicated z-dependence of H now involves the beam density ρ, which in turn is determined by the stationary distribution itself. Clearly some self-consistency requirement is to be imposed.

Continuing the Gaussian example, the stationary distribution maintains its Gaussian distribution in δ,

$$\psi(z, \delta) = \frac{1}{\sqrt{2\pi}\sigma_\delta} \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \rho(z) \; .$$

The Gaussian form and the value of σ_δ are arbitrary if the collective behaviour is governed by the Vlasov equation, as in the case of a proton beam. However, if the beam behaviour is governed, as for an electron beam, by the Fokker–Planck equation, then this Gaussian distribution with a specific value for σ_δ will be the unique solution of the stationary beam distribution.

This distribution matches the stationary solution

$$\psi(z, \delta) \propto \exp\left(-\frac{\omega_s}{\eta^2 c^2 \sigma_\delta^2} H\right) \; .$$

Self-consistency then imposes a transcendental equation for $\rho(z)$, called the *Haissinski equation*,

$$\rho(z) = \rho(0) \exp\left[-\frac{1}{2}\left(\frac{\omega_s z}{\eta c \sigma_\delta}\right)^2 + \frac{r_0}{\eta \sigma_\delta^2 \gamma C} \int_0^z \mathrm{d}z'' \int_{z''}^\infty \mathrm{d}z' \, \rho(z') W_0'(z'' - z')\right] \; .$$

In the limit of zero beam intensity, the solution reduces to the bi-Gaussian form, where $\sigma_z = \eta c \sigma_\delta / \omega_s$. For high beam intensities, $\rho(z)$ deforms from a Gaussian. The Haissinski equation is solved numerically for $\rho(z)$ once $W_0'(z)$ is known and σ_δ is specified. Figure 15 shows the result for the electron damping ring for the SLAC Linear Collider. The bunch shape is Gaussian at low beam intensities, and distorts as the beam intensity is increased. The calculations agree with the measurements.

Note that the distribution leans forward ($z > 0$) as the beam intensity increases. This effect comes from the parasitic loss of the beam bunch, and is a consequence of the real (resistive) part of the impedance. Since the SLAC damping ring is operated above transition, the bunch moves forward so that the parasitic energy loss can be compensated by the RF voltage.

Note also that the bunch length increases as the beam intensity increases. The bunch shape distortion comes mainly from the imaginary part of the impedance. That the bunch lengthens is a consequence of the fact that the imaginary part of the impedance is mostly inductive.

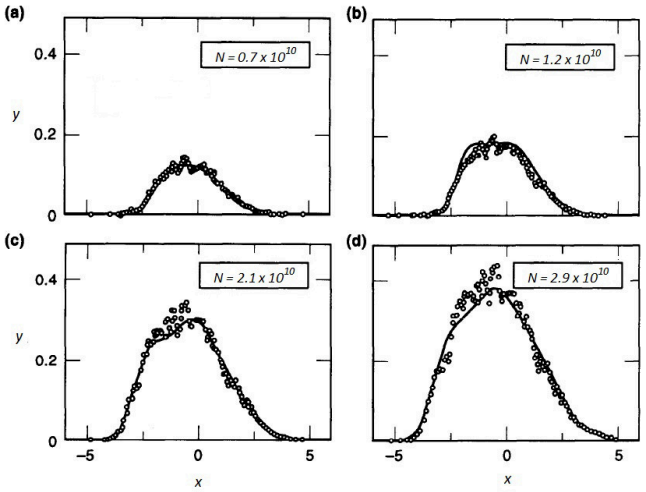


Fig. 15: Potential-well distortion for the electron damping ring for the SLAC Linear Collider. The bunch shape is Gaussian at low beam intensities and distorts as the beam intensity is increased. The horizontal axis is $x = -z/\sigma_{z0}$, where σ_{z0} is the unperturbed rms bunch length. The vertical scale gives $y = 4\pi e\rho(z)/V'_{\rm RF}(0)\sigma_{z0}$. (Courtesy Karl Bane, 1992.)

Fig. 16: Kenneth Robinson (1925–1979)

17 Robinson instability

The Robinson instability (named for Kenneth Robinson, Fig. 16) is one of the most basic instability mechanisms. It is a longitudinal instability that occurs in circular accelerators. The main contributor to this instability is the longitudinal impedance due to the RF accelerating cavities. These cavities are tuned to have a resonant frequency $\omega_{\rm R}$ for its fundamental accelerating mode. This mode is what the klystrons feed into but, at the same time, it is also a big source of impedance. Since we must have this RF mode to accelerate the beam, we must accept its large impedance and live with it.

The real part of this impedance narrowly peaks at $\omega_{\rm R}$, the width $\Delta\omega/\omega_{\rm R} \approx \pm 1/Q$. Typically, $Q \sim 10^4$ (or 10^9 for superconducting cavities).

By design, $\omega_{\rm R}$ is very close to an integer multiple of the revolution frequency ω_0. This necessarily means that the wakefield excited by the beam in the cavities contains a major frequency component near $\omega_{\rm R} \approx h\omega_0$, and the impedance $Z_0^{\parallel}(\omega)$ has sharp peaks at $\pm h\omega_0$, where h is an integer called the *harmonic number*.

As we will soon show, the exact value of $\omega_{\rm R}$ relative to $h\omega_0$ is of critical importance for the

stability of the beam. Above transition, the beam will be unstable if $\omega_{\rm R}$ is slightly above $h\omega_0$ and stable if slightly below. Below transition, it is the other way around.

Consider a one-particle model. The beam is just a big point charge Ne, without internal structures, and consider its longitudinal motion under the influence of its own longitudinal wakefield. Let z_n be the longitudinal displacement of the beam at the accelerating RF cavity in the nth revolution. The rate of change of z_n is related to the relative energy error δ_n of the beam in the same nth revolution by

$$\frac{\mathrm{d}}{\mathrm{d}n} z_n = -\eta C \delta_n \ .$$

The storage ring is above transition if $\eta > 0$ and below transition if $\eta < 0$.

The energy error also changes with time. In the absence of wakefields, its equation of motion is

$$\frac{\mathrm{d}}{\mathrm{d}n} \delta_n = \frac{(2\pi\nu_{\rm s})^2}{\eta C} z_n \ ,$$

where $\nu_{\rm s}$ is the unperturbed synchrotron tune. If we combine these two equations, we get a simple harmonic oscillation for both z_n and δ_n, i.e., the normal synchrotron oscillation.

But for an intense beam, we have to add the wakefield term,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}n} \delta_n &= \frac{(2\pi\nu_{\rm s})^2}{\eta C} z_n + \frac{eV(z_n)}{E} \\ &= \frac{(2\pi\nu_{\rm s})^2}{\eta C}\ z_n - \frac{Nr_0}{\gamma} \sum_{k=-\infty}^{n} W_0'(kC - nC + z_n - z_k) \ , \end{aligned}$$

where W_0' is the longitudinal wake function accumulated over one turn of the accelerator. The summation over k is over the wakefields left behind by the beam from all revolutions prior to the nth. The equation of motion now becomes

$$\frac{\mathrm{d}^2 z_n}{\mathrm{d}n^2} + (2\pi\nu_{\rm s})^2 z_n = \frac{Nr_0\eta C}{\gamma} \sum_{k=-\infty}^{n} W_0'(kC - nC + z_n - z_k) \ .$$

If the beam bunch has an oscillation amplitude much shorter than the wavelength of the fundamental cavity mode, one can expand the wake function,

$$W_0'(kC - nC + z_n - z_k) \approx W_0'(kC - nC) + (z_n - z_k) W_0''(kC - nC) \ .$$

The first term is a static term independent of the motion of the beam. It describes the parasitic loss effect discussed earlier and can be taken care of by a constant shift in z_n. We will drop this term altogether because here we are interested only in the dynamical effects. The second term does involve the dynamics of the beam. The quantity $z_n - z_k$ is the difference of the z terms and – although we will not make such an approximation – resembles a time derivative $\mathrm{d}z/\mathrm{d}n$, which in turn suggests an instability, since a $\mathrm{d}z/\mathrm{d}n$ term in a $\mathrm{d}^2 z/\mathrm{d}n^2$ equation indicates a possible exponential growth (or damping) of z.

We now need to solve this equation for z_n as a function of n. To do so, let

$$z_n \propto \mathrm{e}^{-\mathrm{i}n\Omega T_0} \ ,$$

where Ω is the mode frequency of the beam oscillation and is a key quantity yet to be determined.

Substituting into the equation of motion, we find an algebraic equation for Ω,

$$\Omega^2 - \omega_{\rm s}^2 = -\frac{Nr_0\eta c}{\gamma T_0} \sum_{k=-\infty}^{\infty} (1 - \mathrm{e}^{-\mathrm{i}k\Omega T_0}) W_0''(kC) \ ,$$

where $\omega_{\rm s} = \nu_{\rm s}\omega_0$ is the unperturbed synchrotron oscillation frequency. Now the wake function can be expressed in terms of the impedance by a Fourier transform,

$$\Omega^2 - \omega_{\rm s}^2 = -{\rm i}\,\frac{Nr_0\eta}{\gamma T_0^2}\sum_{p=-\infty}^{\infty}\left[p\omega_0 Z_0^{\parallel}(p\omega_0) - (p\omega_0+\Omega)Z_0^{\parallel}(p\omega_0+\Omega)\right] .$$

Given the impedance, this equation can in principle be solved for Ω. Note that Ω appears on both sides of the equation. Here, however, we take a perturbative approach and assume that Ω does not deviate much from $\omega_{\rm s}$ for modest beam intensities. We thus replace Ω with $\omega_{\rm s}$ on the right-hand side of the equation. Quantity Ω is then easily solved.

In general, Ω is complex. The real part of Ω is the perturbed synchrotron oscillation frequency of the collective beam motion, and the imaginary part gives the growth rate (or damping rate if negative) of the motion. We then obtain a *mode frequency shift*,

$$\begin{aligned}\Delta\Omega =&\ {\rm Re}(\Omega-\omega_{\rm s})\\ =&\ \frac{Nr_0\eta}{2\gamma T_0^2\omega_{\rm s}}\sum_{p=-\infty}^{\infty}\left[p\omega_0{\rm Im}Z_0^{\parallel}(p\omega_0) - (p\omega_0+\omega_{\rm s}){\rm Im}Z_0^{\parallel}(p\omega_0+\omega_{\rm s})\right. ,\end{aligned}$$

and an *instability growth rate*,

$$\tau^{-1} =\ {\rm Im}(\Omega-\omega_{\rm s}) = \frac{Nr_0\eta}{2\gamma T_0^2\omega_{\rm s}}\sum_{p=-\infty}^{\infty}(p\omega_0+\omega_{\rm s}){\rm Re}Z_0^{\parallel}(p\omega_0+\omega_{\rm s}l) .$$

The imaginary part of the impedance contributes to the collective frequency shift. The real part contributes to the instability growth rate. Note that when we measure the synchrotron frequency in an actual operation, what shows up in the beam spectrum is not $\omega_{\rm s}$, but Ω.

So far, our result holds for arbitrary impedance. We now consider the resonator impedance for the fundamental cavity mode. The only significant contributions to the growth rate come from two terms in the summation, namely $p = \pm h$, assuming that $\omega_{\rm R}/Q \ll \omega_0$,

$$\tau^{-1} \approx \frac{Nr_0\eta h\omega_0}{2\gamma T_0^2\omega_{\rm s}}\left[{\rm Re}Z_0^{\parallel}(h\omega_0+\omega_{\rm s}) - {\rm Re}Z_0^{\parallel}(h\omega_0-\omega_{\rm s})\right] .$$

Beam stability requires $\tau^{-1} \le 0$. That is, the real part of the impedance must be less at frequency $h\omega_0+\omega_{\rm s}$ than at frequency $h\omega_0-\omega_{\rm s}$ if $\eta > 0$, and the other way around if $\eta < 0$. This condition gives the Robinson stability criterion that, above transition, the resonant frequency $\omega_{\rm R}$ of the fundamental cavity mode should be slightly detuned downwards from an exact integral multiple of ω_0. Below transition, it should be the other way around, as sketched in Fig. 17.

Physically, the Robinson instability comes from the fact that the revolution frequency of an off-momentum beam is not given by ω_0 but by $\omega_0(1-\eta\delta)$. To illustrate the physical origin, consider a beam executing synchrotron oscillation above transition. Owing to the energy error of the beam, the impedance samples the beam signal at a frequency slightly below $h\omega_0$ if $\delta > 0$, and slightly above $h\omega_0$ if $\delta < 0$. To damp this synchrotron oscillation, we need to let the beam lose energy when $\delta > 0$ and gain energy when $\delta < 0$. This can be achieved by having an impedance that decreases with increasing frequency in the neighbourhood of $h\omega_0$. The Robinson stability criterion then follows.

When $\tau^{-1} > 0$, the beam is unstable because any accidental small synchrotron oscillation would grow exponentially. When $\tau^{-1} < 0$, the Robinson mechanism leads to exponential damping of any synchrotron oscillations of the beam. Robinson damping (or antidamping) can be rather strong. When the Robinson criterion is met, the synchrotron oscillation of the beam is 'Robinson damped' and this damping will help stabilize the beam against similar instabilities due to other impedance sources.

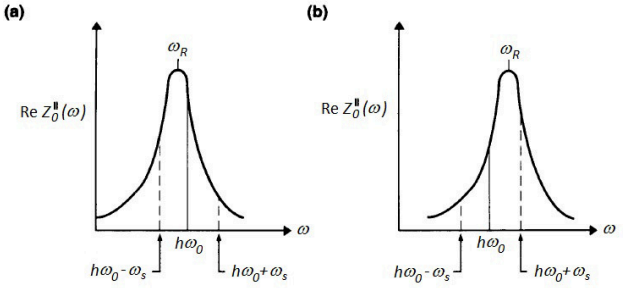


Fig. 17: Resonance frequency above and below transition

Table 4: Comparison of Robinson and strong head–tail instabilities

	Robinson instability	Strong head–tail instability
Dimension	Longitudinal	Transverse
Mode	$m = 0$	$m = 1$
Wakefield	Long-range	Short-range
Impedance	Sharply peaked	Broad-band
Model	One-particle	Two-particle
Threshold behaviour	No	Yes

18 Strong head–tail instability

We next introduce the strong head–tail instability, to be discussed using a two-macroparticle model. It was first observed and analysed at PEP. When the intensity exceeds a threshold, the beam becomes unstable. Below the threshold, the beam motion is perturbed but remains stable. Table 4 compares Robinson and strong head–tail instabilities.

The physical mechanism of the strong head–tail instability is closely related to beam break-up in linacs. Consider an idealized beam with two macroparticles, each with charge $Ne/2$ and each executing synchrotron oscillation. We assume that their synchrotron oscillations have equal amplitude but opposite phases. During time $0 < s/c < T_{\rm s}/2$, where $T_{\rm s} = 2\pi/\omega_{\rm s}$ is the synchrotron oscillation period, particle 1 leads particle 2; the equations of motion are

$$\begin{aligned} y_1'' + \left(\frac{\omega_\beta}{c}\right)^2 y_1 &= 0\,, \\ y_2'' + \left(\frac{\omega_\beta}{c}\right)^2 y_2 &= \frac{Nr_0W_0}{2\gamma C} y_1\,, \end{aligned} \tag{8}$$

where ω_β is the unperturbed betatron oscillation frequency, whether horizontal or vertical. During $T_{\rm s}/2 < s/c < T_{\rm s}$, we have the same equations with indices 1 and 2 switched. Then during $T_{\rm s} < s/c < 3T_{\rm s}/2$, Eq. (8) applies again, etc.

In writing down Eq. (8), we have assumed for simplicity that the wake function (integrated over the accelerator circumference), $W_1(z)$, is a constant, and yet it vanishes before the beam completes one revolution,

$$W_1(z) = \begin{cases} -W_0, & \text{if } 0 > z > -(\text{bunch length})\,, \\ 0, & \text{otherwise}\ . \end{cases} \tag{9}$$

The property of wake functions requires that $W_0 > 0$. This short-range wake function corresponds to a broad-band impedance.

The solution for y_1 in Eq. (8) is simply a free betatron oscillation,

$$\tilde{y}_1(s) = \tilde{y}_1(0) \mathrm{e}^{-\mathrm{i}\omega_\beta s/c}\,,$$

where

$$\tilde{y}_1 = y_1 + \mathrm{i}\frac{c}{\omega_\beta}\, y_1' \,.$$

Substituting $\tilde{y}_1(s)$ into the equation for y_2 yields

$$\tilde{y}_2(s) = \tilde{y}_2(0)\mathrm{e}^{-\mathrm{i}\omega_\beta s/c} + \mathrm{i}\frac{Nr_0W_0c}{4\gamma C\omega_\beta}\left[\frac{c}{\omega_\beta}\tilde{y}_1^*(0)\sin\frac{\omega_\beta s}{c} + \tilde{y}_1(0)s\mathrm{e}^{-\mathrm{i}\omega_\beta s/c}\right] . \tag{10}$$

The first two terms describe free betatron oscillation. The third term, proportional to s, is a resonantly driven response. This analysis is similar to the beam break-up instability.

Equation (10) can be simplified if $\omega_\beta \gg \omega_\mathrm{s}$. The second term can then be dropped. The solution during the period $0 < s/c < T_\mathrm{s}/2$ can be written in a matrix form,

$$\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}_{s=cT_\mathrm{s}/2} = \mathrm{e}^{-\mathrm{i}\omega_\beta T_\mathrm{s}/2}\begin{bmatrix}1 & 0\\ \mathrm{i}\Upsilon & 1\end{bmatrix}\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}_{s=0} ,$$

with a positive, dimensionless parameter,

$$\Upsilon = \frac{\pi Nr_0W_0c^2}{4\gamma C\omega_\beta\omega_\mathrm{s}} \,.$$

The time evolution during $T_\mathrm{s}/2 < s/c < T_\mathrm{s}$ can be obtained by exchanging indices 1 and 2. The total transformation for one full synchrotron period is

$$\begin{aligned}\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}_{cT_\mathrm{s}} &= \mathrm{e}^{-\mathrm{i}\omega_\beta T_\mathrm{s}}\begin{bmatrix}1 & \mathrm{i}\Upsilon\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ \mathrm{i}\Upsilon & 1\end{bmatrix}\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}_0\\ &= \mathrm{e}^{-\mathrm{i}\omega_\beta T_\mathrm{s}}\begin{bmatrix}1-\Upsilon^2 & \mathrm{i}\Upsilon\\ \mathrm{i}\Upsilon & 1\end{bmatrix}\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}_0 .\end{aligned}$$

As time evolves, the phasors $\tilde{y}_1$ and $\tilde{y}_2$ are repeatedly transformed by the 2×2 matrix of this map. The stability of the system is determined by the eigenvalues of this matrix. The two eigenvalues (a + mode and a $-$ mode) are

$$\lambda_\pm = \mathrm{e}^{\pm\mathrm{i}\phi} \quad , \quad \sin\frac{\phi}{2} = \frac{\Upsilon}{2} \,, \tag{11}$$

with eigenvectors

$$V_\pm = \begin{bmatrix}\pm\mathrm{e}^{\pm\mathrm{i}\phi/2}\\ 1\end{bmatrix} .$$

Stability requires $\phi =$ real, or

$$\Upsilon \le 2 \,.$$

For weak beams, $\Upsilon \ll 1$, we have $\phi \approx \Upsilon$. Near the instability, ϕ approaches π as Υ approaches 2.

After a moment of reflection, we see that the instability when $\Upsilon > 2$ causes a rather severe disruption of the beam, as seen by the fact that, during half a synchrotron period, the motion of the trailing particle has grown by an amount more than twice the amplitude of the free-oscillating leading particle. For $\Upsilon \le 2$, the growths made during the half synchrotron periods when the particle is trailing do not accumulate and the beam is stable. As the beam intensity increases so that $\Upsilon > 2$, the growths of the particles then do accumulate and bootstrap into an instability. This *threshold* behaviour is very different from the linac case, in which the beam head is always stable and the beam tail is always unstable. One can imagine that, by periodically exchanging the roles of leading and trailing particles, the two-particle beam is made more stable. The more frequently the roles are exchanged, the more stable is the beam, as evidenced by $\Upsilon \propto 1/\omega_\mathrm{s}$. Synchrotron oscillation is thus an effective stabilizing mechanism in circular

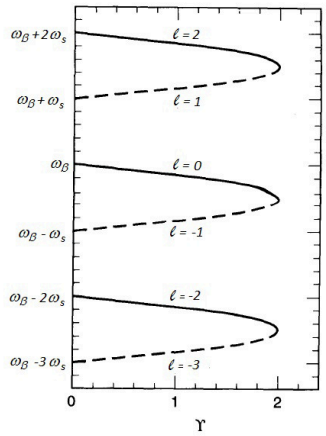


Fig. 18: Solid curves are the spectrum of the + mode; dashed curves are that of the − mode. Instability occurs at the point where the mode frequencies merge.

accelerators. Strong betatron focusing and a high beam energy also help stabilize the beam, as indicated by $\Upsilon \propto 1/(\gamma\omega_\beta)$.

In an accelerator, the beam signal comes from the beam position monitors that detect the centre of charge $y_1 + y_2$ of the beam, and it would be useful to examine its frequency spectrum. To do that, consider a two-particle beam in a pure eigenstate $V_\pm$ at time $s/c = 0$. In the stable region, the subsequent motion of the beam centre of charge is

$$(\tilde{y}_1 + \tilde{y}_2)(s) = \exp\left[-\mathrm{i}\left(\omega_\beta \mp \frac{\phi\omega_\mathrm{s}}{2\pi}\right)\frac{s}{c}\right]\sum_{\ell=-\infty}^{\infty} C_\ell \mathrm{e}^{-\mathrm{i}\ell\omega_\mathrm{s}s/c} ,$$

$$C_\ell = 2\mathrm{i}\Upsilon\,\frac{1\pm(-1)^\ell}{(2\pi\ell\mp\phi)^2}(1\mp\mathrm{e}^{\pm\mathrm{i}\phi/2}) .$$

The $\pm$ modes as observed by a beam position monitor therefore contain the following frequencies:

$$+\text{ mode:}\quad \omega_\beta + \ell\omega_\mathrm{s} - \frac{\phi}{2\pi}\,\omega_\mathrm{s} , \quad \ell = \text{even} ,$$

$$-\text{ mode:}\quad \omega_\beta + \ell\omega_\mathrm{s} + \frac{\phi}{2\pi}\,\omega_\mathrm{s} , \quad \ell = \text{odd} .$$

Note that each mode contains a multiplicity of frequencies when observed continuously in time.

For weak beams, the two macroparticles oscillate in phase in the + mode and out of phase in the − mode. As Υ increases, the mode frequencies shift and the particle motions become more complicated; each mode contains a combination of in-phase and out-of-phase motions. At the stability limit $\Upsilon = 2$, the frequencies of the two modes merge into each other and become imaginary, which means instability (Fig. 18).

To detect internal beam motion in addition to the centre of charge motion, one uses a streak camera. One such observation, made on the large electron–positron collider at CERN, is shown in Fig. 19. It shows the turn-by-turn pictures of a beam executing a transverse head–tail oscillation. The bunch is seen from the side and one observes a vertical head–tail oscillation ($\ell = 1$). The horizontal scale is 500 ps for the total image. The vertical scale is uncalibrated. Figure 19 shows the same bunch each turn from top to bottom.

The strong head–tail instability is one of the cleanest instabilities that can be observes in electron storage rings. One can measure the threshold beam intensity when the beam becomes unstable

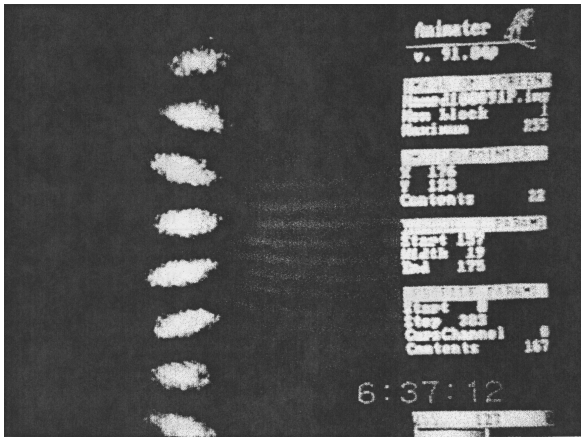


Fig. 19: Beam executing a transverse head–tail oscillation (Courtesy Albert Hofmann and Edouard Rossa, 1992)

transversely and associate the observation with $\Upsilon = 2$. Another approach is to measure the 'betatron frequency' (what is measured is the frequency of the $\ell = 0$ spectral line) as the beam intensity is varied. From our two-particle analysis, the initial slope of this frequency shift is

$$\left(\frac{\mathrm{d}\omega_\beta}{\mathrm{d}N}\right)_{N=0} = -\frac{\omega_\mathrm{s}}{2\pi}\left(\frac{\mathrm{d}\phi}{\mathrm{d}N}\right)_{N=0} = -\frac{r_0 W_0 c^2}{8\gamma C\omega_\beta}\,.$$

By measuring the instability threshold or the initial slope of the betatron frequency, information on the short-range wakefield or broad-band impedance can be obtained.

At the instability threshold, the measured betatron frequency has shifted by $\omega_\mathrm{s}/2$, according to the two-particle model. The measured value of $(\mathrm{d}\omega_\beta/\mathrm{d}N)_{N=0}$ can be used to predict the instability threshold N_th by

$$N_\mathrm{th} = -\frac{\omega_\mathrm{s}}{\pi}\frac{1}{(\mathrm{d}\omega_\beta/\mathrm{d}N)_{N=0}}\,.$$

The eventual instability threshold can thus be estimated by measuring ω_β at low beam intensities.

The two-particle model also predicts that the $\ell = 0$ frequency always shifts *down* as the beam intensity is increased. Physically, this is because, for short bunches, the sign of the wake force is such that the bunch tail is always deflected further away from the vacuum chamber axis if the beam is transversely displaced. With the head and the tail moving together in the $\ell = 0$ mode, the wake force acts as a defocusing effect and the mode frequency shifts down.

The centre of charge signal of the beam as a function of time after the beam receives an initial transverse kick can be analysed for a two-particle model. Figure 20 shows a result compared with experimental observation at PEP. The agreement indicates that the highly idealized two-particle model describes this instability mechanism remarkably well. The signal exhibits damping because of radiation damping.

19 Landau damping

Many collective instability mechanisms act on a high-intensity beam in an accelerator, demanding a wide range of sometimes conflicting stability conditions. Yet the beam as a whole seems basically stable, as evidenced by the existence of a wide variety of working accelerators. One reason for this fortunate outcome is *Landau damping*, which provides a natural stabilizing mechanism against collective instabilities, if particles in the beam have a small spread in their natural (synchrotron or betatron) frequencies.

The spread in ω_β has several sources. A dependence of ω_β on the energy of the particle, together with an energy spread in the beam, leads to a spread in ω_β. Non-linearities in the focusing system cause

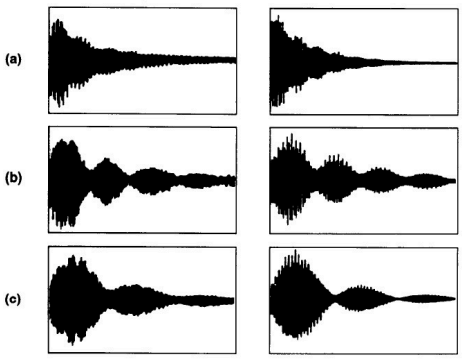


Fig. 20: Beam position monitor signal as a function of time after the beam is kicked. Left: PEP data with (a) $N/N_{\rm th} = 0.86$, (b) $N/N_{\rm th} = 0.93$, and (c) $N/N_{\rm th} = 0.988$. Right: two-particle model with (a) $\Upsilon/2 = 0.77$, (b) $\Upsilon/2 = 0.96$, and (c) $\Upsilon/2 = 0.99$.

a dependence of ω_β on the particle's betatron amplitude. A spread in betatron amplitudes then leads to a spread in ω_β.

The source of spread in $\omega_{\rm s}$ depends on whether the beam is bunched or unbunched. For bunched beams, a spread can result from non-linearity in the RF focusing voltage. For unbunched beams, the dependence of the revolution frequency on the particle energy plays a similar role.

Consider a simple harmonic oscillator with natural frequency ω driven by a sinusoidal force of frequency Ω,

$$\ddot{x} + \omega^2 x = A\cos\Omega t \ ,$$

with initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$. The solution is

$$x(t > 0) = -\frac{A}{\Omega^2 - \omega^2}(\cos\Omega t - \cos\omega t) \ . \tag{12}$$

The $\cos\Omega t$ term gives the main term responding to the driving force; the $\cos\omega t$ term comes from matching the initial conditions.

The explicit inclusion of the initial conditions plays an important role. Otherwise, one could have carelessly written the solution

$$x(t) = -\frac{A}{\Omega^2 - \omega^2}\cos\Omega t, \quad \text{or} \quad x(t) = -\frac{A}{\Omega^2 - \omega^2}{\rm e}^{-{\rm i}\Omega t} \ . \tag{13}$$

Equation (13) contains a singularity at $\Omega = \omega$, while Eq. (12) is well behaved there. This singularity is the source of subtleties and at this point is to be avoided. As we will see later, by applying some mathematical tricks, it is possible to bypass the explicit inclusion of the initial conditions and go straight to Eq. (13) but at this point we will stay with Eq. (12).

Consider now an ensemble of oscillators (each oscillator represents a single particle in the beam) which have a spectrum $\rho(\omega)$ satisfying $\int_{-\infty}^{\infty} {\rm d}\omega\rho(\omega) = 1$. Now subject this ensemble of particles to the driving force $A\cos\Omega t$ with all particles starting with initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$. The ensemble average response is

$$\langle x\rangle(t > 0) = -\int_{-\infty}^{\infty} {\rm d}\omega\rho(\omega)\frac{A}{\Omega^2 - \omega^2}(\cos\Omega t - \cos\omega t) \ .$$

For simplicity, let us consider a narrow beam spectrum around a frequency ω_x and a driving frequency near the spectrum, i.e., $\Omega \approx \omega_x$. The beam response is then

$$\langle x\rangle(t) = -\frac{A}{2\omega_x}\int_{-\infty}^{\infty} {\rm d}\omega\rho(\omega)\frac{1}{\Omega - \omega}(\cos\Omega t - \cos\omega t) \ .$$

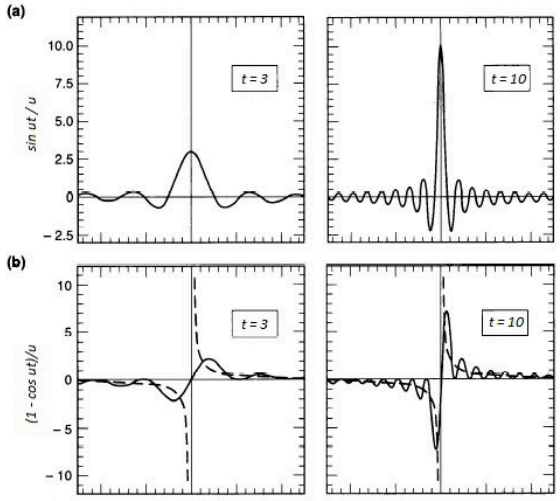


Fig. 21: The functions $\sin(ut)/u$, $(1-\cos ut)/u$ are shown in (a), (b) for two values t = 3 and 10. The dashed curves in (b) are for the function $1/u$. The sole function of $(1-\cos u)$ in (b) is to suppress the singularity at $u = 0$.

Changing variable from ω to $u = \omega - \Omega$ leads to

$$\begin{aligned}\langle x\rangle(t) &= \frac{A}{2\omega_x}\int_{-\infty}^{\infty} \mathrm{d}u \frac{\rho(u+\Omega)}{u}[\cos\Omega t - \cos(\Omega t + ut)] \\ &= \frac{A}{2\omega_x}\left[\cos\Omega t \int_{-\infty}^{\infty} \mathrm{d}u\rho(u+\Omega)\frac{1-\cos ut}{u} + \sin\Omega t\int_{-\infty}^{\infty}\mathrm{d}u\rho(u+\Omega)\frac{\sin ut}{u}\right] .\end{aligned}$$

All integrals are well behaved at $u = 0$.

The beam response contains a $\cos\Omega t$ term and a $\sin\Omega t$ term, but their coefficients are time dependent. The next step is to show that those coefficients approach well behaved limits. To do so, one first observes

$$\begin{aligned}\lim_{t\to\infty}\frac{\sin ut}{u} &= \pi\delta(u) , \\ \lim_{t\to\infty}\frac{1-\cos ut}{u} &= \mathrm{P.V.}\left(\frac{1}{u}\right) .\end{aligned}$$

The proof is illustrated in Fig. 21.

If we are not interested in the transient effects immediately following the onset of the driving force, we obtain

$$\langle x\rangle(t) = \frac{A}{2\omega_x}\left[\cos\Omega t\ \mathrm{P.V.}\int \mathrm{d}\omega\frac{\rho(\omega)}{\omega-\Omega} + \pi\rho(\Omega)\sin\Omega t\right] .$$

This expression explicitly contains a $\cos\Omega t$ term and a mysterious $\sin\Omega t$ term.

The sign of the $\cos\Omega t$ term relative to the driving force depends on the sign of P.V.$\int \mathrm{d}\omega\rho(\omega)/(\omega-\Omega)$. A system is referred to as 'capacitive' or 'inductive' based on whether its sign is positive or negative.

The $\sin\Omega t$ term has a definite sign relative to the driving force because $\rho(\Omega)$ is always positive. In particular, $\mathrm{d}\langle x\rangle/\mathrm{d}t$ is always in phase with the force, indicating that work is being done on the system. The system always reacts to the force 'resistively'.

To proceed, write the beam response in complex notation,

$$\text{driving force} = A\mathrm{e}^{-\mathrm{i}\Omega t} ,$$

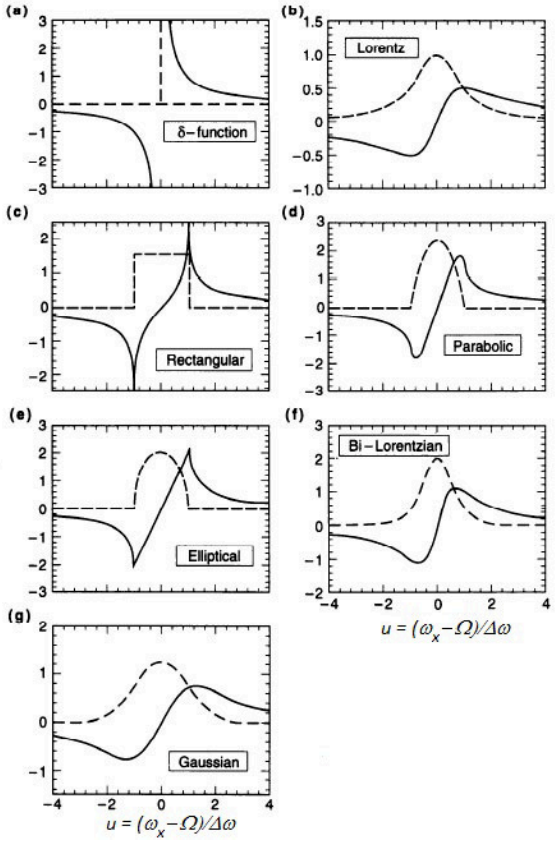


Fig. 22: The functions $f(u)$ (solid curves) and $g(u)$ (dashed curves) for various spectral distributions

$$\langle x\rangle = \frac{A}{2\omega_x\Delta\omega}\mathrm{e}^{-\mathrm{i}\Omega t}\left[f(u)+\mathrm{i}g(u)\right] ,$$

where $u=(\omega_x-\Omega)/\Delta\omega$ with $\Delta\omega$ the width of the spectral spread, and

$$f(u) = \Delta\omega\ \mathrm{P.V.}\int \mathrm{d}\omega\frac{\rho(\omega)}{\omega-\Omega} ,$$
$$g(u) = \pi\Delta\omega\ \rho(\omega_x-u\Delta\omega) .$$

The dimensionless complex quantity $f+\mathrm{i}g$ is the *beam transfer function* (Fig. 22).

For the δ-function spectrum, there is no frequency spread, Landau damping is lost,

$$f(u)=\frac{1}{u}, \quad \text{and} \quad g(u)=\pi\delta(u) .$$

For the Lorentz spectrum,

$$f(u)=\frac{u}{1+u^2}, \quad \text{and} \qquad g(u)=\frac{1}{1+u^2} .$$

We now introduce a mathematical trick. It turns out that one can 'derive' the same result by venturing with Eq. (13). In complex notation, Eq. (13) gives

$$\langle x\rangle = \frac{A}{2\omega_x}\mathrm{e}^{-\mathrm{i}\Omega t}\int \mathrm{d}\omega\frac{\rho(\omega)}{\omega-\Omega} .$$

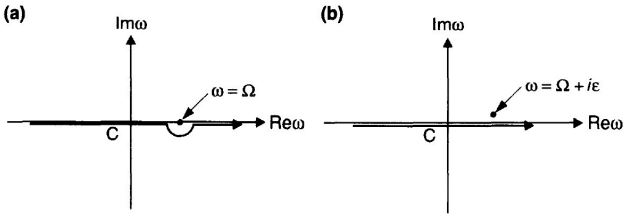


Fig. 23: Contour of integration

Our detailed examinations provides a well defined way to deal with the otherwise undefined integral, i.e.,

$$\int \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} \quad \rightarrow \quad \mathrm{P.V.} \int \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} + \mathrm{i}\pi\rho(\Omega) \; , \tag{14}$$

or, more symbolically,

$$\frac{1}{\omega - \Omega} \quad \rightarrow \quad \mathrm{P.V.} \left(\frac{1}{\omega - \Omega} \right) + \mathrm{i}\pi\delta(\omega - \Omega) \; .$$

Again, it is necessary to include an out-of-phase term—with a definite sign—as evidenced by the imaginary term $\mathrm{i}\pi\rho(\Omega)$, even though the expression on the left-hand side seems to be for a real quantity.

The right-hand side of Eq. (14), in fact, is equal to the left-hand side, provided one takes the integration to be executed in the complex ω-plane and the contour of integration, C, is as illustrated in Fig. 23(a). The connection (Eq. (14)) now reads

$$\int \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} \quad \rightarrow \quad \int_C \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} \; .$$

The straight line portion of C gives the principal value term in $\langle x \rangle$ and the semicircular portion gives the pole contribution $\mathrm{i}\pi\rho(\Omega)$.

Equivalently, one could consider the integration along the real axis of the ω-plane, but move the pole at $\omega = \Omega$ up by an infinitesimal amount,

$$\int d\omega \frac{\rho(\omega)}{\omega - \Omega} \quad \rightarrow \quad \int_{-\infty}^{\infty} \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega - \mathrm{i}\epsilon} \; ,$$

or

$$\frac{1}{\omega - \Omega} \quad \rightarrow \quad \frac{1}{\omega - \Omega - \mathrm{i}\epsilon} \; ,$$

or simply

$$\Omega \quad \rightarrow \quad \Omega + \mathrm{i}\epsilon \; . \tag{15}$$

It is now a matter of taste whether to regard our main conclusion (Eq. (14)) as a result of a simple derivation starting with Eq. (13) and then make a profound connection (Eq. (15)), or to regard it as a result of a detailed calculation that takes initial conditions into account.

20 One-particle model for bunched beams – transverse

The results obtained in the previous section applied to circular accelerators lead to Landau damping of collective instabilities. To demonstrate this for a bunched beam, consider a one-particle model, except that now the N individual particles have a spread in their natural frequencies. The fact that they form one macroparticle even though they have different frequencies is a result of the bunch executing a collective motion.

The driving force on the individual particles comes from the centre of charge displacement of the beam as a whole, $\langle y \rangle$, through the wakefield. For a single particle whose betatron frequency is ω,

$$y''(s) + \left(\frac{\omega}{c}\right)^2 y(s) = -\frac{Nr_0}{\gamma C} \sum_{k=1}^{\infty} \langle y \rangle (s - kC)\, W_1(-kC) \;,$$

Consider the situation when y-motion of the macroparticle is just at the edge of exponential growth, owing to a collective instability. We have

$$\langle y \rangle (s) = B \mathrm{e}^{-\mathrm{i}\Omega s/c} \;, \tag{16}$$

where Ω carries an imaginary part $\mathrm{i}\epsilon$, where ϵ is infinitesimally positive.

It is not very interesting to search for damped, stable solutions. Finding stable solutions does not assure beam stability, but finding one unstable solution reveals the beam to be unstable.

We now have

$$y''(s) + \left(\frac{\omega}{c}\right)^2 y(s) = -\frac{BNr_0}{\gamma C} \mathcal{W} \mathrm{e}^{-\mathrm{i}\Omega s/c} \;,$$

where

$$\mathcal{W} = \sum_{k=1}^{\infty} W_1(-kC) \mathrm{e}^{\mathrm{i}\omega_\beta k T_0} \;,$$

or, in terms of impedance,

$$\mathcal{W} = -\frac{\mathrm{i}}{T_0} \sum_{p=-\infty}^{\infty} Z_1^{\perp}(p\omega_0 + \omega_\beta) \;.$$

We have assumed that the mode frequency shift is small so that $\Omega \approx \omega_\beta$, where ω_β is the centre of the beam frequency spectrum.

The beam is driven by a sinusoidal driving force. Our analysis of Landau damping gives the beam response,

$$\langle y \rangle = -\frac{BNr_0 \mathcal{W} c}{2\omega_\beta \gamma T_0} \mathrm{e}^{-\mathrm{i}\Omega s/c} \left[\mathrm{P.V.} \int \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} + \mathrm{i}\pi\rho(\Omega) \right] \;.$$

But we had already assumed that the collective beam motion is given by Eq. (16). This means that the mode frequency Ω is not arbitrary. For the beam motion to be non-trivial, Ω must satisfy a self-consistency condition, the *dispersion relation*,

$$1 = -\frac{Nr_0 \mathcal{W} c}{2\omega_\beta \gamma T_0} \left[\mathrm{P.V.} \int \mathrm{d}\omega \frac{\rho(\omega)}{\omega - \Omega} + \mathrm{i}\pi\rho(\Omega) \right] \;,$$

or

$$-\frac{Nr_0 \mathcal{W} c}{2\omega_\beta \gamma T_0 \Delta\omega} = \frac{1}{f(u) + \mathrm{i}g(u)} \;.$$

If the beam does not have a natural frequency spread, we have $f(u) = 1/u$, $g(u) = 0$. The complex mode frequency shift is found to be

$$(\Omega - \omega_\beta)_{\text{no Landau damping}} = \frac{Nr_0 c \mathcal{W}}{2\omega_\beta \gamma T_0} \;.$$

We shall designate this quantity as ξ_1; it contains essentially the beam intensity, multiplied by the impedance, divided by the focusing strength and the magnetic rigidity.

For a beam with natural frequency spread, the dispersion relation is

$$-\frac{\xi_1}{\Delta\omega} = \frac{1}{f(u) + \mathrm{i}g(u)} \;. \tag{17}$$

The left-hand side of Eq. (17) contains information about the beam intensity and the impedance. The right-hand side contains information about the beam frequency spectrum. For a given impedance, the left-hand side is obtained by calculating the complex mode frequency shift ξ_1 in the absence of Landau damping. Without Landau damping, the stability condition is simply Im $\xi_1 < 0$.

Once its left-hand side is obtained, Eq. (17) can, in principle, be used to determine Ω in the presence of Landau damping when the beam is at the edge of instability. However, the exact value of Ω is not useful. The useful question to ask is under what conditions the beam becomes unstable, regardless of the exact value of Ω. Equation (17) can be used in a reversed manner to address this question. To do so, consider the real parameter $u = (\omega_\beta - \Omega)/\Delta\omega$ and observe the locus traced out in the complex $\mathcal{D}_1$-plane as u is scanned from ∞ to $-\infty$, where

$$\mathcal{D}_1 = \frac{1}{f(u) + \mathrm{i}g(u)} \; .$$

This locus defines a *stability boundary diagram* (Fig. 24). The left-hand side of Eq. (17), a complex quantity, is then plotted in the complex $\mathcal{D}_1$-plane as a single point. If this point lies on the locus, it means the solution of Ω for Eq. (17) is real, and this ξ_1 value is such that the beam is just at the edge of instability. If it lies on the inside of the locus (the side that contains the origin of the $\mathcal{D}_1$-plane), the beam is stable. If it lies on the outside of the locus, the beam is unstable.

The dispersion relation is particularly simple for the Lorentz spectrum (Fig. 24(b)),

$$\Omega = \omega_\beta + \xi_1 - \mathrm{i}\Delta\omega \; .$$

The stability condition Im$\Omega < 0$ therefore becomes

$$\mathrm{Im}\xi_1 < \Delta\omega \; .$$

The fact that the stable region is always enlarged by the frequency spread demonstrates the Landau damping mechanism. Its origin can be traced back to the fact that $g(u)$ is always positive, which in turn comes from the fact that the beam continues to absorb energy from the driving force without having to let $\langle y \rangle$ grow.

For a given spectral shape, the tolerable $\xi_1 \propto \Delta\omega$; the larger the frequency spread, the stronger the Landau damping. For a given $\Delta\omega$, the effectiveness of Landau damping is different for different spectral shapes. The Lorentz spectrum, having a long distribution tail, is most forgiving, while the δ-function spectrum is not effective.

For practical accelerator operations, there may be information on the value of the half-width-at-half-height $\Delta\omega_{\frac{1}{2}}$, but not enough detailed information on the shape of the frequency spectrum. For those applications, we introduce a simplified stability criterion,

$$|\xi_1| = \frac{Nr_0c}{2\omega_\beta\gamma T_0^2} \left| \sum_{p=-\infty}^{\infty} Z_1^\perp(p\omega_0 + \omega_\beta) \right| < \frac{1}{\sqrt{3}} \Delta\omega_{\frac{1}{2}} \; , \tag{18}$$

where the factor $1/\sqrt{3}$ is chosen so that it coincides with the semicircular portion of the boundary for the elliptical spectrum. The stability diagram of this simplified model is shown in Fig. 24(h).

Equation (18) says that if the mode frequency shift or growth rate, calculated without Landau damping, is larger than the frequency spread of the beam, Landau damping will not rescue the beam from instability.

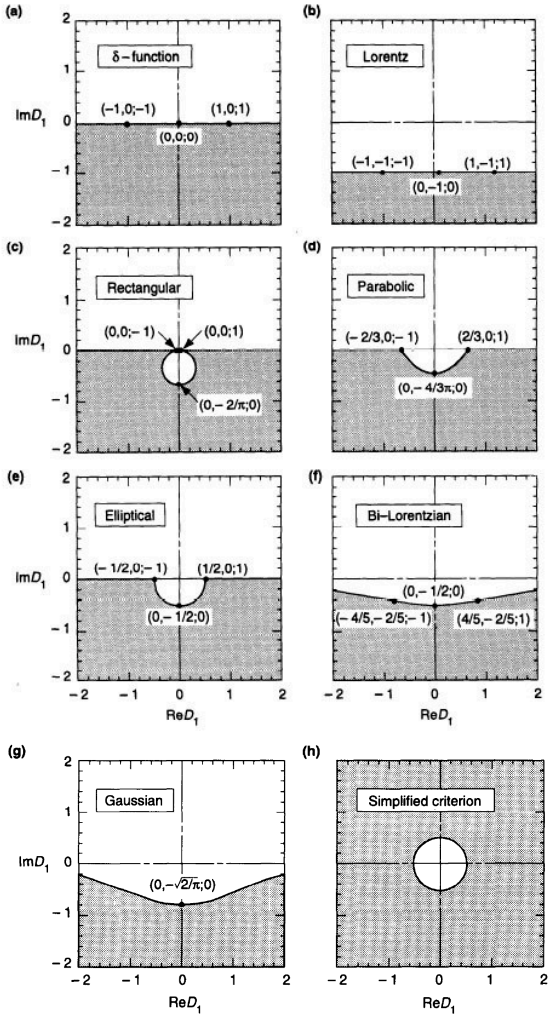


Fig. 24: The stability boundary diagrams for various spectra. Shaded regions are unstable. The coordinates labelled refer to (Re$\mathcal{D}_1$, Im$\mathcal{D}_1$; u). The value of u can be used to obtain Ω. (a) δ-function spectrum, no Landau damping. (h) is the simplified criterion (Eq. (18)).

21 One-particle model for bunched beams – longitudinal

A similar analysis can also be performed for the longitudinal Robinson instability using a one-particle model,

$$z''(s) + \left(\frac{\omega_{\rm s}}{c}\right)^2 z(s) = \frac{Nr_0\eta}{\gamma C}\sum_{k=1}^{\infty}\left[\langle z\rangle(s) - \langle z\rangle(s-kC)\right]W_0''(-kC)$$
$$= \frac{Nr_0\eta}{\gamma C}B{\rm e}^{-{\rm i}\Omega s/c}\mathcal{W}\,,$$

where we have introduced

$$\langle z\rangle(s) = B{\rm e}^{-{\rm i}\Omega s/c}$$

and

$$\mathcal{W} = \sum_{k=1}^{\infty}\left(1 - {\rm e}^{{\rm i}\omega_{\rm s}kT_0}\right)W_0''(-kC)$$

$$= \frac{\mathrm{i}}{C} \sum_{p=-\infty}^{\infty} \left[p\omega_0 Z_0^{\parallel}(p\omega_0) - (p\omega_0 + \omega_\mathrm{s}) Z_0^{\parallel}(p\omega_0 + \omega_\mathrm{s}) \right] \ .$$

Self-consistency then gives rise to a dispersion relation

$$\frac{N r_0 \eta \mathcal{W} c^2}{2\omega_\mathrm{s} \gamma C \Delta\omega} = \frac{1}{f(u) + \mathrm{i} g(u)} \ ,$$

similar to the transverse case except that the frequency spectrum now refers to synchrotron frequency, and the complex mode frequency shift in the absence of Landau damping is

$$\xi_1 = -\frac{N r_0 \eta \mathcal{W} c^2}{2\omega_\mathrm{s} \gamma C} \ .$$

The simplified stability criterion reads

$$|\xi_1| = \frac{N r_0 \eta c^2}{2\omega_\mathrm{s} \gamma C^2} \left| \sum_{p=-\infty}^{\infty} \left[p\omega_0 Z_0^{\parallel}(p\omega_0) - (p\omega_0 + \omega_\mathrm{s}) Z_0^{\parallel}(p\omega_0 + \omega_\mathrm{s}) \right] \right| < \frac{1}{\sqrt{3}} \Delta\omega_{\frac{1}{2}} \ .$$

The conclusion that the longitudinal Landau damping behaves analogously to the transverse case, however, is valid only for *bunched* beams for which $\omega_\mathrm{s} \neq 0$. The analyses depend on the assumption that the mode frequency shift $|\Omega|$ is small compared with the unperturbed natural frequency $\omega_\beta, \omega_\mathrm{s}$. For unbunched beams, $\omega_\mathrm{s} = 0$, the longitudinal analysis gives results very different from its transverse counterpart.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Bench Measurements and Simulations of Beam Coupling Impedance

U. Niedermayer
Institut für Theorie Elektromagnetischer Felder, Technische Universität Darmstadt, Germany

Abstract
After a general introduction, the basic principles of wake-field and beam-coupling-impedance computations are explained. This includes time domain, frequency domain, and methods that do not include excitations by means of a particle beam. The second part of this paper deals with radio frequency bench measurements of beam coupling impedances. The general procedure of the wire measurement is explained, and its features and limitations are discussed.

Keywords
Wake field; beam coupling impedance; electromagnetic simulations; bench measurements; wire measurements.

1 Introduction: Maxwell's equations, wakes and impedance

A complete macroscopic description of electromagnetic (EM) fields as function of position $\vec{r} \in \Omega \subset \mathbb{R}^3$ and time $t \in \mathbb{R}$ is given by Maxwell's equations

$$\nabla \times \vec{E}(\vec{r},t) = -\partial_t \vec{B}(\vec{r},t) \tag{1a}$$
$$\nabla \times \vec{H}(\vec{r},t) = \vec{J}_s(\vec{r},t) + \vec{J}(\vec{r},t) + \partial_t \vec{D}(\vec{r},t) \tag{1b}$$
$$\nabla \cdot \vec{D}(\vec{r},t) = \varrho_s(\vec{r},t) \tag{1c}$$
$$\nabla \cdot \vec{B}(\vec{r},t) = 0 \tag{1d}$$

and material equations

$$\vec{D}(\vec{r},t) = \varepsilon(\vec{r})\vec{E}(\vec{r},t) \tag{2a}$$
$$\vec{B}(\vec{r},t) = \mu(\vec{r})\vec{H}(\vec{r},t) \tag{2b}$$
$$\vec{J}(\vec{r},t) = \kappa(\vec{r})\vec{E}(\vec{r},t)\ , \tag{2c}$$

where ϱ_s and $\vec{J}_s$ denote the source charge and current densities, respectively. The material distribution is given by the permittivity ε, the conductivity κ and the permeability μ, which are assumed to be linear and isotropic. At first, the material is also assumed to be non-dispersive. By means of Gauss' and Stokes' theorems, Maxwell's equations can also be written in integral form, which is more general, since the differentiability requirements of the field vector functions can be relaxed.

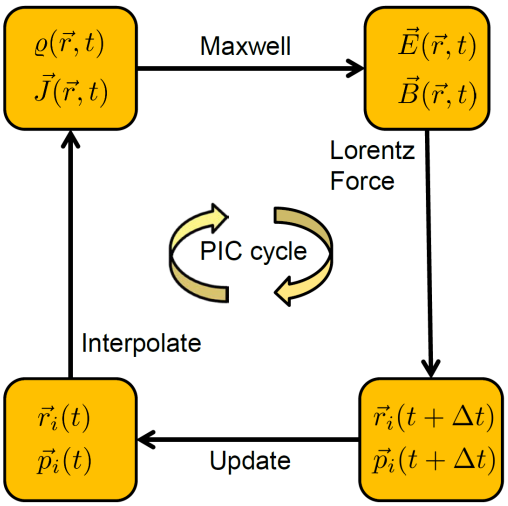
The force acting on a point-like charged particle $i = 1, \ldots, N$ is

$$\vec{F}_i(t) = q\left(\vec{E}(\vec{r}_i,t) + \vec{v}_i(t) \times \vec{B}(\vec{r}_i,t)\right), \tag{3}$$

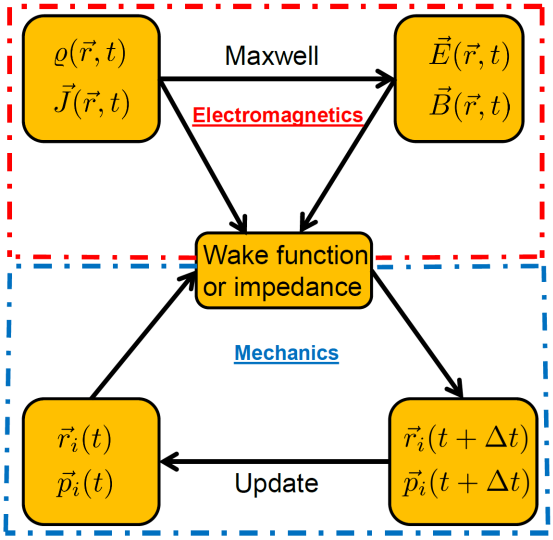
where q and $\vec{v}_i$ are the particle's charge and velocity, respectively. For a description of particle motion in both their own and external fields, Maxwell's equations have to be coupled with the equations of mechanics,

$$\vec{F}_i(t) = \partial_t \vec{p}_i(t) \tag{4}$$
$$\vec{p}_i(t) = \gamma_i(t) m_i \vec{v}_i(t)\ , \tag{5}$$

https://doi.org/10.23730/CYRSP-2017-003.81



(a) Description of particle motion in EM-fields. For full analytic self-consistency $\Delta t \to 0$ is required. Numerical approaches can, dependent on the time scale of the problem, allow finite Δt and still be self-consistent.



(b) Description of particle motion with wake fields. The EM forces are precomputed using a rigid beam as source. For synchrotrons, the time step is usually one revolution period, which is a good approximation for multi-turn phenomena.

Fig. 1: PIC simulation loop versus wake-field approach

where the relativistic mass and velocity factors are $\gamma = (1-\beta^2)^{-1/2}$ and $\beta = |\vec{v}|/c$, and $\vec{p}_i(t)$ is the ith particle's momentum. In order to simulate the dynamics of many particles self-consistently, mostly particle in cell (PIC) or Vlasov–Maxwell solvers are used (cf. Fig. 1(a)). The wake function approach is different, since Maxwell's equations are entirely decoupled from the equations of motion (cf. Fig. 1(b)). Therefore one obtains self-consistency only for phenomena which evolve slower than the wake-field kicks.

1.1 Wake functions

Usually one makes two assumptions that decouple mechanics from electromagnetics (cf. Fig. 1(b)).

1. **Rigid Beam Approximation** Although the leading charge loses energy, its velocity remains unchanged. This is exactly fulfilled for an ultrarelativistic beam which carries infinite energy.
2. **Kick Approximation** The wake force continuously acting on the trailing charge is lumped in a single kick *after* the passage through the device. This means that the trailing charge is also assumed to be rigid during the passage.

These approximations are justified by the different time scales of the particle passage (fast) and the evolution of wake-field effects (slow).

We define the wake function as (see also, for example, Ref. [1, 2])

$$\begin{aligned}\vec{W}(\vec{r}_2^{\perp},\vec{r}_1^{\perp},s) :&= \frac{1}{q_1 q_2}\int_{-\infty}^{\infty}\vec{F}\left(\vec{r}_2, z_2, t=\frac{z_2+s}{v}\right)\mathrm{d}z_2 \\ &= \frac{1}{q_1}\int_{-\infty}^{\infty}\left[\vec{E}+\vec{v}\times\vec{B}\right]\left(\vec{r}_2, z_2, t=\frac{z_2+s}{v}\right)\mathrm{d}z_2\,, \end{aligned} \tag{6}$$

such that a positive value indicates momentum or energy gain for the test charge. The integral in Eq. (6) exists only if the assumed infinitely long pipe connections (see Fig. 2) do not cause any wake fields, which requires the following conditions:

– smooth pipe (no geometric wake fields);
– perfectly conducting pipe (no resistive wake fields);
– ultrarelativistic beam (no space-charge wake fields).

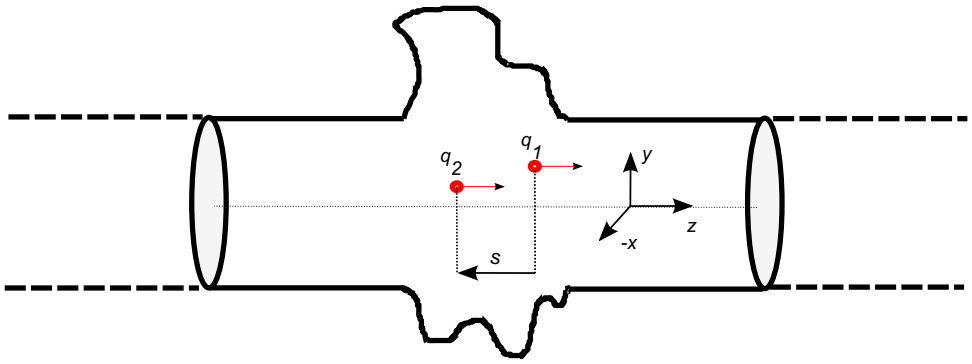


Fig. 2: Longitudinal cut through an accelerator structure and centred coordinate system. The charge q_1 is usually referred to as source or leading charge and q_2 is referred to as test or trailing charge.

When all these conditions are fulfilled, the infinite integration in Eq. (6) can be replaced by a finite one, since the scattered fields from the 3D region decay in the pipe below the waveguide cutoff frequency. Above the cutoff frequency, waveguide modes do not interact with the particle beam in average, since the fields are periodic and $v_{\mathrm{beam}} \leq c < v_{\mathrm{phase}}^{\mathrm{mode}}$. An estimate of the decay length for a given threshold can be found in Ref. [3]. For non-ultrarelativistic velocity, sophisticated boundary conditions are required for the entry and exit of the beam in a finite-sized computational domain, see [4]. In order to ensure that the integral in Eq. (6) is finite, even if the third assumption is not fulfilled, the integration can be performed over a finite length and the space-charge interaction in the infinitely long pipe can be described by a space-charge wake function per length. Note that the three components of Eq. (6) are not independent, but connected by the Panofsky–Wenzel theorem [5], which can be conveniently expressed as $\nabla' \times \vec{W}(\vec{r}_2^{\perp}, s) = 0$, where the curl acts on the relative position of the trailing charge, i.e. $\nabla' = (\partial_{x_2}, \partial_{y_2}, -\partial_s)^T$. The longitudinal wake function reads

$$W_{\parallel}(s) = \frac{1}{q_1 q_2} \int_{-\infty}^{\infty} \vec{E}\left(\vec{r}_2 = 0, z_2, t = \frac{z_2 + s}{v}\right) \mathrm{d}z_2 \tag{7}$$

and is directly connected with the energy loss of the trailing charge by $\Delta E(s) = -q^2 W_{\parallel}(s)$. The energy loss of a test charge caused not only by a single charge but by a whole bunch can be found by convolution with the beam distribution as

$$W_{\parallel}^{\mathrm{pot}}(s) = \int_{-\infty}^{\infty} W_{\parallel}(s')\lambda(s - s')\,\mathrm{d}s' \tag{8}$$

and is called the wake potential. If λ is normalized to 1, the energy loss of a test charge becomes $\Delta E(s) = -Nq^2 W_{\parallel}^{\mathrm{pot}}(s)$, where N is the number of particles in the bunch.

In order to quantify the transverse deflection of the beam by wake fields one makes the assumption of small displacements from the beam axis ($d_{x/y_{1/2}}$ of the leading and trailing charge in the x/y direction, respectively) and defines

$$W_{\perp,x}^{\mathrm{drive}}(s) = \frac{1}{q_1 d_{x_1}} \int_{-\infty}^{\infty} \left[\vec{E} + \vec{v} \times \vec{B}\right] \left(\vec{r}_2^{\perp} = 0, z, \frac{z_2 + s}{v}\right) \mathrm{d}z_2 \tag{9}$$

$$W_{\perp,x}^{\mathrm{det}}(s) = \frac{1}{q_1 d_{x_2}} \int_{-\infty}^{\infty} \left[\vec{E} + \vec{v} \times \vec{B}\right] \left(\vec{r}_2^{\perp} = d_{x_2}\vec{e}_x, z, \frac{z_2 + s}{v}\right) \mathrm{d}z_2 \tag{10}$$

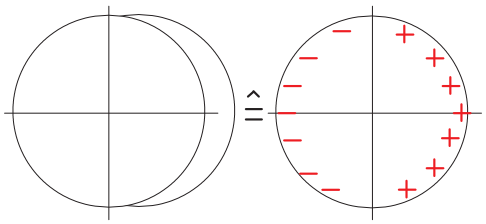


Fig. 3: Illustration of a coherent transverse oscillation (left-hand side) represented by a dipole moment (right-hand side).

in units of $\mathrm{VA}^{-1}\mathrm{s}^{-1}\mathrm{m}^{-1}$). The linearized transverse kick is thus $\Delta p_x(s) = q^2/\beta c(d_{x_1} W_{\perp,x}^{\mathrm{drive}}(s) + d_{x_2} W_{\perp,x}^{\mathrm{det}}(s))$. The dipolar or driving wake acts coherently, since the force experienced by a test charge does not depend on the test charge's position in a dipole field. Contrarily, the detuning wake acts incoherently, as it depends on the displacement of the test charge linearly.

If the beam is not ultrarelativistic, space-charge effects need to be taken into account and the transverse beam dimensions need to be considered as finite. We assume the beam to be transversely uniform of radius a. A coherent displacement can be written as

$$\sigma(\vec{r}^{\perp}) = \frac{1}{\pi a^2}\Theta(a - |\vec{r}^{\perp} - \vec{d}^{\perp}|)\,, \tag{11}$$

where $\vec{d}^{\perp}$ is the transverse displacement vector and Θ denotes the Heaviside step function. This displacement can be approximated in polar coordinates ϱ, φ by (Ref. [6])

$$\sigma(\varrho, \varphi) \approx \frac{1}{\pi a^2}(\Theta(a - \varrho) + \delta(a - \varrho)(d_{\mathrm{x}} \cos\varphi + d_{\mathrm{y}} \sin\varphi) + \cdots) \tag{12}$$

which in first order is a dipole moment, see Fig. 3.

1.2 Frequency domain

In this framework, the following convention for the Fourier transform is used:

$$\underline{F}(\omega) = \mathcal{F}\{f(t)\}(\omega) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t. \tag{13}$$

The inverse transform reads

$$f(t) = \mathcal{F}^{-1}\{\underline{F}(\omega)\}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \underline{F}(\omega)\mathrm{e}^{+\mathrm{i}\omega t}\,\mathrm{d}\omega\,, \tag{14}$$

and between the two domains Plancherel's theorem holds, $2\pi(f, g) = (\underline{G}, \underline{F})$, where the brackets denote a scalar product $(f, g) = \int f g^*\,\mathrm{d}x$ and * is complex conjugation. The duration and bandwidth of a mean value free signal $f(t)$ are defined with $\|\cdot\|_2 = \sqrt{(\cdot,\cdot)}$ as (see Ref. [7])

$$T = \frac{\|tf(t)\|_2}{\|f(t)\|_2}, \quad B = \frac{\|\omega\underline{F}(\omega)\|_2}{\|\underline{F}(\omega)\|_2}\,. \tag{15}$$

The two fulfil the Küpfmüller uncertainty principle, see Ref. [8], $T \cdot B \geq 1/2$, where equality is achieved for a Gaussian pulse. Applying Eq. (13) to Eqs. (1), Maxwell's equations read in the frequency domain

$$\nabla \times \underline{\vec{E}}(\vec{r}, \omega) = -\mathrm{i}\omega\underline{\vec{B}}(\vec{r}, \omega) \tag{16a}$$

$$\nabla \times \underline{\vec{H}}(\vec{r}, \omega) = \underline{\vec{J}}_s(\vec{r}, \omega) + \underline{\vec{J}}(\vec{r}, \omega) + \mathrm{i}\omega\underline{\vec{D}}(\vec{r}, \omega) \tag{16b}$$

$$\nabla \cdot \underline{\vec{D}}(\vec{r},\omega) = \underline{\varrho}_s(\vec{r},\omega) \tag{16c}$$

$$\nabla \cdot \underline{\vec{B}}(\vec{r},\omega) = 0\,, \tag{16d}$$

where the material relations can easily include dispersive materials as

$$\underline{\vec{D}}(\vec{r},\omega) = \varepsilon_0 \underline{\varepsilon}_r(\vec{r},\omega)\underline{\vec{E}}(\vec{r},\omega) \tag{17a}$$

$$\underline{\vec{B}}(\vec{r},\omega) = \mu_0 \underline{\mu}_r(\vec{r},\omega)\underline{\vec{H}}(\vec{r},\omega) \tag{17b}$$

$$\underline{\vec{J}}(\vec{r},\omega) = \kappa(\vec{r},\omega)\underline{\vec{E}}(\vec{r},\omega)\,. \tag{17c}$$

Combining Eqs. (16) leads to the curl–curl equation

$$\nabla \times \underline{\mu}^{-1} \nabla \times \underline{\vec{E}} + \mathrm{i}\omega\kappa\underline{\vec{E}} - \omega^2\varepsilon\underline{\vec{E}} = -\mathrm{i}\omega\underline{\vec{J}}_s \tag{18}$$

and the continuity equation

$$\nabla \cdot (\underline{\vec{J}}_s + \kappa\underline{\vec{E}}) + \mathrm{i}\omega\underline{\varrho}_s = 0\,. \tag{19}$$

The beam current density is modelled as a convection current density $\underline{\vec{J}}_s = \underline{\varrho}_s v\vec{e}_z$. Therefore, the spatial Fourier correspondence $\partial_z \to -\mathrm{i}k_z$ is given for the source fields (beam in free space) by $k_z = \omega/v$. In a longitudinally homogeneous and smooth two-dimensional (2D) structure, this property must hold also for the fields scattered back from the wall.

The beam coupling impedance is defined as the Fourier transform of the wake function

$$\underline{\vec{Z}}(\vec{r}_1^{\perp},\vec{r}_2^{\perp},\omega) = -\int_{-\infty}^{\infty} \vec{W}(\vec{r}_1^{\perp},\vec{r}_2^{\perp},s)\mathrm{e}^{-\mathrm{i}\omega s/v}\,\frac{\mathrm{d}s}{v} = -\frac{1}{q_1 q_2}\int_{-\infty}^{\infty} \underline{\vec{F}}(\vec{r}_1^{\perp},\vec{r}_2^{\perp},z,\omega)\mathrm{e}^{+\mathrm{i}\omega z/v}\,\mathrm{d}z\,, \tag{20}$$

where $\underline{\vec{F}}(\omega) = q(\underline{\vec{E}}(\omega) + \vec{v}\times\underline{\vec{B}}(\omega))$ is the spectral density of the force. If the beam has a finite transverse size, the longitudinal impedance can be written as (see Ref. [9])

$$\underline{Z}_{\parallel}(\omega) = -\frac{1}{q^2}\int_{\mathrm{beam}} \underline{\vec{E}}\cdot\underline{\vec{J}}_s^{*}\,\mathrm{d}V\,. \tag{21}$$

The transverse impedance is usually defined with an extra $(-\mathrm{i})$, in order to relate the real part to instability growth and the imaginary part to phase shift, as it is the case for the longitudinal impedance. The dipolar transverse impedance reads

$$\underline{Z}^{\mathrm{drive}}_{\perp,x/y}(\omega,\vec{r}_2) = -\frac{(-\mathrm{i})}{q_1 d_{x_1/y_1}}\int_{-\infty}^{\infty}\left[\underline{\vec{E}}(\omega) + \vec{v}\times\underline{\vec{B}}(\omega)\right]_{x/y}\mathrm{e}^{\mathrm{i}\omega z/v}\,\mathrm{d}z\,, \tag{22}$$

which can be generalized using the Panofsky–Wenzel theorem to (see Ref. [9])

$$\underline{Z}^{\mathrm{drive}}_{\perp,x/y}(\omega) = -\frac{v}{\omega(qd_{x/y})^2}\int_{\mathrm{beam}} \underline{\vec{E}}\cdot\underline{\vec{J}}^{*}_{s,d_{x/y}}\,\mathrm{d}V\,, \tag{23}$$

where $\underline{\vec{J}}_{s,d_{x/y}}$ is the dipolar component of the source current density in FD, cf. Eq. (12). Note that the formulations of Eqs. (21) and (23) are particularly convenient for evaluation on a mesh, since evaluation errors are averaged out by the integration.

1.3 Helmholtz decomposition

In a simply connected domain $\Omega \subset \mathbb{R}^3$, any differentiable vector field $\underline{\vec{A}} : \Omega \to \mathbb{C}^3$ can be written as $\underline{\vec{A}} = \underline{\vec{A}}_{\mathrm{curl}} + \underline{\vec{A}}_{\mathrm{div}}$, where $\underline{\vec{A}}_{\mathrm{curl}}$ and $\underline{\vec{A}}_{\mathrm{div}}$ are uniquely determined by demanding $\nabla \times \underline{\vec{A}}_{\mathrm{div}} = 0$ and $\nabla \cdot \underline{\varepsilon}\,\underline{\vec{A}}_{\mathrm{curl}} = 0$ for a piecewise smooth non-vanishing function $\underline{\varepsilon} : \Omega \to \mathbb{C}$. See, for example, Ref. [p. 86] [10] for a proof.

If the domain Ω is not simply connected, the Helmholtz decomposition has to be generalized to the so-called Hodge decomposition, i.e. a third field becomes constituent of $\underline{\vec{A}}$. This so-called harmonic field satisfies both $\nabla \times \underline{\vec{A}}_{\mathrm{harm}} = 0$ and $\nabla \cdot \underline{\varepsilon}\, \underline{\vec{A}}_{\mathrm{harm}} = 0$ and is yet non-zero.

Applying the Helmholtz decomposition with $\underline{\vec{E}}_{\mathrm{div}} = -\nabla \underline{\Phi}$ to Eq. (18) we find

$$-\nabla \cdot \underline{\varepsilon} \nabla \underline{\Phi} = \underline{\varrho}_s \tag{24a}$$

$$\nabla \times \underline{\nu} \nabla \times \underline{\vec{E}}_{\mathrm{curl}} - \omega^2 \underline{\varepsilon}\, \underline{\vec{E}}_{\mathrm{curl}} = \underline{\vec{R}}\,, \tag{24b}$$

where $\underline{\varrho}_s = (\mathrm{i}/\omega) \nabla \cdot \underline{\vec{J}}_s$ and $\underline{\vec{R}} = -\omega^2 \underline{\varepsilon} \nabla \underline{\Phi} - \mathrm{i}\omega \underline{\vec{J}}_s$. It is crucial here that due to the continuity equation $\nabla \cdot \underline{\vec{R}} = 0$ holds, i.e. all vector fields in Eq. (24b) are divergence free. Moreover, for a beam in z-direction, the charge can be written as $\underline{\varrho}_s = \underline{J}_{s,z}/v$.

2 Impedance simulations

Beam coupling impedance simulations can be sorted into three main groups, namely Time Domain (TD), Frequency Domain (FD), and methods without a particle beam. Most common are explicit TD methods, since they require only matrix-vector multiplications for time stepping. They are usually based on finite differences time domain (FDTD, Yee 1966 [11]) or finite integration technique (FIT, Weiland 1977 [12]), which result in a coinciding space discretization on a Cartesian mesh. However, note that, in general, mesh and method are independent, e.g. FIT or the finite element method (FEM) can be applied on both structured and unstructured, tetrahedral or hexahedral, or even mixed meshes.

Table 1: Examples of time domain wake-field codes. More detailed summaries can be found e.g. in [13, 14]

Code	Method	Website	Availability
CST PS	FIT	`www.cst.de`	commercial
GdfidL	FDTD	`www.gdfidl.de`	commercial
Echo	FIT	`www.desy.de/fel-beam/s2e/codes.html`	free
PBCI	FIT, DG-FEM	`www.temf.de` , see also [13]	on request
ABCI	FIT	`abci.kek.jp/abci.htm`	free
ACE3P	implicit FEM	`slac.stanford.edu`	free in USA

Explicit TD simulations are suitable at medium and high frequency, and particularly in perfectly conducting structures. They are disadvantageous for low frequencies and low velocity of the beam. Also dispersively lossy materials are difficult to treat in TD, since a convolution with the impulse response, i.e. the inverse FT of the material dispersion curve, is required. In FD the beam velocity and dispersive material data are just parameters. However, a system of linear equations (SLE) has to be solved for each frequency point, which is costly when the matrix is large and ill-conditioned.

In the following we focus on the FIT in TD and both FIT and FEM in FD. More specialized techniques in TD are the boundary element method (BEM) [15, 16], the finite volume (FV) method [17, 18], discontinuous Galerkin finite element method (DG-FEM), and implicit methods. In FD, there is no direct advantage from diagonal material matrices, which favours the FEM on an unstructured mesh. However, FIT is also used, since the structured mesh makes the implementation of Floquet boundary conditions simple. Particularly in the absence of materials, BEM is also an attractive option in FD [19, 20]. An overview of some commonly used TD codes is given in Table 1.

The most common method without particle beam is the computation of eigenmodes, which can be related to the wake function as discussed in Ref. [21]. Eigenmode computations with FEM for perfectly electric conducting (PEC) structures including losses through beam pipes and couplers are described in Ref. [22], a FIT algorithm for dispersively lossy tensorial materials used in e.g. ferrite cavities is

presented in Ref. [23]. Methods with excitations other than particle beams are based on current path [24, 25] or plane wave [26] excitation, but they require special interpretations to obtain the beam coupling impedance.

2.1 (Explicit) time domain

Due to the minimal duration–bandwidth product, the excitation is usually a Gaussian bunch

$$\lambda(z,t) = \frac{q}{\sqrt{2\pi}\sigma_s} \mathrm{e}^{-\frac{1}{2}\left(\frac{z-vt}{\sigma_s}\right)^2} , \tag{25}$$

which rigidly moves through the structure. The spectrum of this pulse is obtained from the FT over $s = vt - z$ as

$$|\underline{\lambda}(\omega)| = \frac{q_\omega}{\sqrt{2\pi}\sigma_\omega} \mathrm{e}^{-\frac{1}{2}\left(\frac{\omega}{\sigma_\omega}\right)^2} , \tag{26}$$

where $\sigma_\omega = v/\sigma_s$ and the normalization is $q_\omega = q\sqrt{2\pi}\sigma_\omega/v$. The duration and bandwidth are

$$T = \frac{\sigma_s}{\sqrt{2}v}, \qquad B = \frac{v}{\sigma_s\sqrt{2}} , \tag{27}$$

resulting in $TB = 1/2$. The point charge impedance is obtained by the convolution theorem as

$$\underline{\vec{Z}}(\omega) = \frac{\mathcal{F}\{\vec{W}^{\mathrm{pot}}(s)\}(\omega)}{\mathcal{F}\{\lambda(s)\}(\omega)} , \tag{28}$$

where the numerically obtained wake potential is transformed by (equidistant) discrete Fourier transform (DFT). The choice of the bunch length σ_s does not necessarily depend on the real bunch length in the accelerator, but rather on the frequency of interest. The maximum frequency at which a reasonable excitation amplitude is present, is roughly $2\sigma_f$, i.e. the spectrum is mainly located at $\sigma_f = v/(2\pi\sigma_s)$, the so-called frequency associated with the bunch length. Shorter bunches increase the maximum frequency, but they decrease the frequency resolution at low frequency. The frequency resolution depends on the total number of points employed for an equidistant DFT as $\Delta f = \frac{1}{N_{\mathrm{DFT}}\Delta t}$ and the total integrated wake length is $L_{\mathrm{W}} = vN_{\mathrm{DFT}}\Delta t$. Bunch length and wake length are the parameters to be set for the impedance simulation.

The FIT is based on evaluating the integral form of Maxwell's equations on a given mesh[1], i.e.

$$\begin{aligned}
\widehat{\mathbf{e}}_i &= \int_{L_i} \vec{E}\cdot\mathrm{d}\vec{s} & \widehat{\mathbf{h}}_i &= \int_{\tilde{L}_i} \vec{H}\cdot\mathrm{d}\vec{s} & \mathbf{q}_i &= \int_{\tilde{V}_i} \varrho\,\mathrm{d}V \\
\widehat{\widehat{\mathbf{d}}}_i &= \int_{\tilde{A}_i} \vec{D}\cdot\mathrm{d}\vec{A} & \widehat{\widehat{\mathbf{b}}}_i &= \int_{A_i} \vec{B}\cdot\mathrm{d}\vec{A} & \widehat{\widehat{\mathbf{j}}}_i &= \int_{\tilde{A}_i} \vec{J}\cdot\mathrm{d}\vec{A}\,.
\end{aligned} \tag{29}$$

The resulting $3N_p$-dimensional vectors are the electric and magnetic edge voltages $\widehat{\mathbf{e}}$ and $\widehat{\mathbf{h}}$, face fluxes $\widehat{\widehat{\mathbf{d}}}$ and $\widehat{\widehat{\mathbf{b}}}$, and the face current $\widehat{\widehat{\mathbf{j}}}$ and volume charge $\mathbf{q}$, which are connected by the continuity equation. Combining the integrals of Eqs. (29) to closed loops or closed surfaces (see Fig. 4) results in the Maxwell-Grid-Equations (MGE)

$$\mathbf{C}\widehat{\mathbf{e}} = -\partial_t\widehat{\widehat{\mathbf{b}}} \tag{30a}$$

$$\widetilde{\mathbf{C}}\widehat{\mathbf{h}} = \widehat{\widehat{\mathbf{j}}}_s + \widehat{\widehat{\mathbf{j}}} + \partial_t\widehat{\widehat{\mathbf{d}}} \tag{30b}$$

$$\widetilde{\mathbf{S}}\widehat{\widehat{\mathbf{d}}} = \mathbf{q}_s \tag{30c}$$

$$\mathbf{S}\widehat{\widehat{\mathbf{b}}} = 0\,, \tag{30d}$$

[1] For simplicity, we restrict ourselves here to Cartesian meshes.

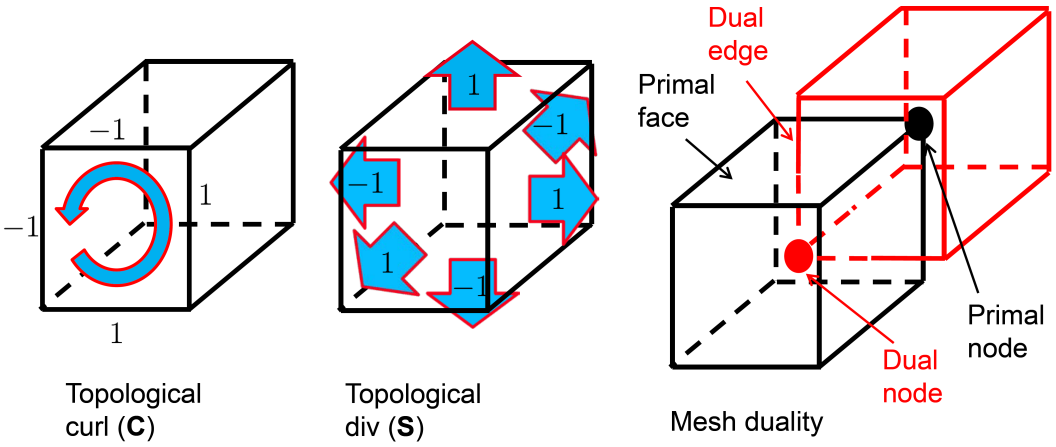


Fig. 4: Topological FIT mesh properties

where $\mathbf{C}$ and $\mathbf{S}$ are purely topological curl and divergence operators.The operators $\widetilde{\mathbf{C}}, \widetilde{\mathbf{S}}$ in Eqs. (30) represent evaluation on a dual grid, which has the property that dual vertices and edges intersect primal volumes and faces with same index, respectively, and vice versa. The MGE Eqs. (30) are exact, since they represent an evaluation of Maxwell's equations on the given grid topology. The numerical approximations required to solve the MGE are included in the material matrices ($\nu = \mu^{-1}$)

$$\widehat{\mathbf{h}} = \mathbf{M}_\nu \widehat{\widehat{\mathbf{b}}}\,, \qquad \widehat{\widehat{\mathbf{d}}} = \mathbf{M}_\varepsilon \widehat{\mathbf{e}}\,, \qquad \widehat{\widehat{\mathbf{j}}} = \mathbf{M}_\kappa \widehat{\mathbf{e}}\,, \tag{31}$$

which are diagonal matrices due to the dual orthogonal mesh and given by

$$[\mathbf{M}_\nu]_{n,n} = \overline{\nu}_n \frac{\tilde{L}_n}{A_n} \approx \frac{\int_{\tilde{L}_n} \vec{H} \cdot \mathrm{d}\vec{s}}{\int_{A_n} \vec{B} \cdot \mathrm{d}\vec{A}} \tag{32}$$

$$[\mathbf{M}_\varepsilon]_{n,n} = \overline{\varepsilon}_n \frac{\tilde{A}_n}{L_n} \approx \frac{\int_{\tilde{A}_n} \vec{D} \cdot \mathrm{d}\vec{A}}{\int_{L_n} \vec{E} \cdot \mathrm{d}\vec{s}}, \tag{33}$$

where $L_n, \tilde{L}_n, A_n, \tilde{A}_n$ denote the length and area of the nth primal and dual cell edge and face, respectively.

The discretization of time derivatives can be done by forward (explicit) or backward (implicit) finite differences. Implicit methods are unconditionally stable, but they require solving a SLE in each time step. Explicit methods are much lighter in computation, but they are only conditionally stable.

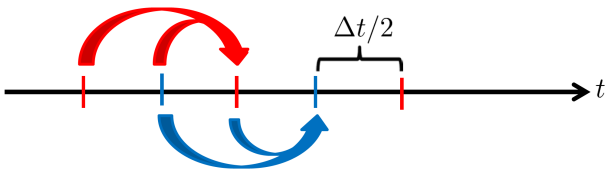


Fig. 5: Leap-frog method. Magnetic fields are allocated on half integer time indices between two electric field steps.

The most commonly used explicit method is the so-called 'leap-frog' method, introduced by Yee [11] in 1966. It consists of a (staggered) central-difference quotient featuring second-order accuracy

$$\widehat{\mathbf{h}}^{n+1/2} = \widehat{\mathbf{h}}^{n-1/2} - \Delta t \mathbf{M}_\mu^{-1} \mathbf{C} \widehat{\mathbf{e}}^n \tag{34a}$$

$$\widehat{\mathbf{e}}^{n+1} = \widehat{\mathbf{e}}^n - \Delta t \mathbf{M}_\varepsilon^{-1} (\widetilde{\mathbf{C}} \widehat{\mathbf{h}}^{n+1/2} - \widehat{\widehat{\mathbf{j}}}_s^{n+1/2})\,, \tag{34b}$$

which is visualized in Fig. 5. The stability of the scheme is connected to the grid dispersion relation, which describes the velocity of a plane wave on the grid as dependent on the direction of the wave vector. It reads for a particular Cartesian cell $(\Delta x, \Delta y, \Delta z)$

$$\left(\frac{\sin\frac{k_x\Delta x}{2}}{\Delta x/2}\right)^2 + \left(\frac{\sin\frac{k_y\Delta y}{2}}{\Delta y/2}\right)^2 + \left(\frac{\sin\frac{k_z\Delta z}{2}}{\Delta z/2}\right)^2 = \mu\varepsilon\left(\frac{\sin\frac{\omega\Delta t}{2}}{\Delta t/2}\right)^2 , \tag{35}$$

reproducing the continuous dispersion relation $k_x^2 + k_y^2 + k_z^2 = \omega^2\mu\varepsilon$ in the limit $\Delta x, \Delta y, \Delta z, \Delta t \to 0$. In order to fulfil Eq. (35) with real valued frequency and wavenumbers, at least

$$\Delta t \leq \min_i \sqrt{\frac{\mu_i\varepsilon_i}{\frac{1}{\Delta x_i^2} + \frac{1}{\Delta y_i^2} + \frac{1}{\Delta z_i^2}}} \tag{36}$$

must hold, where the minimum is taken over all mesh cells. This is also referred to as the Courant–Friedrichs–Lewy (CFL) criterion [27] for the time step Δt. It can be shown, see Ref. [28], that Eq. (36) is also a sufficient condition for stability on the time step. Particularly for short bunches, the grid dispersion can lead to unphysical effects, such as a positive wake potential at the head of the bunch or numerical Cherenkov radiation. Therefore dedicated schemes have been developed, which e.g. do not have dispersion in the direction of beam propagation [29].

At low frequencies, the CFL criterion together with the Küpfmüller uncertainty principle pose a strong requirement on the time step. Since longer wakes have to be computed at low frequencies but the time step is tied on the space step, a large number of time steps need to be computed, which is a massive oversampling of a low-frequency wave. This is a major drawback of explicit TD methods at low frequencies.

2.2 Frequency domain

The MGE can be written in FD as

$$\mathbf{C}\widehat{\underline{\mathbf{e}}} = -\mathrm{i}\omega\widehat{\underline{\mathbf{b}}} \tag{37a}$$

$$\widetilde{\mathbf{C}}\widehat{\underline{\mathbf{h}}} = \widehat{\underline{\mathbf{j}}}_s + \widehat{\underline{\mathbf{j}}} + \mathrm{i}\omega\widehat{\underline{\mathbf{d}}} \tag{37b}$$

$$\widetilde{\mathbf{S}}\widehat{\underline{\mathbf{d}}} = \underline{\mathbf{q}}_s \tag{37c}$$

$$\mathbf{S}\widehat{\underline{\mathbf{b}}} = 0 , \tag{37d}$$

which can be combined with the complex material relations to a $3N_p \times 3N_p$ SLE

$$\left(\widetilde{\mathbf{C}}\mathbf{M}_{\underline{\nu}}\mathbf{C} + \mathrm{i}\omega\mathbf{M}_\kappa - \omega^2\mathbf{M}_\varepsilon\right)\widehat{\underline{\mathbf{e}}} = -\mathrm{i}\omega\widehat{\underline{\mathbf{j}}}_s . \tag{38}$$

Since the FIT is a mimetic discretization, the Helmholtz decomposition can also be applied on the discrete level, preserving the properties discussed in Section 1.3. The monopolar beam source current can be written as (i_z is the z-index and $\tilde{z}_i$ the corresponding coordinate)

$$\widehat{\underline{\mathbf{j}}}_{s,z}^{\mathrm{mono}}(i_z) = \int \vec{\underline{J}}_s \cdot \mathrm{d}\vec{A}_z = q\mathrm{e}^{-\mathrm{i}\omega\tilde{z}_i/v} \tag{39}$$

and a dipole source is obtained by $\widehat{\underline{\mathbf{j}}}_{s,z}^{\mathrm{dip}}(i_z) = \widehat{\underline{\mathbf{j}}}_{e,z}^{\mathrm{mono}}(x = -d_x/2) - \widehat{\underline{\mathbf{j}}}_{e,z}^{\mathrm{mono}}(x = +d_x/2)$. For the entry and exit of the beam in the computational domain dedicated boundary conditions are required. A simplified way to implement these is to use Floquet boundary conditions, i.e. to connect the entry and the exit of the beam with the proper phase advance $\exp(\mathrm{i}\omega L/v)$, where L is the length of the computational domain. This has the same effect as if the structure would be repeated infinitely often in a chain. It is

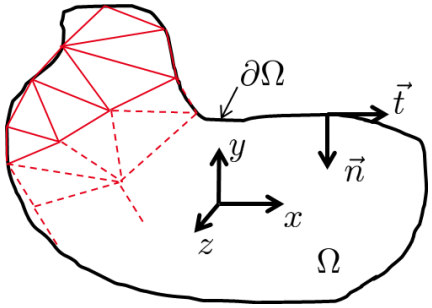


Fig. 6: Computational domain for the 2D impedance solver

valid for frequencies below the beam pipe cutoff, i.e. when the subsequent (repeated) structures do not interact with each other.

The impedances can be evaluated according to Eqs. (21) and (23) as

$$\underline{Z}_{\parallel}(\underline{\widehat{\mathbf{e}}}(\omega)) = -\frac{1}{q^2}\underline{\widehat{\mathbf{e}}} \cdot \underline{\widehat{\mathbf{j}}}^*_{\text{mono}} \tag{40}$$

$$\underline{Z}_{\perp}(\underline{\widehat{\mathbf{e}}}(\omega)) = -\frac{v}{\omega(qd_x)^2}\underline{\widehat{\mathbf{e}}} \cdot \underline{\widehat{\mathbf{j}}}^*_{\text{dip}} \,, \tag{41}$$

which can be seen as functionals of the discrete solution of Maxwell's equations in FD.

In FD one cannot use the advantages of diagonal material matrices, and therefore FEM on unstructured meshes can be more advantageous than FIT. This becomes particularly clear in the discretization of dipole sources for non-ultrarelativistic beams, cf. Eq. (12). We will briefly discuss a 2D FEM approach which is particularly useful for beam pipes and long kicker magnets.

The finite element method is based on decomposing the computational domain Ω in finite-sized subdomains Ω_e, i.e. the elements. A function in an appropriate space can be approximated by a finite basis, such that each element is the support of one basis (ansatz) function. Since such an approximation is (weakly) differentiable only once, a second-order PDE has to be brought in a 'weak formulation'. This is obtained by multiplying with all test functions[2] of an appropriate test-function space, integrating over the whole domain and transferring one (exterior) derivative by means of partial integration. Finally an SLE is obtained, which has number of ansatz functions as columns and number of test functions as rows.

We consider again Eq. (18), but $\vec{\underline{E}} : \Omega \to \mathbb{C}^3$ with $\Omega \subset \mathbb{R}^2$ being a simply connected domain as shown in Fig. 6. Each triangle shall be uniformly filled with material $\underline{\mu}, \underline{\varepsilon}$, where $\underline{\varepsilon} = \varepsilon_0\varepsilon_r - \mathrm{i}\kappa/\omega$. In order to allow the normal component of the electrical field to jump on a material border while the tangential is continuous, Nédélec edge elements are employed. These elements (at lowest order) are not suited to model the divergence of a field. Thus a Helmholtz split needs to be performed, i.e. Eqs. (24) are discretized.

The discretization of the complex Poisson equation Eq. (24a) is done by nodal functions[3]

$$N_k(\xi, \eta) = a_k\xi + b_k\eta + c_k \,, \tag{42}$$

which fulfil

$$N_k(\xi_i, \eta_i) = \delta_{i,k} \,, \tag{43}$$

where i and k are local vertex indices and (ξ, η) are local coordinates. The lowest-order Nédélec edge elements of the first kind can be obtained from the nodal elements by (see e.g. Ref. [30])

$$\vec{w}_i(\xi, \eta) = N_k\nabla_\perp N_l - N_l\nabla_\perp N_k \,, \tag{44}$$

[2] In the standard (Galerkin) approach, the test functions are identical to the ansatz functions.

[3] For simplicity we consider the ansatz functions as purely real, i.e. the PDEs need to be split into real and imaginary part, which are coupled in the presence of losses.

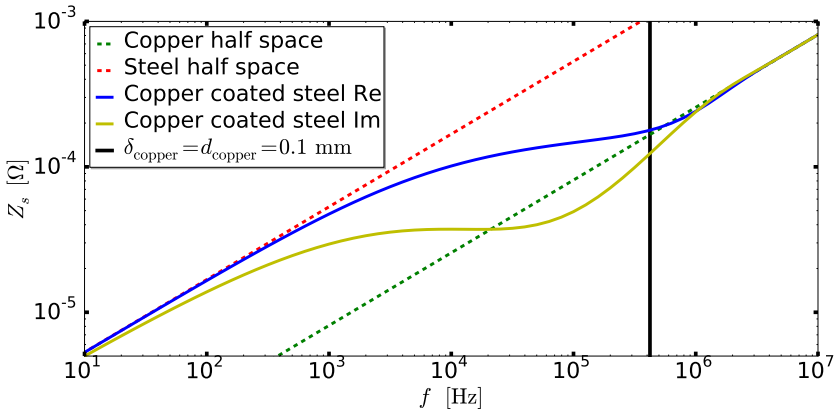


Fig. 7: Surface impedance for a thick steel surface, coated by a thin copper layer

and fulfil

$$\frac{1}{|l_k|}\int_{l_k} \vec{w}_i \cdot \vec{t}_k \,\mathrm{d}s = \delta_{i,k} \tag{45}$$

with $\vec{t}_k$ being the tangential unit vector of edge l_k, which is located at the opposite of vertex k. Instead of going further into details about impedance computation with FEM, we refer to [31] and references therein.

In FD and particularly FEM it is also fairly simple to include a thick conducting wall by means of a surface impedance boundary condition (SIBC), i.e. $\vec{n}\times\vec{n}\times\underline{\vec{E}} = \underline{Z}_s\vec{n}\times\underline{\vec{H}}$ with the surface impedance $Z_s = \sqrt{\underline{\mu}/\underline{\varepsilon}}$. For smooth metal surfaces one finds

$$\underline{Z}_s = \frac{1+i}{\kappa\delta_s} \qquad \text{with the skin depth} \qquad \delta_{\mathrm{s}} = \sqrt{\frac{2}{\mu\kappa\omega}}\,. \tag{46}$$

This can also be generalized for a metal coating (thickness d, permeability μ_1 and conductivity κ_1) on a metal surface (permeability μ_2 and conductivity κ_2) as

$$\underline{Z}_s = \frac{1+i}{\kappa_1\delta_{s1}}\frac{M^{(+)}\mathrm{e}^{\mathrm{i}k_{z1}d} + M^{(-)}\mathrm{e}^{-\mathrm{i}k_{z1}d}}{M^{(+)}\mathrm{e}^{\mathrm{i}k_{z1}d} - M^{(-)}\mathrm{e}^{-\mathrm{i}k_{z1}d}}\,, \tag{47}$$

where

$$k_{z1,2} = \frac{1-\mathrm{i}}{\delta_{s1,2}}\,, \quad M^{(+)} = 1+\sqrt{\frac{\mu_1\kappa_2}{\mu_2\kappa_1}}\,, \quad M^{(-)} = 1-\sqrt{\frac{\mu_1\kappa_2}{\mu_2\kappa_1}}\,. \tag{48}$$

The surface impedance of a copper ($\kappa = 70$MS) coated steel ($\kappa = 1.4$MS) surface is plotted in Fig. 7.

2.3 Two examples in two dimensions

The beam-induced heat load in a ferrite kicker module depends crucially on the ferrite yoke gap. If there is no gap, the magnetic circuit is closed and the longitudinal impedance is much larger.

In the presence of a gap, the longitudinal impedance can be further decreased by increasing the path length for the magnetic field lines outside the ferrite. Achieving this by increasing the gap thickness can be disadvantageous for the kick-field quality. The gap can also be filled highly conductive material, i.e. copper, which influences the kick field only weakly.

Figure 8 shows the longitudinal impedance of a GSI SIS-100 transfer kicker module (design outline) from 2D FIT and FEM simulations, as a comparison between a vacuum and a copper-filled magnet gap. The difference between the two is significant, i.e. two orders of magnitude. The heat power values

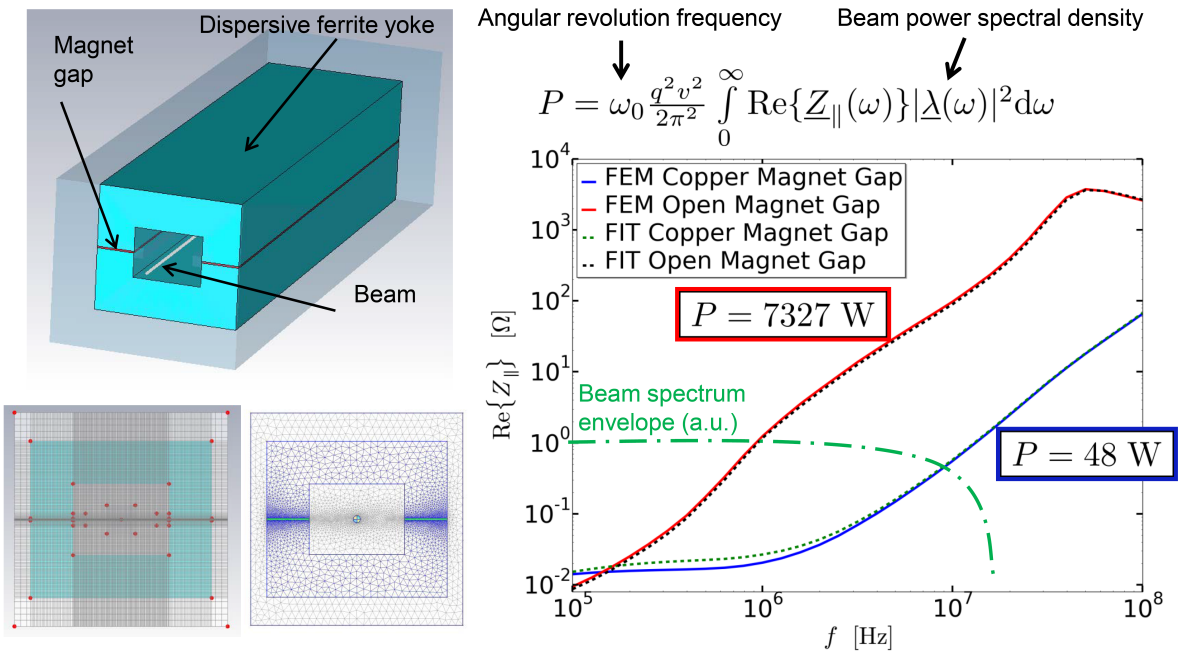


Fig. 8: Outline of a GSI SIS-100 transfer kicker. The plot shows the longitudinal impedance computed with FIT on Cartesian and FEM on triangular mesh, for both open and copper filled magnet gaps.

are for the shortest (3.7 m) single Gaussian bunch with $N = 2 \times 10^{13}$ protons. Note that these numbers are for CW operation, in practice they have to be scaled with the duty factor which is in the range of 0.5 at most.

The second example is the transverse impedance of a thin-layered round beam pipe, see Fig. 9. At low frequency the wall has to be meshed since field transmission plays a significant role, especially for frequencies below the maximum. For frequencies above the skin frequency f_s the skin depth becomes very small and cannot be properly meshed. By means of the SIBC, resolving the skin depth can be avoided and high frequencies can be reached.

3 Impedance bench measurements

The transmission-line measurement technique is based on replacing the beam by a wire and measuring the attenuation of a TEM wave. It was introduced in 1974 by Sands and Rees [32] in order to determine the beam energy-loss factors in the TD using a broadband pulse with a shape similar to the particle bunch. Nowadays, modern vector network analysers (VNA) allow sweeping a narrow-banded sinusoidal signal, to obtain the impedance directly for a particular frequency range. Especially when particular beam instability sidebands are under investigation, the FD method is advantageous.

The motivation of the wire measurement comes from a the fields of a point charge moving at velocity βc which read in the rest frame of the charge as

$$\vec{E}'(\varrho', z', t') = \frac{q}{4\pi\varepsilon} \left(\frac{\varrho'}{\sqrt{\varrho'^2 + z'^2}^3} \vec{e}_\varrho + \frac{z'}{\sqrt{\varrho'^2 + z'^2}^3} \vec{e}_z \right) . \tag{49}$$

Lorentz-boosting to the laboratory frame one obtains

$$\vec{E}(\varrho, z, t) = \frac{q}{4\pi\varepsilon} \left(\frac{\gamma\varrho}{\sqrt{\varrho^2 + (\beta\gamma ct)^2}^3} \vec{e}_\varrho + \frac{-\beta\gamma ct}{\sqrt{\varrho^2 + (\beta\gamma ct)^2}^3} \vec{e}_z \right) , \quad H_\varphi = \frac{\beta}{Z_0} E_\varrho \tag{50}$$

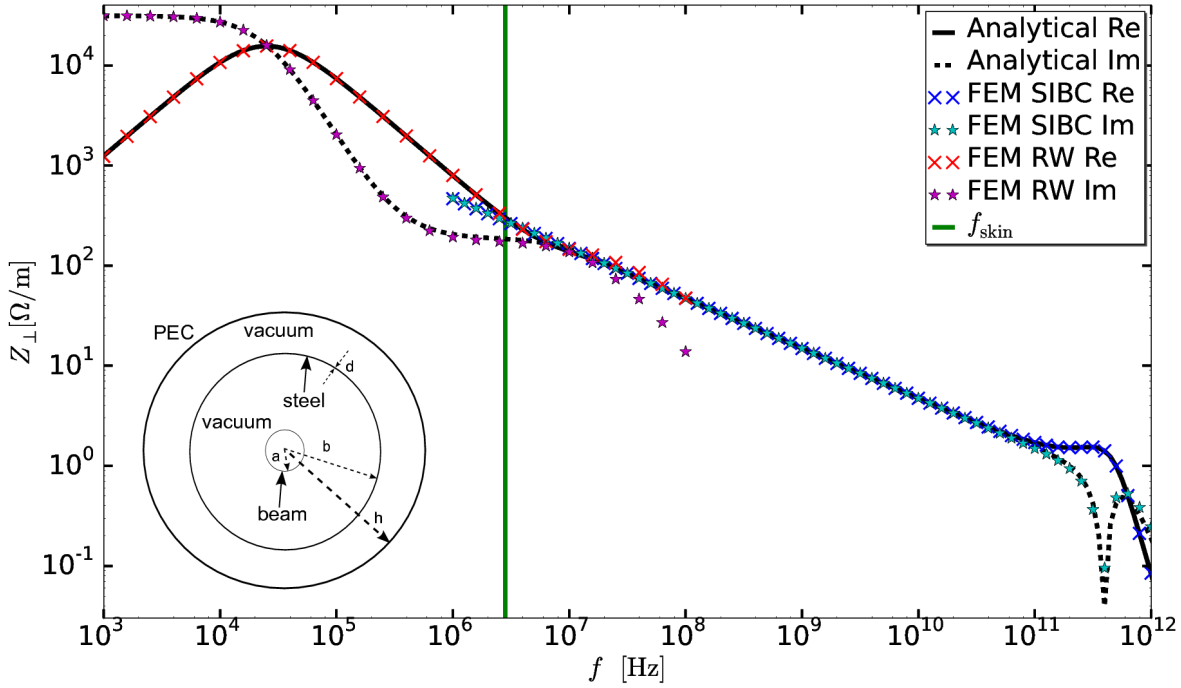


Fig. 9: Transverse impedance of the thin beam pipe with radius $b = 4\,\text{cm}$, thickness $d = 0.3\,\text{mm}$, outer boundary radius $h = 1\,\text{m}$, length $l = 1\,\text{m}$, conductivity $\kappa = 1\,\text{MS/m}$ and $\beta = 0.999999$. The analytical curve originates from Rewall [48]. At low frequencies, the resistive wall is resolved by the mesh (RW-marks, red and magenta), while at high frequency the surface impedance boundary condition was applied (SIBC-marks, blue and cyan).

and subsequent Fourier transform results in

$$\underline{E}_z(\varrho, z, \omega) = \mathrm{i}q\frac{\mu_0}{2\pi}\frac{\omega}{\beta^2\gamma^2}\mathrm{K}_0\left(\frac{\omega}{\beta\gamma\mathrm{c}}\varrho\right) \xrightarrow{\gamma\to\infty} 0 \tag{51a}$$

$$\frac{1}{\beta}Z_0\underline{H}_\varphi = \underline{E}_\varrho(\varrho, z, \omega) = q\frac{\mu_0}{2\pi}\frac{\omega}{\beta^2\gamma}\mathrm{K}_1\left(\frac{\omega}{\beta\gamma\mathrm{c}}\varrho\right) \xrightarrow{\gamma\to\infty} \mathrm{q}\frac{\mathrm{Z}_0}{2\pi\varrho} \tag{51b}$$

which is a TEM mode in the ultrarelativistic limit. From Eq. (51b), we can see that the wire technique corresponds only to an ultrarelativistic beam, since the wave impedance for a beam is

$$Z_{\text{wave}} = \frac{\underline{E}_\varrho}{\underline{H}_\varphi} = \frac{Z_0}{\beta}, \tag{52}$$

but for a TEM mode one has always $Z_{\text{wave}} = Z_0$. Moreover, the wire needs to be thin, in order to have most of the field unperturbed close to the wire. In the wide sense, this corresponds to the rigid-beam approximation.

3.1 Basics of RF vector network analysis and impedance matching

In Radio Frequency (RF) systems, the integral of the electric field strength depends on the taken path, thus voltages are not uniquely defined. To account for that, power flow parameters

$$a_i := \frac{1}{2\sqrt{Z_c}}(\underline{U}_i + Z_c\underline{I}_i)\,, \quad b_i := \frac{1}{2\sqrt{Z_c}}(\underline{U}_i - Z_c\underline{I}_i) \tag{53}$$

in units of $\sqrt{\mathrm{W}}$ are used rather than voltages and currents. The out-flowing power of a linear, time-invariant device with $i = 1, \ldots, N$ ports is determined from the inflowing power by $\vec{b} = \mathbf{S}\vec{a}$, where the

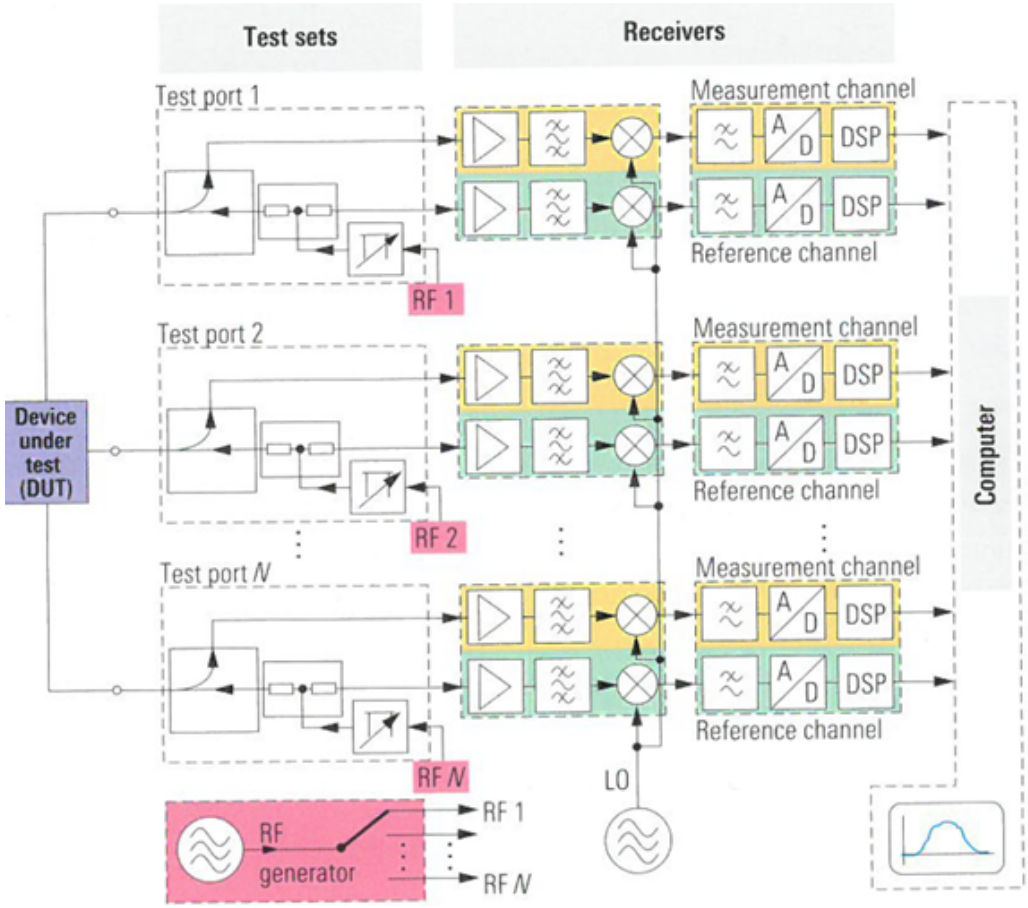


Fig. 10: Block schematic of a modern VNA. Picture adapted from Ref. [33]

scattering parameters S_{ij} are defined by

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0 \;\; \forall k \neq j} . \tag{54}$$

For simplicity, we assume the characteristic impedance for all ports to be equal and consider only the TEM mode. The scattering matrix can be determined by numerical simulation or by measurement with a vector network analyser (VNA), see Fig. 10 for a block schematic of a VNA. Prior to measurement, the VNA needs to be calibrated at a particular reference plane close to the device under test (DUT), in order to exclude effects of the cables. Phase-stable cables are crucial for measurements at higher frequencies, but such precision measurement cables are costly. A good alternative are semi-rigid cables, which maintain their shape and therefore have only little phase drift. In order to connect the $50\,\Omega$ lines of the VNA with the wire for beam impedance measurement, the characteristic impedance [34]

$$Z_c = \frac{Z_0}{2\pi} \ln \frac{b}{a} , \tag{55}$$

where b and a are the radii of the outer and inner conductor of a coaxial cable, needs to be matched. A mismatched transition would lead to a reflection [34]

$$r = \frac{Z_{c,1} - Z_{c,2}}{Z_{c,1} + Z_{c,2}} \tag{56}$$

which is for the transition of a thin wire in a measurement box (as depicted in Fig. 11) to a conventional $50\,\Omega$ line in the range of 80%. Thus, a matching network is required, since otherwise multiple reflections

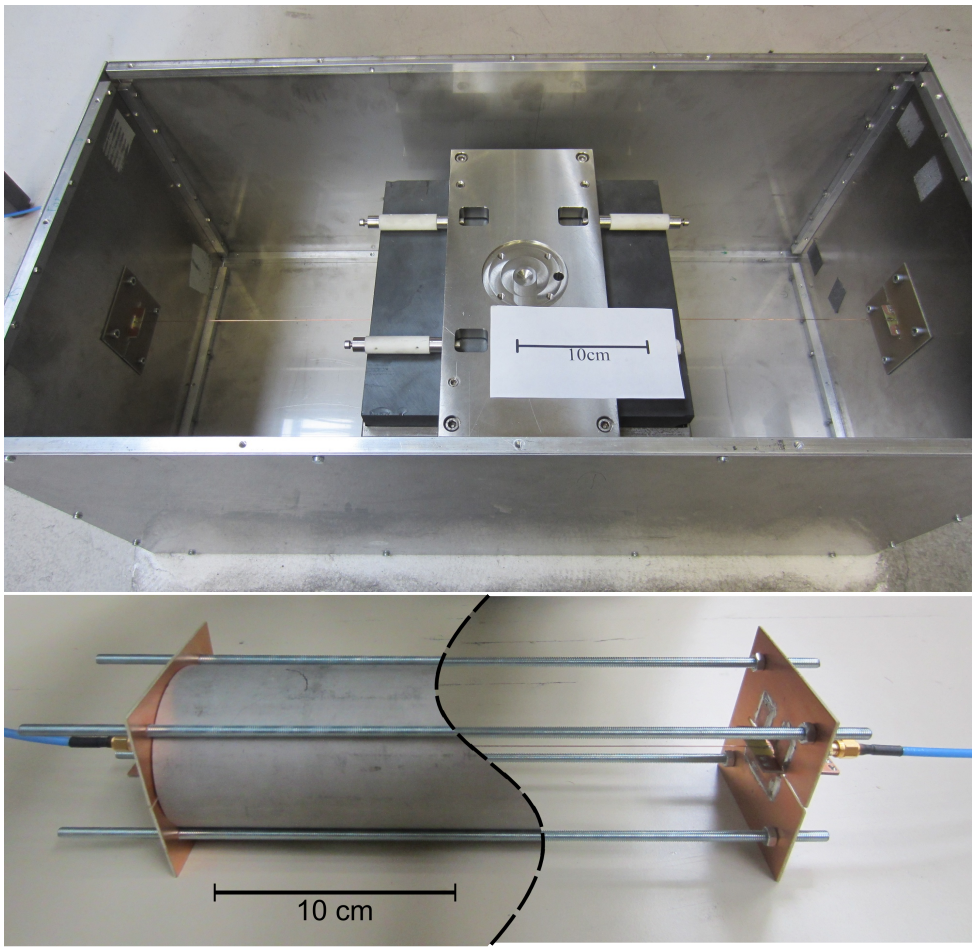
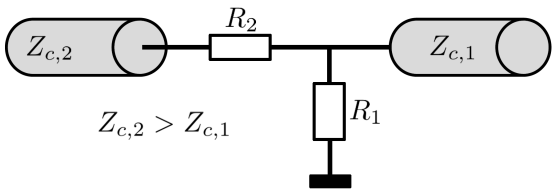
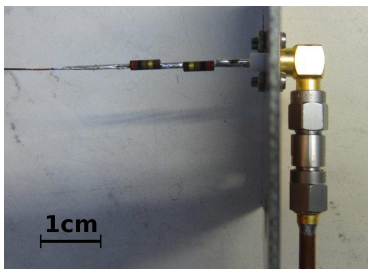


Fig. 11: Large ($Z_c = 433 \pm 18\,\Omega$) and small ($Z_c = 299 \pm 12\,\Omega$) measurement box

would be in the same range of amplitude as the primary signal. The simplest way to construct a matching network is to use RF-resistors (carbon composite) in a way that each side sees its own characteristic impedance, e.g. as depicted in Fig. 12(a).



(a) Matching with a simple voltage divider.



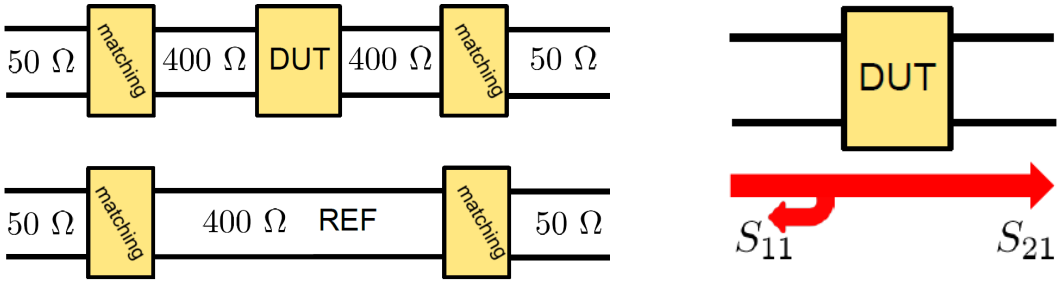
(b) Matching with resistors and a 10 dB attenuator [35].

Fig. 12: Different types of resistive matching

Here, the two resistors R_1 and R_2 have to fulfil the two matching conditions

$$R_1 \parallel (R_2 + Z_{c,2}) = Z_{c,1} \tag{57a}$$

$$R_2 + R_1 \parallel Z_{c,1} = Z_{c,2}\,, \tag{57b}$$



(a) REF measurement in order to de-embed the DUT.

(b) Quantities of interest for the de-embedded DUT.

Fig. 13: De-embedding by subsequent DUT and REF measurements

where $x \parallel y = xy/(x+y)$ is the abbreviation for parallel circuits. Unfortunately, practical resistors have an inductance, which makes it impossible to fulfil Eqs. (57) at higher frequencies in a broadband manner. This can be partly overcome by taking an attenuator instead of R_1, see Fig. 12(b). Commercially available attenuators are broadband matched to $50\,\Omega$ and do not suffer from the inductance problem.

The frequency-dependent attenuation and phase shift of the matching network is calibrated out by a reference measurement, such that for an assumed perfect matching the de-embedded transmission is $S_{21}^{\mathrm{de-embed}}(\omega) = S_{21}^{\mathrm{DUT}}(\omega)/S_{21}^{\mathrm{REF}}(\omega)$. Here, only the reflection of the matching network needs to be close to zero, but it is allowed to be lossy, within the dynamic range of the VNA. The way how S_{21}^{DUT} and S_{21}^{REF} are measured is illustrated in Fig. 13. Obviously, it is advantageous to have a setup as small as possible, in order to shift the eigenmodes (resonances) of the box to frequencies as high as possible.

3.2 Wire method

There is a crucial difference between the beam and the wire setup: the TEM wave experiences an attenuation, which is not negligible and actually the quantity to be measured by the S_{21}-parameter. Thus, lumped (short) and distributed (long) impedances require different interpretations of the measured S_{21}-parameters. Mathematically, lumped and distributed impedances can be identified by their distribution along the z-axis, i.e.

$$\frac{\partial \underline{Z}_{\parallel}^{\mathrm{lumped}}(\omega,z)}{\partial z} = \underline{Z}_{\parallel}^{\mathrm{total}}(\omega)\delta(z-z_0) \tag{58a}$$

$$\frac{\partial \underline{Z}_{\parallel}^{\mathrm{dist}}(\omega,z)}{\partial z} = \frac{\underline{Z}_{\parallel}^{\mathrm{total}}(\omega)}{l}\,. \tag{58b}$$

In a real accelerator components, there is always a mixture of both. The impedance discontinuity (geometric impedance) at the beginning of the DUT is always lumped, while the body of the DUT (resistive wall) is often almost equally distributed.

The modelling of lumped impedances is just a localized impedance element in the longitudinal direction, while distributed impedances can be represented by a TEM line with an impedance element $Z_{\parallel}/l$ equally distributed to each infinitely short transmission-line element, see Fig. 14. We call a device a lumped (distributed) impedance if $\underline{Z}^{\mathrm{lumped}} \gg \underline{Z}^{\mathrm{dist}}$ ($\underline{Z}^{\mathrm{lumped}} \ll \underline{Z}^{\mathrm{dist}}$).

The scattering matrices for the de-embedded DUT are given by Ref. [34, 36]

$$\mathbf{S}^{\mathrm{lump}} = \frac{1}{2Z_c^{\mathrm{REF}} + Z^{\mathrm{lump}}} \begin{pmatrix} Z^{\mathrm{lump}} & 2Z_c^{\mathrm{REF}} \\ 2Z_c^{\mathrm{REF}} & Z^{\mathrm{lump}} \end{pmatrix} \tag{59}$$

(a) Lumped impedance **(b)** Distributed impedance

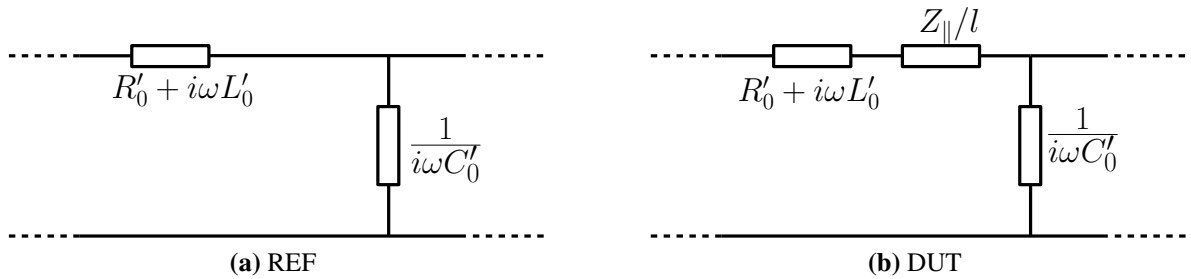
Fig. 14: Different modelling approaches for the de-embedded accelerator component

$$\mathbf{S}^{\mathrm{dist}} = \frac{\begin{pmatrix} (Z_c^{\mathrm{DUT}^2} - Z_c^{\mathrm{REF}^2})\sin(k_{\mathrm{z}}^{\mathrm{DUT}} l) & -2iZ_c^{\mathrm{DUT}} Z_c^{\mathrm{REF}} \\ -2iZ_c^{\mathrm{DUT}} Z_c^{\mathrm{REF}} & (Z_c^{\mathrm{DUT}^2} - Z_c^{\mathrm{REF}^2})\sin(k_{\mathrm{z}}^{\mathrm{DUT}} l) \end{pmatrix}}{(Z_c^{\mathrm{DUT}^2} + Z_c^{\mathrm{REF}^2})\sin(k_{\mathrm{z}}^{\mathrm{DUT}} l) - 2iZ_c^{\mathrm{DUT}} Z_c^{\mathrm{REF}} \cos(k_{\mathrm{z}}^{\mathrm{DUT}} l)} \tag{60}$$

for the lumped and distributed impedance, respectively. How to derive the DUT impedance from the measured S-parameters will be discussed for lumped and distributed impedances in the following subsections.

3.2.1 Distributed-impedance measurement

In transmission-line theory, a ladder-replacement-circuit model as shown in Fig. 15 can be derived. Here, L_0', C_0', and R_0' are inductance, capacitance, and resistance per length, respectively. The distributed beam coupling impedance can be seen as an additional longitudinal element $Z_\parallel/l$.



(a) REF **(b)** DUT

Fig. 15: Transmission-line replacement circuit for distributed impedance

From the transmission-line parameters the propagation constants and characteristic impedances can be calculated as (see Ref. [34])

$$k_{\mathrm{z}}^{\mathrm{DUT}} = \omega\sqrt{C_0'L_0'}\sqrt{1 - \mathrm{i}\frac{R_0' + \underline{Z}_\parallel/l}{\omega L_0'}} \tag{61a}$$

$$k_{\mathrm{z}}^{\mathrm{REF}} = \omega\sqrt{C_0'L_0'}\sqrt{1 - \mathrm{i}\frac{R_0'}{\omega L_0'}} \tag{61b}$$

$$\underline{Z}_c^{\mathrm{DUT}} = \sqrt{\frac{R_0' + \mathrm{i}\omega L_0' + \underline{Z}_\parallel/l}{\mathrm{i}\omega C_0'}} \tag{61c}$$

$$\underline{Z}_c^{\mathrm{REF}} = \sqrt{\frac{R_0' + \mathrm{i}\omega L_0'}{\mathrm{i}\omega C_0'}} \approx \sqrt{\frac{L_0'}{C_0'}} =: Z_c. \tag{61d}$$

This system can be solved for $Z_\parallel$ as

$$\underline{Z}_\parallel^{\mathrm{dist}} = iZ_c^{\mathrm{REF}} l \cdot (k_{\mathrm{z}}^{\mathrm{DUT}} - k_{\mathrm{z}}^{\mathrm{REF}}) \cdot \left(1 + \frac{k_{\mathrm{z}}^{\mathrm{DUT}}}{k_{\mathrm{z}}^{\mathrm{REF}}}\right). \tag{62}$$

Since the DUT setup is a combination of three transmission lines (cf. Fig. 14), obtaining the propagation constants can be involved, when a reflection takes place at the DUT. When this reflection is small, i.e. $Z_c^{\mathrm{DUT}} \simeq Z_c^{\mathrm{REF}}$, Eq. (60) simplifies to

$$S_{21} = S_{12} = \mathrm{e}^{-\mathrm{i}k_z l} \ , \ \ S_{11} = S_{22} = 0 \ , \tag{63}$$

which can be easily inverted. Otherwise a reflection-corrected S_{21}-parameter can be introduced, which is by definition

$$S_{21}^{\mathrm{C}} := \mathrm{e}^{-\mathrm{i}k_z l} \ . \tag{64}$$

The new S_{21}^{C}-parameter can be obtained by solving Eq. (60) for k_{z}, which can be achieved through replacing sine and cosine by exponentials. The quadratic equation thereby derived for S_{21}^{C} is called the Wang–Zhang formula [37] and is

$$(S_{21}^{\mathrm{C}})^2 + \frac{S_{11}^2 - S_{21}^2 - 1}{S_{21}} S_{21}^{\mathrm{C}} + 1 = 0 \ , \tag{65}$$

where only one of the two solutions, which fulfils $|S_{21}^{\mathrm{C}}| < 1$, is physical. Solving Eq. (65) requires the knowledge of the S_{11}-parameter, which is in practice difficult to measure because of multiple reflections between the DUT and the matching section (cf. Fig. 13). Nonetheless, S_{11} can be determined easily in simulations with waveguide ports.

The wavenumber k_{z} is found from the complex logarithm of either the original (see Eq. (63)) or the corrected (see Eq. (64)) S_{21}-parameter. It can be inserted into Eq. (62) to obtain (see Ref. [36])

$$\underline{Z}_{\parallel}^{\mathrm{dist}} = Z_c \cdot \ln\left(\frac{S_{21}^{\mathrm{REF}}}{S_{21}^{\mathrm{DUT}}}\right) \cdot \left[1 + \frac{\ln(S_{21}^{\mathrm{DUT}})}{\ln(S_{21}^{\mathrm{REF}})}\right] \ , \tag{66}$$

which is called 'improved-log-formula'[4] in the literature. This formula is exact for ideally distributed impedances, but it does not apply to lumped impedances, since the replacement circuit in Fig. 15 requires many such transmission-line elements in succession. The dependence on the electrical length of the reference $\Theta_{\mathrm{z}}^{\mathrm{REF}} = k_{\mathrm{z}}^{\mathrm{REF}} l = \omega l / c$ can be pointed out explicitly by rewriting Eq. (66) as (see Ref. [40])

$$\underline{Z}_{\parallel}^{\mathrm{dist}} = Z_c \cdot \ln\left(\frac{S_{21}^{\mathrm{REF}}}{S_{21}^{\mathrm{DUT}}}\right) \cdot \left[2 + \frac{i}{\Theta_{\mathrm{z}}^{\mathrm{REF}}} \ln\left(\frac{S_{21}^{\mathrm{DUT}}}{S_{21}^{\mathrm{REF}}}\right)\right] \ . \tag{67}$$

This formula contains only the logarithm of the ratio, i.e. the difference term in Eq. (62). Besides the implicit dependence of k_z^{DUT} on l, the $\ln(S_{21}^{\mathrm{DUT}}/S_{21}^{\mathrm{REF}})$-term is linear in l. Thus, for distributed impedances, the square bracket in Eq. (67) does not depend on the length explicitly.

When inserting the lumped impedance S-parameters (see Eq. (59)) into the improved-log-formula (see Eq. (66) or (67)) one observes that the second term in the bracket is not independent of the length anymore, i.e. Θ_z does not cancel. This shows explicitly the inapplicability of the improved-log-formula to lumped impedances since the length is not defined for lumped impedances. In other words, if a differentially short transmission-line element is assigned a finite impedance value (lumped impedance) and this is integrated over a finite length, then the result must diverge. The convergence of $\underline{Z}_{\parallel}^{\mathrm{dist}}$ to $\underline{Z}_{\parallel}$ in the limit of decreasing wire radius is discussed in Ref. [35].

3.2.2 Lumped impedance measurement

The determination of lumped impedances is significantly simpler than the one for distributed impedances, since the reflection does not influence the transmission measurement result. In fact, the reflection can

[4]Historically, first the lumped-element formulas [32, 38], then the simplified transmission-line formula 'Log-formula' [39], and finally the full transmission-line formula 'Improved Log-formula' [36] were derived.

even be used as an alternative method to determine a lumped impedance. However, Hahn and Pedersen argued, in Ref. [38], that the reflection method is inferior to the transmission method. From solving Eq. (59) for $\underline{Z}^{\mathrm{lump}}$ one obtains the so-called Hahn–Pedersen lumped impedance formula [38],

$$\underline{Z}^{\mathrm{lump}}_{\parallel,\mathrm{HP}} = 2Z_c \frac{S_{21}^{\mathrm{REF}} - S_{21}^{\mathrm{DUT}}}{S_{21}^{\mathrm{DUT}}} \,. \tag{68}$$

In modern VNAs this impedance measurement formula is already built-in, i.e. the impedance can be directly displayed for the simplified case $S_{21}^{\mathrm{REF}} = 1$. Equation (68) is an improvement of the original Sands–Rees pulse-energy-loss formula [32] (see also Ref. [41])

$$\underline{Z}^{\mathrm{lump}}_{\parallel,\mathrm{SR}} = 2Z_c \frac{S_{21}^{\mathrm{REF}} - S_{21}^{\mathrm{DUT}}}{S_{21}^{\mathrm{REF}}} \,. \tag{69}$$

Note that there is no theoretical limit on the impedance magnitude for the determination of purely lumped impedances. A proof, that the measured lumped impedance converges to the beam impedance for decreasing wire radius, is outlined in [42].

3.2.3 Mixed-impedance measurement

Both the lumped (Hahn–Pedersen) and the distributed (improved-log) formulas apply only to their respective types of impedance and give incorrect results for the other. However, practical accelerator components consist of both types, and it is impossible to disentangle them. Thus, a transmission-line measurement interpretation formula is required that applies to both. Such a formula is the (Walling-) log-formula [39]

$$\underline{Z}^{\mathrm{log}}_{\parallel} = 2Z_c \cdot \ln\left(\frac{S_{21}^{\mathrm{REF}}}{S_{21}^{\mathrm{DUT}}}\right) \,, \tag{70}$$

which is obtained from Eq. (67) by neglecting the second term in the square bracket. The requirement for this neglect can be conveniently expressed as

$$\frac{k_{\mathrm{z}}^{\mathrm{DUT}}}{k_{\mathrm{z}}^{\mathrm{REF}}} = \frac{\underline{Z}_c^{\mathrm{DUT}}}{\underline{Z}_c^{\mathrm{REF}}} \approx 1 \,, \tag{71}$$

i.e. the log-formula is valid if the presence of the DUT does not change the characteristic impedance significantly. Contrary-wise, it must be invalid for a long distributed device causing a large attenuation, i.e. a large distributed impedance.

The systematic error of the log-formula for distributed impedances can be quantified by solving it for the logarithm and inserting into Eq. (67). The quadratic equation thereby obtained,

$$\underline{Z}^{\mathrm{dist}}_{\parallel} = \underline{Z}^{\mathrm{log}}_{\parallel} + \frac{\underline{Z}^{\mathrm{log}^2}_{\parallel}}{4i\Theta_{\mathrm{z}}^{\mathrm{REF}} Z_c} \,, \tag{72}$$

has the two solutions

$$\underline{Z}^{\mathrm{log}}_{\parallel} = 2i\Theta_{\mathrm{z}} Z_c \left(-1 \pm \sqrt{1 + \frac{\underline{Z}^{\mathrm{dist}}_{\parallel}}{i\Theta_{\mathrm{z}}^{\mathrm{REF}} Z_c}}\right) . \tag{73}$$

Only the positive solution is physical and gives the length independent error estimate

$$\frac{\underline{Z}^{\mathrm{log}}_{\parallel}}{\underline{Z}^{\mathrm{dist}}_{\parallel}} = 1 + \frac{i}{4}\frac{\underline{Z}^{\mathrm{dist}}_{\parallel}}{\Theta_{\mathrm{z}}^{\mathrm{REF}} Z_c} - \frac{1}{8}\left(\frac{\underline{Z}^{\mathrm{dist}}_{\parallel}}{\Theta_{\mathrm{z}}^{\mathrm{REF}} Z_c}\right)^2 + \cdots \,, \tag{74}$$

which agrees with Hahn's estimate [43] to first order. The systematic error of the log-formula (see Eq. (70)) for lumped impedance can be estimated by inserting the lumped impedance S-parameters (see Eq. (59)),

$$\underline{Z}_{\parallel}^{\log} = -2Z_c \ln \frac{1}{1+\frac{\underline{Z}^{\mathrm{lump}}}{2Z_c}} \,. \tag{75}$$

Taylor expansion results in

$$\frac{\underline{Z}_{\parallel}^{\log}}{\underline{Z}^{\mathrm{lump}}} = 1 - \frac{1}{2}\frac{\underline{Z}^{\mathrm{lump}}}{2Z_c} + \frac{1}{3}\left(\frac{\underline{Z}^{\mathrm{lump}}}{2Z_c}\right)^2 - \cdots \,, \tag{76}$$

i.e. the log-formula reproduces lumped impedances, for $\underline{Z}^{\mathrm{lump}} \ll 2Z_c$. Finally, one can conclude that the log formula is valid for both lumped and distributed impedance, provided the lumped part does not exceed the characteristic impedance of the REF, and the distributed part does not change the characteristic impedance significantly. Obviously, this is true for a small impedance magnitude.

3.3 Transverse impedance

Since the transverse impedance can be measured in a manner similar to the longitudinal one, only the different aspects are discussed. There are two principal methods to measure the transverse impedance: the displaced-wire method and the twin-wire method. In order to enhance the extremely small signals in the twin-wire method at low frequencies, it can be extended to the coil method, which requires a quasi-stationary interpretation.

3.3.1 Displaced-wire method

The displaced-wire technique is based on measuring the dipolar longitudinal impedance and using the Panofsky–Wenzel theorem to obtain the transverse impedance. In a structure with x and y symmetry, the dipolar longitudinal impedance has a quadratic dependence on the transverse offset from the centre (see e.g. Ref. [44]). It can be measured in the same way as the monopolar longitudinal impedance, but with a displaced wire. Subsequently, a parabola can be fitted on the measured results for each frequency point at different transverse positions [45]. However, since a displaced wire measures both the driving and detuning impedance, the driving impedance in one plane can only be obtained if the detuning impedance vanishes, i.e. in a structure that is invariant under 90° rotation [46]. For rectangular structures, the detuning impedance can be cancelled by measuring the impedance in both horizontal and vertical planes and adding the two, but this yields only the sum of both driving impedances.

3.3.2 Twin-wire method

The setup is the same as for the longitudinal impedance, but with two symmetrically driven wires on the differential TEM mode. The characteristic impedance (REF) for the differential TEM mode, i.e. the voltage between the two conductors divided by the current in one conductor, is given by (cf. Ref. [47])

$$Z_c^{\mathrm{dip}} = \frac{Z_0}{\pi} \ln\left(\frac{d+\sqrt{d^2-a^2}}{a} \cdot \frac{b^2 - d\sqrt{d^2-a^2}}{b^2 + d\sqrt{d^2-a^2}}\right) \,, \tag{77}$$

where a is the wire radius, b is outer shield radius and $2d = \Delta$ is the wire distance. With respect to this characteristic impedance, symmetric S-parameters can be defined. The symmetric S_{21}-parameter can be measured best with a four-port VNA, which internally converts the 4×4 S-matrix to a 2×2 matrix for the symmetric signals. There are also approaches to use splitters and combiners with a two-port VNA, but the limited bandwidth of those components makes the calibration (after the hybrids) an involved endeavour. For a four-port VNA the calibration plane can just be chosen before the matching section

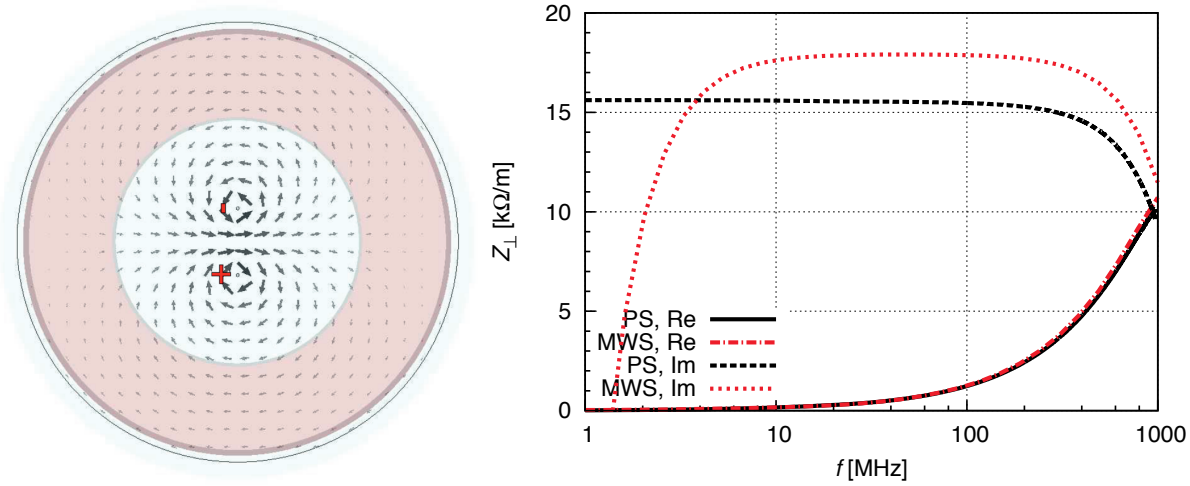


Fig. 16: Magnetic field of dipole TEM eigenmode obtained by the multi-pin-portmode-solver. The wire distance is $\Delta = 10\,\text{mm}$ with an uncertainty of 10% for the measurement. The plot shows the impedance from S-parameter simulation (MWS) compared with wake-field simulation (PS).

(as for the single-wire measurement) and the 18 calibration steps (open, short, match, through) can be significantly eased by using an auto-cal kit.

The twin-wire approximation provides for the ultrarelativistic dipolar transverse impedance [44]

$$\underline{Z}_\perp(\omega) \approx \frac{c}{\omega\Delta^2}\delta\underline{Z}_\parallel(\omega) = \frac{c}{\omega\Delta^2}\cdot 2Z_c^{\text{dip}}\frac{S_{21,\text{dip}}^{\text{REF}} - S_{21,\text{dip}}^{\text{DUT}}}{S_{21,\text{dip}}^{\text{DUT}}}\,, \tag{78}$$

where $\delta\underline{Z}_\parallel$ is the impedance obtained by the conversion formula Eq. (68) for the differential mode S_{21}-parameter and characteristic impedance Z_c^{dip}. Since the EM-fields are mostly confined between the two wires, the de-embedded $S_{21,\text{dip}}$ is very small and one does not have to distinguish between lumped and distributed transverse impedance.

A comparison of the transverse impedance from wake-field and S-parameter simulation for a dispersive ferrite ring is shown in Fig. 16. The agreement between the two is reasonably good, except at low frequency, where the computational accuracy is insufficient. The same is visible also for the lab measurement, as plotted in Fig. 17. Since the DUT alters the EM-fields only slightly in the twin-wire

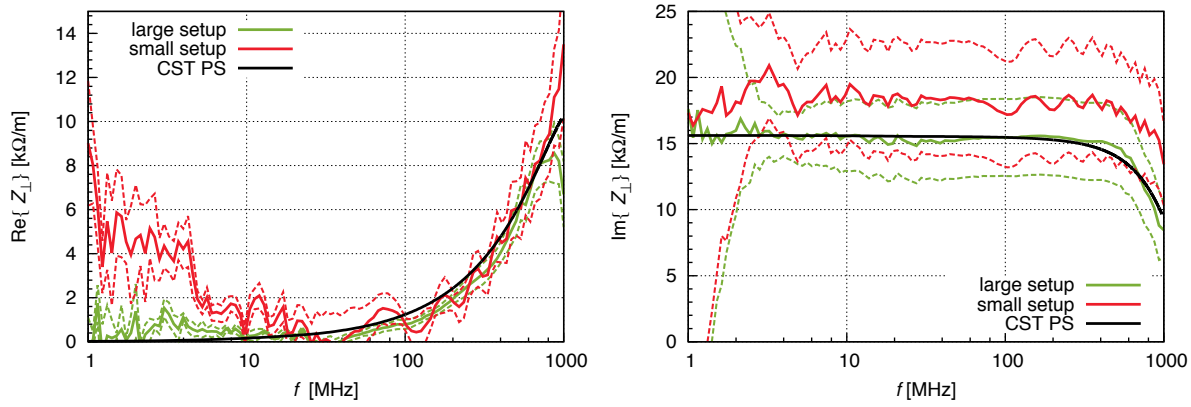


Fig. 17: Transverse impedance of the ferrite ring: measurement vs wake-field simulation. The dashed lines denote error bars.

measurement, the dominating parasitic reflections are the same for DUT and REF measurements. Thus,

they are almost entirely removed by the de-embedding. However, since the difference between DUT and REF measurement is so small, temperature drifts and noise are the main issues. The temperature drift can be reduced by taking metal film resistors instead of carbon resistors for the matching network, which have a higher inductance but a smaller temperature coefficient. The thereby enlarged mismatch is less critical than the temperature drift for the twin-wire measurement. The noise can be reduced by averaging many subsequent DUT and REF measurements.

Figure 17 shows that the result for the ferrite ring in both the large and the small measurement setup is reasonably good at medium and high frequency. However, at low frequencies the method becomes impracticable. This can be improved by employing the coil method instead of the twin-wire method.

3.3.3 Coil method

In order to enhance the extremely small signals in the twin-wire method at low frequencies, the two wires can be replaced by a multi-turn coil. Both the magnetic flux and the induced voltage are magnified by the number of turns N and thus Eq. (78) has to be replaced by

$$\underline{Z}_{\perp}^{\mathrm{coil}} \approx \frac{c}{\omega \Delta^2 N^2} \delta \underline{Z} \,, \tag{79}$$

where the coil impedance difference $\delta \underline{Z} = \underline{Z}^{\mathrm{DUT}} - \underline{Z}^{\mathrm{REF}}$ can be determined by a LCR-meter (we used the Agilent E4980A, 20 Hz-2 MHz). The REF measurement is performed just by measuring the coil outside the DUT in free space. Figure 18 shows the setup and two measurement coils. The coil method has an upper frequency limit, at which the inter-turn capacitance causes a resonance which lies usually in the range of 1 MHz. It can be increased by taking fewer turns and increasing the turn distance. At extremely low frequency, the accuracy is limited by the instrument noise and the temperature drift of the coil resistance. Thus, it makes sense to use different coils: a temperature-stable one (e.g. constantan wire) with many turns for low frequency, and one with few turns and high conductivity (copper wire) for higher frequencies.

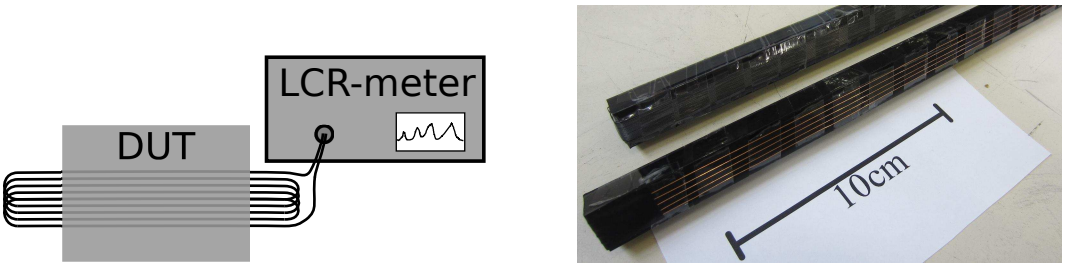


Fig. 18: Coil measurement setup (left-hand side) and different coils (right-hand side). Courtesy of L. Eidam [35]

Since ferrites usually have very small impedance contributions at such low frequencies, the coil method is benchmarked using a steel beam pipe of 2 mm wall thickness and 3.3 cm radius. The real part of the transverse impedance of the pipe, measured with the coils depicted in Fig. 18 is plotted in Fig. 19. Besides the noise at extremely low frequency, the measured real part of the impedance agrees well with the analytic prediction by Rewall [48].

The phase of the coil current does not depend on the longitudinal position, therefore the coil method corresponds to entirely 2D ($\partial_z = 0$) source fields. This does not represent an ultrarelativistic beam, but rather a current-only model, which is sometimes also referred to as 'radial model' and is discussed in [25,49]. The consequence of this is that the imaginary part of the impedance is superimposed by the 'image inductance', i.e. the magnetic part of the indirect transverse space-charge impedance. For a circular pipe of radius b it is given by

$$Z_{\perp}^{\mathrm{image}} = -\mathrm{i} \frac{Z_0 l}{2\pi b^2} \,. \tag{80}$$

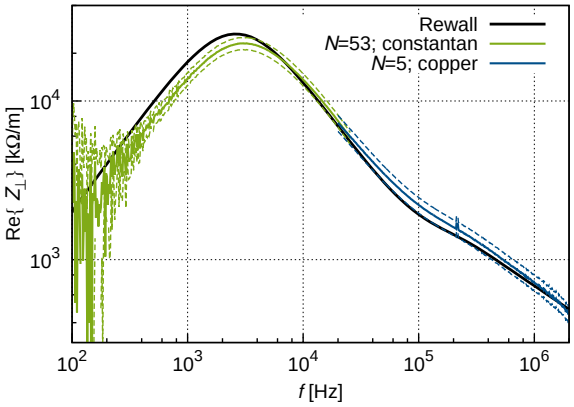
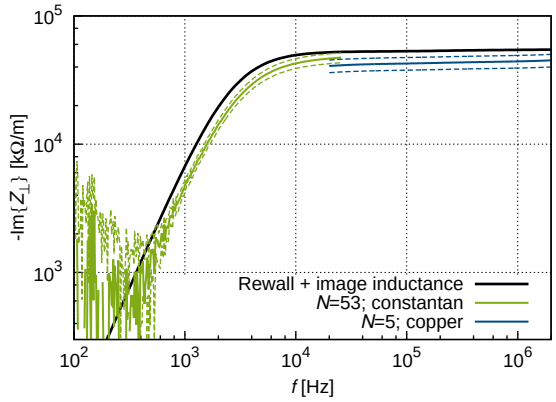
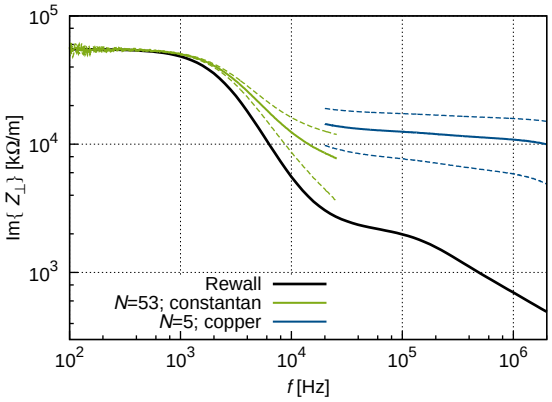


Fig. 19: Coil measurement of real transverse impedance of a tubular-beam pipe vs analytical calculation by Rewall. The dashed lines are error bars.



(a) Imaginary part of the measurement vs Rewall analytical result plus image inductance $-\mathrm{i}5.5 \times 10^4\,\Omega/\mathrm{m}$ from Eq. (80).



(b) Imaginary part of the measurement minus image inductance vs Rewall analytical result.

Fig. 20: Imaginary part from the same measurement as in Fig. 19. Due to the high bias of the image inductance, the imaginary resistive wall impedance measurement result becomes very inaccurate.

The measurement result of the imaginary part of the impedance is shown in Fig. 20. Since the image inductance is much higher than the imaginary part of the resistive wall impedance, small relative measurement errors lead to large relative errors for the imaginary restive wall impedance. Thus, the coil method is effective only for the determination of the real part of the transverse restive wall impedance.

Acknowledgments

I am grateful to Lewin Eidam for providing the pictures of the bench measurements. Moreover, I would like to acknowledge fruitful discussions with the CERN ICE group. Particularly for the bench measurements, discussions with Fritz Caspers and Manfred Wendt were enlightening.

References

[1] T. Weiland and R. Wanzenberg, Wake Fields and Impedances, in Proc. of the CCAS-CAT-CERN Accelerator School, vol. 1: Advanced Accelerator Physics and Technology Course, Indore, India, 7-16 November 1993, edited by S.S. Ramamurthy and L.M. Rangarajan, DESY-M-91-06 (1995), https://cds.cern.ch/record/221360. Accelerator School (1993).

[2] L. Palumbo, V.G. Vaccaro and M. Zobov, Wake Fields and Impedance, in Proc. of the CAS-CERN Accelerator School: 5th Advanced Accelerator Physics Course, Rhodes, Greece, 20 Sept. – 1 Oct. 1993, edited by T. Stuart, CERN-1995-006 (CERN, Geneva, 1995), pp. 331-390, http://dx.doi.org/10.5170/CERN-1995-006.331.

[3] M. Balk, Ph.D. thesis, TU-Darmstadt, 2005.

[4] M. Balk, R. Schuhmann and T. Weiland, *Nucl. Instrum. Methods Phys. Res.* A **558** (2006) 54. http://dx.doi.org/10.1016/j.nima.2005.11.076

[5] W.K.H. Panofsky and W.A. Wenzel, *Rev. Sci. Instrum.* **27** (1956) 967. http://dx.doi.org/10.1063/1.1715427

[6] A. Al-Khateeb *et al.*, *Phys. Rev. ST Accel. Beams* **10** (2007) 064401. http://dx.doi.org/10.1103/PhysRevSTAB.10.064401

[7] R. Unbehauen, *Systemtheorie* (Oldenbourg, Munich, 1983).

[8] K. Küpfmüller, *Einführung in die theoretische Elektrotechnik* (Springer, Berlin Heidelberg, 1932).

[9] R. Gluckstern, Analytic Methods for Calculating Coupling Impedances, CERN-2000-011 (CERN, Geneva, 2000), http://dx.doi.org/10.5170/CERN-2000-011.

[10] P. Monk, *Finite Element Methods for Maxwells Equations* (Oxford University Press, Oxford, 2003). http://dx.doi.org/10.1093/acprof:oso/9780198508885.001.0001

[11] K. Yee, *IEEE Trans. Antennas Propag.* **14** (1966) 302. http://dx.doi.org/10.1109/TAP.1966.1138693

[12] T. Weiland, *Electronics and Communication* **31** (1977) 116.

[13] E. Gjonaj *et al.*, *ICFA Beam Dynamics Newsletter* **45** (2008).

[14] A.W. Chao et al. (Eds.), *Handbook of Accelerator Physics and Engineering*, 2nd ed. (World Scientific, Singapore, 2013).

[15] H. Kawaguchi and T. Honma, *Nucl. Instrum. Methods Phys. Res.* A **363** (1995) 145. http://dx.doi.org/10.1016/0168-9002(95)00364-9

[16] K. Fujita *et al.*, *IEEE Trans. Nucl. Sci.* **53** (2006) no. 2.

[17] E. Gjonaj, T. Lau and T. Weiland, Wakefield Computations with the PBCI Code using a Non-Split finite Volume Method, Proc. 2009 Particle Accelerator Conf. (JACoW.org, 2009).

[18] A. Tsakanian *et al.*, A new Approach for Resistive Wakefield Calculations in Time Domain, Proc. 6th International Particle Accelerator Conf. (JACoW.org, 2015).

[19] K. Yokoya, Resistive Wall Impedance of Beam Pipes of General Cross Section, KEK Preprint 92-196 (1993).

[20] A. Macridin, P. Spentzouris and J. Amundson, *Phys. Rev. ST Accel. Beams* **16** (2013) 121001. http://dx.doi.org/10.1103/PhysRevSTAB.16.121001

[21] K. Bane, P. Wilson and T. Weiland, Wake Fields and Wake Field Acceleration, SLAC-PUB-3528 (1984).

[22] W. Ackermann and T. Weiland, High Precision Cavity Simulations, Proc. International 2012 Particle Accelerator Conf. (JACoW.org, 2012).

[23] K. Klopfer, W. Ackermann and T. Weiland, *IEEE Trans. Magn.* **51** (2015) no. 1.

[24] T. Kroyer, Simulation of the Low-frequency Collimator Impedance, CERN-AB-Note-2008-017 (2008).

[25] U. Niedermayer and O. Boine-Frankenheim, *Nucl. Instrum. Methods Phys. Res.* A **687** (2012) 51. http://dx.doi.org/10.1016/j.nima.2012.05.096

[26] O. Kononenko and A. Grudiev, *Phys. Rev. ST Accel. Beams* **14** (2011) 111001. http://dx.doi.org/10.1103/PhysRevSTAB.14.111001

[27] R. Courant, K. Friedrichs and H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik, *Math. Ann.* **100** (1928) 32. http://dx.doi.org/10.1007/BF01448839

[28] F. Edelvik, R. Schuhmann and T. Weiland, *Int. J. Numer. Model.* **17** (2004) 407. http://dx.doi.org/10.1002/jnm.547

[29] I. Zagorodnov and T. Weiland, *Phys. Rev. ST Accel. Beams* **8** (2005) 042001. http://dx.doi.org/10.1103/PhysRevSTAB.8.042001

[30] P. Ingelström, *IEEE Trans. Microwave Theory Tech.* **54** (2006) 106. http://dx.doi.org/10.1109/TMTT.2005.860295

[31] U. Niedermayer, O. Boine-Frankenheim and H. De Gersem, *Phys. Rev. ST Accel. Beams* **18** (2015) 1. http://dx.doi.org/10.1103/PhysRevSTAB.18.032001

[32] M. Sands and J. R. Rees, A Bench Measurement of the Energy Loss of a Stored Beam to a Cavity, SLAC-TN-05-051 (1974).

[33] M. Hiebel, *Fundamentals of Vector Network Analysis* (Rohde & Schwarz, Munich, 2005).

[34] D. Pozar, *Microwave Engineering* (John Wiley & Sons, Hoboken, NJ, 1998).

[35] U. Niedermayer, L. Eidam and O. Boine-Frankenheim, *Nucl. Instrum. Methods Phys. Res.* A **776** (2015) 129. http://dx.doi.org/10.1016/j.nima.2014.12.053

[36] V. G. Vaccaro, Coupling Impedance Measurements: An improved wire method, INFN/TC-94/023 (1994).

[37] J. Wang and S. Zhang, *Nucl. Instrum. Methods Phys. Res.* A **459** (2001) 381. http://dx.doi.org/10.1016/S0168-9002(00)01024-X

[38] H. Hahn and F. Pedersen, On Coaxial Wire Measurements of the Longitudinal Coupling Impedance, BNL Report 78-9 (1978).

[39] L. S. Walling *et al.*, *Nucl. Instrum. Methods Phys. Res.* A **281** (1989) 433. http://dx.doi.org/10.1016/0168-9002(89)91474-5

[40] E. Jensen, An improved log-formula for homogeneously distributed impedance, CERN PS/RF/Note 2000–001 (2000).

[41] E. Karantzoulis, An Overview of Impedances and Impedance Measuring Methods for Accelerators, Sincrotrone Trieste, Tech. Rep. (1991).

[42] A. Argan *et al.*, On the Sands and Rees Measurement Method of the Longitudinal Coupling Impedance, in Proc. 8th Particle Accelerator Conf. (JACoW.org, 1999), p. 1599.

[43] H. Hahn, *Phys. Rev. ST Accel. Beams* **3** (2000) 122001. http://dx.doi.org/10.1103/PhysRevSTAB.3.122001

[44] G. Nassibian and F. Sacherer, *Nucl. Instrum. Methods* **159** (1978) 21.

[45] E. Metral *et al.*, Kicker impedance measurements for the future multiturn extraction of the CERN Proton Synchrotron, Proc. European Particle Accelerator Conf., Edinburgh, 2006 (JACoW.org, 2006).

[46] A. Burov and V. Danilov, *Phys. Rev. Lett.* **82** (1999) 2286. http://dx.doi.org/10.1103/PhysRevLett.82.2286

[47] J. Wang and S. Zhang, *Nucl. Instrum. Methods Phys. Res.* A **459** (2001) 381. http://dx.doi.org/10.1016/S0168-9002(00)01024-X

[48] N. Mounet and E. Metral, Electromagnetic field created by a macroparticle in an infinitely long and axisymmetric multilayer beam pipe, CERN–BE–2009–039 (2009).

[49] H. Hahn, *Phys. Rev. ST Accel. Beams* **13** (2010) 012002. http://dx.doi.org/10.1103/PhysRevSTAB.13.012002

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam-Based Impedance Measurements

E. Shaposhnikova
CERN, Geneva, Switzerland

Abstract

Beam-based impedance measurements play an important role in benchmarking existing impedance models of the accelerator, as well as in elaborating them. Impedance measurements can be made with both stable and unstable beams. In the first case, one makes use of changes in stable bunch parameters, such as bunch length, synchrotron frequency distribution, or synchronous phase shift. In the second case, measurements of instability characteristics (threshold, growth rates, bunch spectrum) can be used for impedance search or evaluation, usually by comparison with results of particle simulations.

Keywords

Impedance; beam measurements.

1 Introduction

First of all, let us answer the question; why does one need to measure impedance with a beam? Indeed, nowadays the beam-coupling impedance of various machine elements can be accurately estimated using analytical calculations [1, 2], advanced electromagnetic simulations (various codes are available), or bench measurements in the lab, e.g., see Refs. [3, 4].

Nevertheless, very often one needs to verify the accuracy of an impedance model, based on electromagnetic simulations or measurements, since there are always some very complex machine elements with impedances that are difficult to calculate, simulate, or measure; the material properties of these elements are also not always well known. In addition, non-conformities may also exist, from either fabrication errors or beam-induced damage owing to e.g., operation with high-intensity beams. Measurements with the beam can also be useful in identifying impedance sources that are driving beam instabilities or posing some other intensity limitations.

In what follows, we consider methods mainly used in circular proton accelerators with relatively high beam energy (above the gigaelectronvolt range) and long bunches (above the nanosecond range). The frequency ranges of interest for impedance measurements, and therefore the approaches used in these machines, are quite different from those applied in the synchrotron light sources with picosecond or even femtosecond bunches. Thanks to careful initial design, the impedances of modern rings become smaller and smaller, so that more elaborate methods are required to measure them with beam. However, numerical simulations of various collective effects have also become more advanced and can be used for comparison with beam tests of impedance.

Note that practically all intensity effects could potentially be used for impedance evaluation by comparison of beam measurements with particle simulations or analytical formulae. Only a few methods could be discussed in detail in this paper. Their selection is based on personal experience in using them on CERN machines and also on the fact that some of them are probably less well known outside CERN. Most of the examples are given for measurements in the longitudinal plane, but similar techniques are often applicable in the transverse plane.

This paper consists of two main parts. The first describes measurements with stable bunches that include bunch lengthening, synchrotron frequency shift, and change in debunching time with intensity, and are applied for evaluation of the reactive impedance $\mathrm{Im}Z$. Measurements of synchronous phase shift with intensity can be used to estimate the resistive part of the beam-coupling impedance $\mathrm{Re}Z$. The

https://doi.org/10.23730/CYRSP-2017-003.107

second part of the paper deals with an unstable beam and its characteristics (spectra, growth rates, and thresholds). In all these cases, the impedance evaluation is based on changes in beam characteristics. However, measurements with stable beams are mainly used to test the existing impedance models while measurements with unstable beams often contain important information about parameters of the dominant offending impedance. There is also a separate case (not considered in this paper) when the impedance of a particular element in the ring (e.g., an RF cavity) can be evaluated from the signal excited there by a single bunch with known (measured) profile (see, e.g., Ref. [5,6]). This approach can be considered an intermediate case between beam measurements and bench measurements of impedance in the lab, since the beam characteristics stay unchanged.

Measurements of impedance with a single bunch can give information only about the effective impedance—the actual impedance integrated over the spectrum of the bunch. For a stable bunch, the effective impedance is defined by integration over the stable bunch spectrum centred at zero frequency. Since the width of the bunch spectrum is inversely proportional to the bunch length τ, long bunches 'see' only the low-frequency ($f < 1/\tau$) part of the coupling impedance. For an unstable bunch, the situation is different. The growth rate of some mode depends on the effective impedance, which is now defined by integration over the spectrum of this mode with a non-zero centre frequency.

The total voltage seen by a particle is the sum of the RF voltage $V_{\rm rf}$ and the induced voltage $V_{\rm ind}$,

$$V = V_{\rm rf} + V_{\rm ind}\,. \tag{1}$$

The induced voltage due to beam-coupling impedance $Z(\omega)$ contains two contributions: the first defined by the stable bunch spectrum $\Lambda(\omega)$ and the second by the unstable spectrum $h(\omega)$. In the next section, we will consider the intensity effects defined by the stable bunch spectrum; in Section 3, we will consider the those defined by the unstable spectrum.

2 Impedance measurements with stable beam

In equilibrium, the particle distribution is a function of the Hamiltonian H with a potential well defined by the total voltage (1) seen by the particle (the effect of potential well distortion). The induced voltage can be written in the following form (see, e.g., Refs [7, 8]):

$$V_{\rm ind}(\theta) = -e\omega_0 N_{\rm b} \sum_n \Lambda_n Z_n {\rm e}^{{\rm i}n\theta}\,, \tag{2}$$

where $\omega_0 = 2\pi f_0$ is the revolution frequency, θ is the longitudinal co-ordinate of the particle expressed in radians, $Z_n = Z(n\omega_0)$, and

$$\Lambda_n = \frac{1}{2\pi}\int \lambda(\theta){\rm e}^{-{\rm i}n\theta}{\rm d}\theta \tag{3}$$

is the nth Fourier harmonic of the unperturbed bunch line density $\lambda(\theta)$ containing $N_{\rm b}$ particles and normalized to unity

$$\int \lambda(\theta){\rm d}\theta = 1\,. \tag{4}$$

Equation (2) is sufficient to obtain, in a self-consistent way, the equilibrium particle distribution, which provides measurable dependence of synchrotron frequency $\omega_{\rm s}$, bunch length τ, and synchronous phase on bunch intensity. The Haissinski equation describes the situation for an electron bunch in equilibrium, assuming the Gaussian distribution in energy [9]. However, there is no unique solution for proton bunches and measured bunch profiles should be used to obtain the required equilibrium bunch characteristics.

Assuming a symmetric bunch profile with $\Lambda_n = \Lambda_{-n}$, Eq. (2) becomes

$$V_{\rm ind}(\theta) = -2\pi I_{\rm b} \sum_n \Lambda_n ({\rm Re}Z_n \cos n\theta - {\rm Im}Z_n \sin n\theta)\,, \tag{5}$$

where $I_\mathrm{b} = ef_0N_\mathrm{b}$ is the average beam current. For $n\theta \ll 1$, the right-hand side of Eq. (5) can be expanded to give

$$V_\mathrm{ind}(\theta) \simeq -2\pi I_\mathrm{b} \sum_n \Lambda_n(\mathrm{Re}Z_n - n\theta\,\mathrm{Im}Z_n + ...)\,, \tag{6}$$

with $\mathrm{Re}Z$ leading to the synchronous phase shift and $\mathrm{Im}Z$ contributing to the change of the RF voltage amplitude and, therefore, introducing the synchrotron frequency shift.

Adding the induced voltage to the RF voltage in the linearized longitudinal equation of motion yields

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \omega_{\mathrm{s}0}^2\left[\theta + \frac{V_\mathrm{ind}(\theta)}{V_\mathrm{rf}\,h\cos\phi_\mathrm{s}}\right] = 0\,, \tag{7}$$

where ϕ_s is the synchronous phase ($\phi = h\theta$, with RF harmonic number h) and $\omega_{\mathrm{s}0}$ is the synchrotron frequency of particles with small oscillation amplitude.

In this approximation, we get the following expression for the synchronous phase shift:

$$\Delta\phi_\mathrm{s} = h\Delta\theta \simeq \frac{2\pi I_\mathrm{b}}{V_\mathrm{rf}\cos\phi_\mathrm{s}} \sum_n \Lambda_n\,\mathrm{Re}Z_n\,, \tag{8}$$

which will be discussed in the corresponding section.

For the linear synchrotron frequency, we obtain

$$\omega_\mathrm{s}^2 \simeq \omega_{\mathrm{s}0}^2\left(1 + \frac{2\pi I_\mathrm{b}}{V_\mathrm{rf}h\cos\phi_\mathrm{s}} \sum_n n\Lambda_n\,\mathrm{Im}Z_n\right)\,. \tag{9}$$

We notice already that a dominant space charge impedance ($\mathrm{Im}Z < 0$) below transition ($\cos\phi_\mathrm{s} > 0$) or inductive impedance above transition leads to negative frequency shift. Note that Eqs. (8)–(9) are applicable only for small-amplitude particles and the dependence of the measured synchrotron frequency shift and phase on beam parameters can, in reality, be much more complicated.

2.1 Frequency shifts

For a reactive impedance $\mathrm{Im}\mathrm{Z}_n/n$, which is constant over the stable bunch spectrum, and for small shifts $\Delta\omega_\mathrm{s} \ll \omega_{\mathrm{s}0}$, Eq. (9) can be rewritten in the form

$$\Delta\omega_\mathrm{s} = \omega_\mathrm{s} - \omega_{\mathrm{s}0} \simeq \frac{\pi I_\mathrm{b}\omega_{\mathrm{s}0}}{V_\mathrm{rf}h\cos\phi_\mathrm{s}}\mathrm{Im}Z/n \sum_n n^2\Lambda_n\,. \tag{10}$$

Using the fact that

$$\lambda(\theta) = \sum_n \Lambda_n \mathrm{e}^{-\mathrm{i}n\theta}\,, \tag{11}$$

one can see that, for small-amplitude particles, the sum in Eq. (10) can be replaced by the second derivative of the bunch line density $\lambda(\theta)$, and we obtain

$$\Delta\omega_\mathrm{s} \simeq -\frac{\pi I_\mathrm{b}\omega_{\mathrm{s}0}\mathrm{Im}Z/n}{V_\mathrm{rf}h\cos\phi_\mathrm{s}}\left.\frac{\mathrm{d}^2\lambda}{\mathrm{d}\theta^2}\right|_{\theta=0}\,, \tag{12}$$

where the second derivative is taken at the bunch centre.

Note that for a constant $\mathrm{Im}Z/n$, a similar formula can be obtained directly from the expression for induced voltage $V_\mathrm{ind} = -L\mathrm{d}I/\mathrm{d}t$, where $I(t)$ is the instantaneous beam current and the inductance L can be found from the expression $\mathrm{i}\omega_0 L = \mathrm{Im}Z/n$. Then, Eq. (2), rewritten for $\mathrm{Im}Z/n = \mathrm{const}$, can also be reproduced.

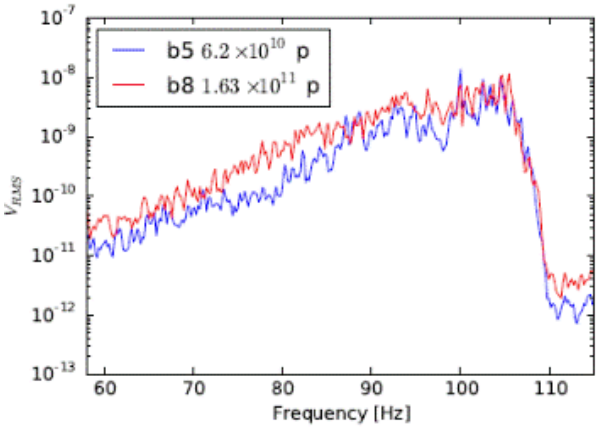


Fig. 1: Quadrupole line of peak-detected Schottky signal, proportional to particle distribution in synchrotron frequency, for two LHC bunches of Beam 1 with similar length of 1.4 ns (4σ Gaussian fit) but different intensities (0.1×10^{11} and 1.1×10^{11}) at 450 GeV/c ($f_{s0} = 55.1$ Hz). The difference $2\delta f_s$ is less than 1.0 Hz and $\delta f_s = 0.35$ Hz is expected from the LHC impedance model [13].

For a parabolic bunch in a linear RF voltage, the expression for synchrotron frequency shift becomes particularly simple:

$$\omega_s^2 = \omega_{s0}^2 + \frac{6\omega_{s0}^2 I_b \mathrm{Im}Z/n}{V_{rf} h \cos\phi_s (\omega_0\tau)^3} \,. \tag{13}$$

To obtain an estimate of $\mathrm{Im}Z/n$ of the ring, the synchrotron frequency shift should be measured as a function of bunch intensity N_b. Different possible methods are described next.

2.1.1 Incoherent synchrotron frequency shift

The incoherent synchrotron frequency shift can be found for a bunch in equilibrium by measuring the distance $2m\Delta f_s$ between positive and negative mth synchrotron sidebands of the longitudinal Schottky spectrum [10]. This method was used in both RHIC rings [11], where the dependence on intensity was obtained from the natural intensity decay during luminosity production. The parabolas were fitted to the top 30% of the averaged bunch profiles to find $\mathrm{d}^2\lambda/\mathrm{d}\theta^2$. The results obtained by this method for the two RHIC rings, blue and yellow, which are very similar, differed by more than a factor of three; the source of this difference is not clear.

The quadrupole ($m = 2$) line of the peak-detected Schottky spectrum contains information about the particle distribution in synchrotron frequency [12] and can be used to observe the synchrotron frequency shift. The measurements of the $m = 2$ line performed at the bottom energy of the CERN LHC for two bunches of similar bunch length and different intensities are shown in Fig. 1. As one can see, the available frequency resolution of 0.2 Hz is insufficient and only an upper limit on $\mathrm{Im}Z/n$ could be obtained ($< 0.2\ \Omega$) [13]. This limit agrees with the current LHC impedance budget of $0.1\ \Omega$.

Another method that can be used to estimate the synchrotron frequency shift, when applied in the LHC, gave similar results. Eight bunches with intensities in the range 0.6×10^{11}–2.0×10^{11} and bunch length in the range 1.2–1.4 ns were excited via a cavity set point by phase modulation $\phi(t) = \phi_0 \sin(2\pi f_{mod} t)$ with modulation frequency f_{mod}, changing in steps of 0.1 Hz from the zero-intensity synchrotron frequency $f_{s0} = 55.1$ Hz. Dipole oscillations of different bunches were observed at excitation frequencies, reaching the synchrotron frequency spread inside these bunches. The results are again in agreement with an expected maximum frequency shift of 0.11 Hz. Owing to the finite length of this excitation (and therefore the frequency bandwidth), a constant offset in synchrotron frequencies was also observed. To improve accuracy, longer excitations were applied for shorter bunches (available at the LHC flat top) in recent machine studies. Finally, the LHC impedance ($\mathrm{Im}Z/n = 0.09\ \Omega$) could be

estimated most accurately from the measurements (using bunches with various lengths and intensities) of thresholds of the loss of Landau damping caused by the incoherent synchrotron frequency shift [13], but these results are not discussed here.

2.1.2 Coherent synchrotron frequency shift

The synchrotron frequency shift can also be measured from excited oscillations of bunches with different intensities N_b. In this case, we are dealing with the coherent synchrotron frequency shift as well as the incoherent shift, since now the bunch spectrum consists of both stationary and oscillating components. The frequency of bunch oscillations $\omega_m = 2\pi f_m$ can be presented in the form (see, e.g., Ref. [8])

$$\omega_m = m(\omega_{\mathrm{s}0} + \Delta\omega_\mathrm{inc}) + \Delta\omega_\mathrm{coh}\,, \tag{14}$$

where $\Delta\omega_\mathrm{inc}$ and $\Delta\omega_\mathrm{coh}$ are the incoherent and coherent synchrotron frequency shifts, respectively. The two last terms in Eq. (14) are defined by the two different effective impedances. Indeed, the incoherent frequency shift $\Delta\omega_\mathrm{inc} \propto \mathrm{Im}Z_0$ and the coherent frequency shift $\Delta\omega_\mathrm{inc} \propto (\mathrm{Im}Z/\omega)_m^\mathrm{eff}$, where

$$(\mathrm{Im}Z/\omega)_m^\mathrm{eff} = \frac{\sum_{p=-\infty}^{\infty} h_m(\omega_{pm}) Z(\omega_{pm})/\omega_{pm}}{\sum_{p=-\infty}^{\infty} h_m(\omega_{pm})} \tag{15}$$

and $\omega_{pm} = p\omega_0 + m\omega_s$.

For a Gaussian bunch with r.m.s. bunch length σ, the spectrum function is

$$h_m(\omega) = (\omega\sigma)^{2m} \mathrm{e}^{-(\omega\sigma)^2} \tag{16}$$

and

$$Z_0 \simeq \sum_{p=-\infty}^{\infty} p\,\mathrm{Im}Z(\omega_{p0}) \mathrm{e}^{-(\omega_{p0}\sigma)^2/2}\,. \tag{17}$$

For dipole oscillations ($m = 1$), the last two terms in Eq. (14) practically cancel each other (exactly, for a parabolic bunch in a linear RF voltage). Thus, for beam measurements we are left with quadrupole ($m = 2$) oscillations, which, for example, can be excited at injection into a mismatched RF voltage or by a non-adiabatic increase of voltage. The frequency of bunch length (or bunch peak amplitude) oscillations can be found from fitting the first 12–13 oscillations with a sine wave or from the maximum frequency in the peak-detected Schottky spectrum [14]. The variation of bunch intensity allows the dependence of oscillation frequency on impedance to be estimated using the expression

$$f_\mathrm{2s} = f_\mathrm{2s}(N_\mathrm{b} = 0) + bN_\mathrm{b}\,, \tag{18}$$

where for $\mathrm{Im}Z/n = \mathrm{const}$ the slope $b \propto \mathrm{Im}Z/n$ [7]. Note that the slope b also strongly depends on bunch length (as $1/\tau^3$ for $\mathrm{Im}Z/n = \mathrm{const}$) and special care should been taken when making the reference impedance measurements, as in the CERN SPS [14], by using bunches with similar bunch lengths and also emittances.

Indeed, single bunches injected into mismatched voltage at 26 GeV/c (above transition) have been used in the CERN SPS to evaluate changes in longitudinal inductive impedance since 1999, see Fig. 2 (left-hand side). The first significant reduction in the inductive impedance (the slope b) could be seen after shielding the $\sim$900 pumping ports in 2000 (compare measurements from 1999 and 2001). This was followed by an impedance increase due to installation in 2003 and 2006 of kickers for beam extraction to the two LHC rings. Later, the impedance of a few kickers was significantly reduced, but the effect was no longer measurable with the beam [14], mainly owing to variation of the injected bunch length (emittance) in measurements. Recently, measurements of synchrotron frequency shift as a function of bunch length allow the frequency dependence of effective impedances to be studied, see Fig. 2 (right-hand side). A comparison of these measurements with particle simulations can serve as a good test of

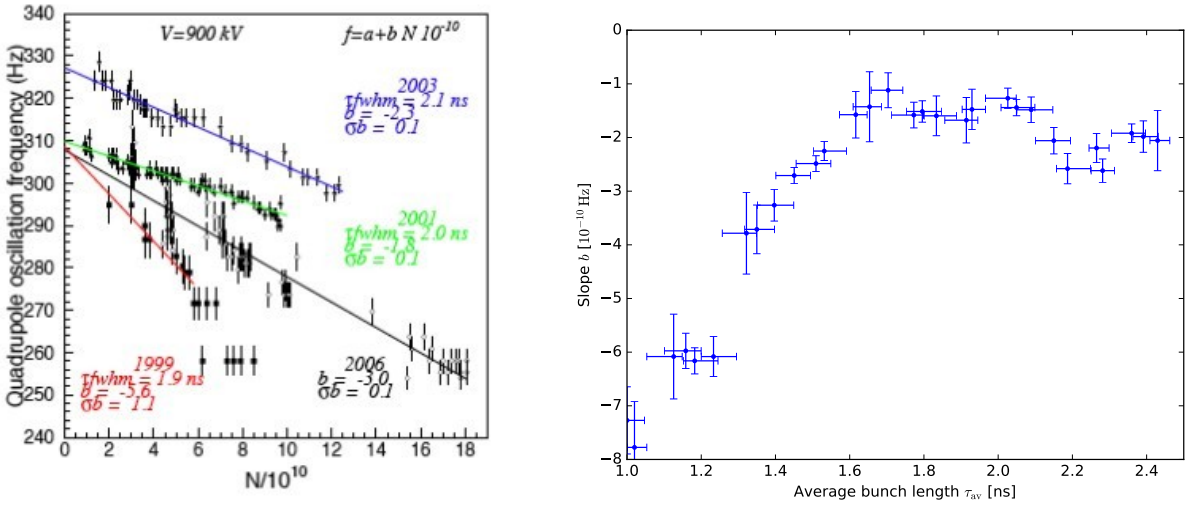


Fig. 2: Left-hand side: measurements of quadrupole frequency shift as a function of intensity (slope b) over a period of years in the CERN SPS following up the impedance evolution of the ring [14]. Right-hand side: recent measurements of the slope b from Eq. (18) in [Hz/10^{10}] as a function of average bunch length [15].

the impedance model of the ring. In the case of the CERN SPS, this comparison reveals some missing inductive impedance $\mathrm{Im}Z/n \sim 0.3\,\Omega$ [15]. The analysis also shows that for the SPS impedance measured using the quadrupole oscillations, the frequency shift in Eq. (14) is dominated by the contribution from the incoherent frequency shift.

Changes in the transverse SPS impedance were also observed over many years by performing measurements of the vertical tune shift with intensity [16].

2.2 Debunching

The voltage induced by the bunch produces 'potential well distortion' and changes the synchrotron frequency distribution when RF is on, but it also affects the beam dynamics when RF is off. The effective reactive impedance of the ring can then also be estimated by measuring the evolution of the bunch parameters during the debunching process [17].

For a parabolic bunch, the variation with time of length τ and the peak line density λ_p in the presence of reactive impedance $\mathrm{Im}Z = \mathrm{const}$ with RF off,

$$\tau(t) = \tau(0)\, r(t)\,, \qquad \lambda_\mathrm{p}(t) = \lambda_\mathrm{p}(0)/r(t)\,, \tag{19}$$

is described by the function $r(t)$. At the beginning of debunching, it has the form [18]

$$r(t) \simeq (1 + \Omega_\mathrm{d}^2 t^2)^{1/2}\,, \qquad \text{with} \quad \Omega_\mathrm{d}^2 = \Omega^2 + s\Omega_N^2\,, \tag{20}$$

where

$$\Omega = \frac{2|\eta|}{\tau(0)|}\frac{\Delta p_m}{p}\,, \qquad \Omega_N^2 = \frac{6N_\mathrm{b}e^2|\eta|}{\pi E\tau^3}\,\mathrm{Im}Z/n\,, \tag{21}$$

$\pm\Delta p_m/p$ is the maximum relative momentum spread in the bunch and $s = \mathrm{sign}(\eta \mathrm{Im}Z)$. The plus sign gives faster debunching, owing to the defocusing effect of inductive impedance above transition, or capacitive below transition, see Fig. 3 (top).

Here, the parameter Ω_N is similar to that contributing to the incoherent synchrotron frequency shift in Eq. (14). For $N_\mathrm{b} = 0$, the matched bunch has $\Omega = \omega_{\mathrm{s}0}$, otherwise it is defined by the RF parameters

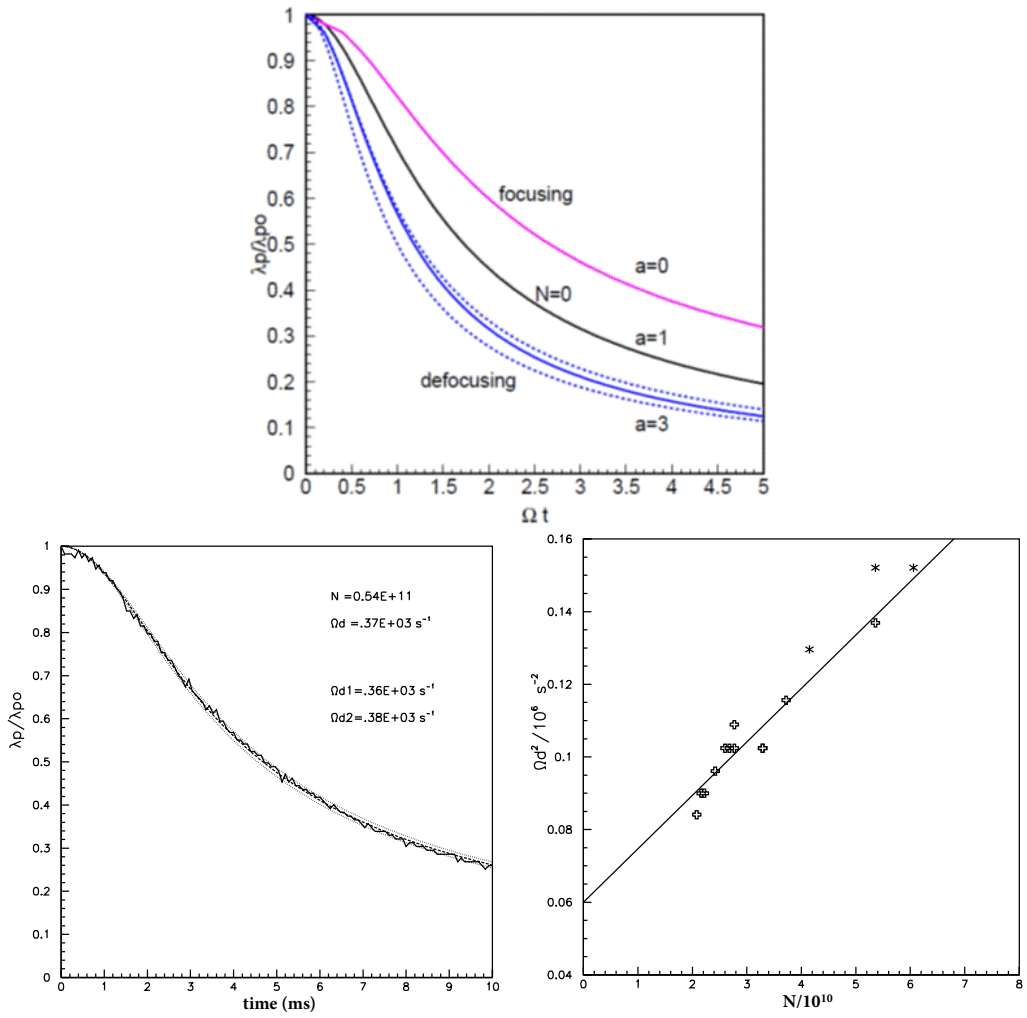


Fig. 3: Top: decay of normalized peak line density $1/r(t)$ with time for different values of parameter $a = 1 + 2s\,\Omega_N^2/\Omega^2$. $a = 1$ corresponds to absence of intensity effects. No debunching will happen for $a < 0$. Dashed lines are approximations of the exact solution valid at $t \ll 1/\Omega_\mathrm{d}$ and $t \gg 1/\Omega_\mathrm{d}$ [17]. Bottom left-hand side: example of measured peak line density variation during debunching in the SPS at 26 GeV/c and its fit using $r(t)$ from Eq. (20). Bottom right-hand side: measured Ω_d^2 as a function of intensity.

(synchrotron frequency) of the injector. This means that if the RF is switched off for a matched bunch with $\Omega^2 = \omega_{\mathrm{s}0}^2 - s\Omega_N^2$ then, as follows from Eq. (20), the debunching time $t_\mathrm{d} = 1/\Omega_\mathrm{d}$ in the first approximation does not depend on intensity.

The debunching time measured as a function of intensity from the decay of peak line density at 26 GeV/c in the CERN SPS is shown in Fig. 3. The estimate of $\mathrm{Im}Z/n$ (18.7 Ω) obtained by this method [17] is slightly larger than values found at that time (before the first impedance reduction) by other methods (with RF on), mainly because longer bunches during debunching sample lower frequencies and therefore a higher inductive impedance.

2.3 Bunch lengthening

Measurements of bunch lengthening with intensity under stable conditions are often used to estimate the effective reactive impedance of the ring. The Haissinski equation [9] describes the equilibrium bunch profile when, owing to the effect of synchrotron radiation, the momentum distribution is Gaussian, and therefore is normally applicable only in lepton machines.

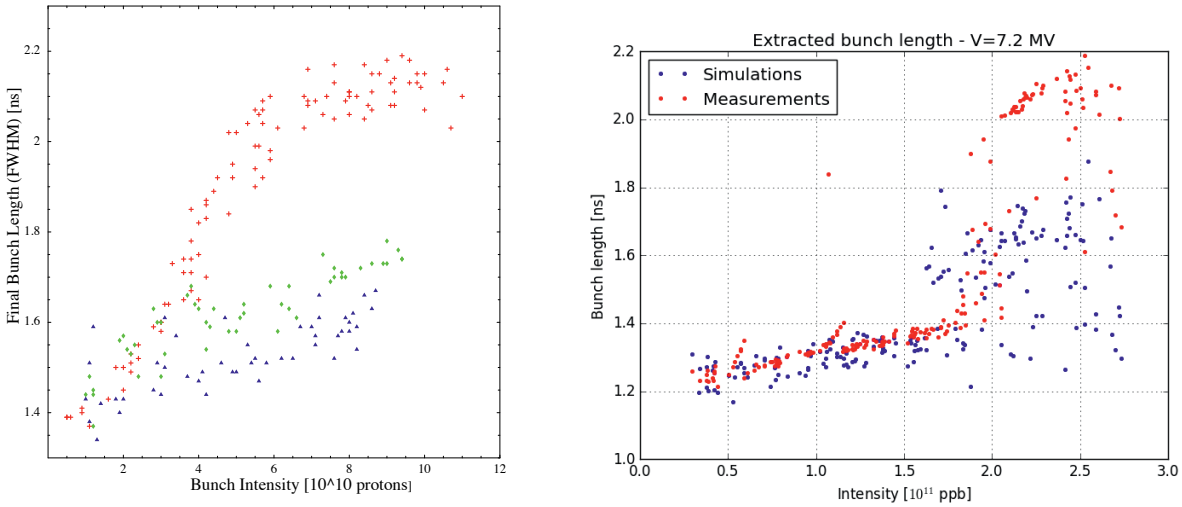


Fig. 4: Left-hand side: bunch length as a function of intensity on the 26 GeV/c SPS flat bottom at 600 ms after injection, before (red dots) and after (green and blue) the first impedance reduction in 1999–2001. Right-hand side: bunch length obtained from recent (2015) measurements (red) on the 450 GeV/c SPS flat top and macroparticle simulations through the whole acceleration cycle using initial bunch parameters from the measurements (blue) and full SPS impedance model [15].

The equation that describes bunch lengthening for protons [7] is based on the fact that longitudinal emittance is an invariant of motion. For parabolic bunches and constant inductive impedance $\mathrm{Im}Z/n$, the bunch length τ normalized to zero-intensity value τ_0 satisfies the following equation

$$\left(\frac{\tau}{\tau_0}\right)^4 + A\frac{\tau}{\tau_0} - 1 = 0\,, \qquad \text{where} \quad A = \frac{24\pi I_\mathrm{b}\mathrm{Im}Z/n}{(\omega_0\tau_0)^3 V_\mathrm{rf} h\cos\phi_\mathrm{s}}\,. \tag{22}$$

Comparison of measured and calculated bunch lengthening with intensity gives an estimate of reactive impedance, assuming $\mathrm{Im}Z/n = \mathrm{const}$. For proton bunches, the experiment is not easy, owing to the need for bunches with variable intensity but constant longitudinal emittance. If these bunches are injected into the ring, the constant RF voltage at injection can only be matched for a certain bunch intensity, leading at other intensities to quadrupole oscillations, filamentation, and emittance blow-up. Macroparticle simulations could be performed for the actual experimental set-up using longitudinal impedance having a more complicated, but realistic structure. Usually, bunch lengthening due to potential well distortion can be easily distinguished from that due to an instability; see two examples of SPS bunch lengthening measurements on the flat bottom and flat top in Fig. 4.

2.4 Synchronous phase shift

Measurements of the synchronous phase shift as a function of intensity are often used to evaluate the resistive impedance of the ring that abstracts particle energy [19]. In the absence of acceleration, the synchronous phase ϕ_s is defined by the expression

$$\Delta\phi_\mathrm{s} = -U/(e\,V_\mathrm{rf}\cos\phi_\mathrm{s})\,, \tag{23}$$

where U is the energy loss per turn and per particle.

Equation (8) describes the phase shift of a single particle with a small synchrotron oscillation amplitude. In the same way as energy loss of a given particle, this phase shift depends on particle oscillation amplitude. Experimentally, only the total energy loss of the whole bunch can be measured. The total energy loss normalized to the number of particles can be found by measuring the synchronous phase

shift $\Delta\phi_{\rm s}$ at different bunch intensities. The measured dependence of energy loss on bunch length can be compared with that calculated from the known resistive impedances and the given bunch distribution [20].

The energy loss of the whole bunch per turn and per particle can be found from the following expression [7]:

$$U = -e^2 N_{\rm b} k\,, \tag{24}$$

with the loss factor:

$$k = \frac{\omega_0}{\pi} \sum_{p=0}^{\infty} {\rm Re}Z(p\omega_0)|\Lambda(p\omega_0)|^2\,. \tag{25}$$

Finally, one obtains

$$\Delta\phi_{\rm s} = \frac{2I_{\rm b}}{V_{\rm rf}\cos\phi_{\rm s}} \sum_{p=0}^{\infty} {\rm Re}Z(p\omega_0)|\Lambda(p\omega_0)|^2\,. \tag{26}$$

For a Gaussian bunch, the spectrum $\Lambda(\omega) = \exp[-(p\omega_0\sigma)^2/2]$ and one can see that Eq. (8) and Eq. (26) give the same result for $\sigma \ll 1/\omega_{\rm r}$.

The shift of the synchronous phase $\Delta\phi_{\rm s}$ can be measured from the distance between the two bunches in the ring or from the phase of the beam signal relative either to the reference RF signal or to the signal from a probe in the RF cavity. When using the reference RF signal (sent from the power amplifier to the cavity), the energy loss due to the cavity fundamental impedance is included. The signal from the probe in the cavity contains information about the sum of the applied RF voltage and the beam-induced voltage, so that in this case the beam-loading effect will be excluded from the measured phase shift. Measuring the distance between a time reference, low-intensity bunch and a witness bunch with varied intensity (see, e.g., Ref. [21]) is similar to the use of the reference RF signal, since the measured loss factor can be dominated by the contribution from the RF cavities.

Measurements of synchronous phase shift made in the CERN SPS after the first impedance reduction using the RF reference signal [22] are shown in Fig. 5. Single bunches with variable intensity were injected in four different RF voltages to obtain the dependence of energy loss on bunch length. In the measurements σ varied in the range 0.6–0.9 ns, so that impedances up to 1 GHz should be taken into account. Contributions to the normalized energy loss $\bar{U} = |U|/(N/10^{10})$ from different SPS impedances with frequencies less than 1 GHz (at the time of measurements) calculated for a Gaussian bunch are shown in Fig. 5 (left-hand side). As can be seen, in these measurements, the energy loss was dominated by the loss in the fundamental impedance of the 200 MHz RF system (shunt impedance $R_{\rm sh} \simeq 4.5\ \Omega$, quality factor $Q = 140$) and the MKE kickers. Contributions due to the main impedance of the 800 MHz cavities, total $R_{\rm sh} = 1.94$ MΩ and $Q = 300$, as well as the high-order mode of the 200 MHz RF system, with $f_{\rm r} = 629$ MHz, $Q = 500$ and $R_{\rm sh} = 604$ kΩ, are much smaller. The contribution to $\bar{U}$ from the resistive wall impedance is about 0.8 keV for a bunch with $\sigma = 0.6$ ns and decreases $\propto \sigma^{-3/2}$ for longer bunches. The measured and estimated total energy losses $\bar{U}$ are presented in Fig. 5 (right-hand side) as a function of bunch length.

The bunch-by-bunch measurements of the beam phase relative to the measured RF phase (probe) were used in the CERN LHC to estimate the energy loss of the proton bunches due to the electron cloud. Very high precision, within one degree, is required to measure the small shifts accurately. To obtain reliable results, the first 12 bunches were used as a reference to exclude other energy losses, from (short-range) impedances. The required accuracy was achieved after corrections for systematic errors and data post-processing [13]. Comparison with simulations gives a good estimate of the electron-cloud density [23]. This diagnostic tool has been available in the CERN Control Centre since 2015 and is used to evaluate results of beam scrubbing of the vacuum chamber, Fig. 6.

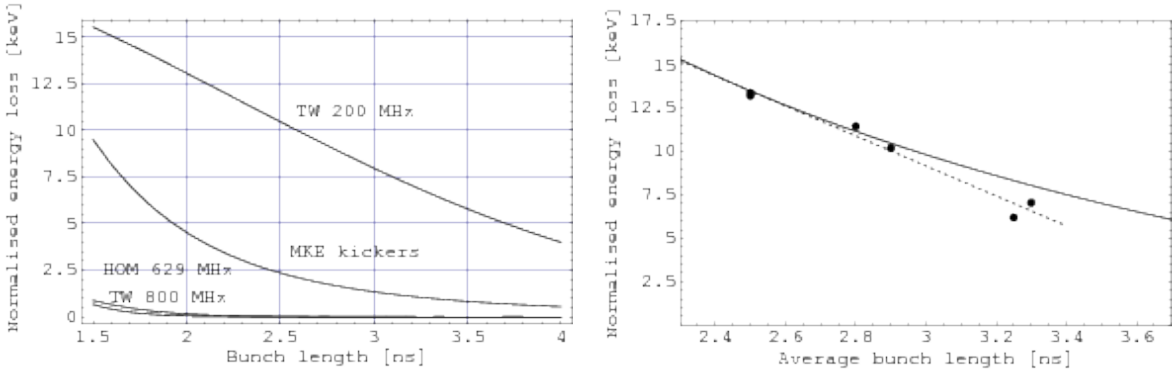


Fig. 5: Left-hand side: contribution to energy loss $\bar{U}$ (keV) from different SPS impedances as a function of 4σ bunch length. Right-hand side: normalized energy loss $\bar{U}$ (keV) calculated from the known SPS impedances (solid line) and measured from the phase shift (circles, measurement points; dashed line, linear fit) for different bunch lengths.

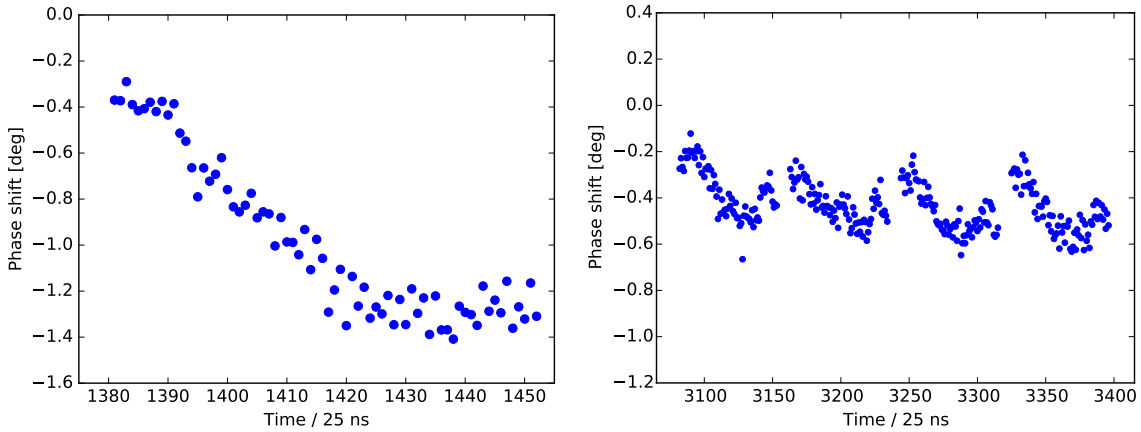


Fig. 6: Bunch-by-bunch synchronous phase shift for similar bunches spaced at 25 ns before (left-hand side) and after (right-hand side) scrubbing of the CERN LHC. Measurements at 450 GeV/c flat bottom [13].

3 Impedance measurements with unstable beam

Measurements with an unstable beam can help to identify the impedance source responsible for the instability. In particular, in some cases, which are considered next, the beam spectrum may contain important information about the frequency of the impedance responsible. Measurements of growth rates and instability thresholds are useful to estimate other parameters (R_{sh} and Q) of these impedances, especially with known resonant frequencies.

3.1 Unstable beam spectra with RF off: single-bunch case

Unstable spectra of long single bunches injected into an accelerator with RF off can be used to identify the longitudinal resonant impedances with large values of R/Q (for details see Refs. [24, 25]). The presence of different resonant impedances leads to line density modulation at the resonant frequencies, which can be detected. The synchrotron motion normally destroys this modulation (with the exception of a very fast instability with growth time significantly smaller then the synchrotron period). Too fast debunching will also modify this modulation. Thus, the parameters of bunches used for these measurements should satisfy certain requirements. For narrow-band resonant impedances with frequency bandwidth $\Delta\omega_{\mathrm{r}} \ll 1/\tau$, the width of the corresponding peak in the unstable spectrum is defined by $1/\tau$, so long bunches give better frequency resolution. Since injected bunches should become unstable, but not debunch too quickly, a

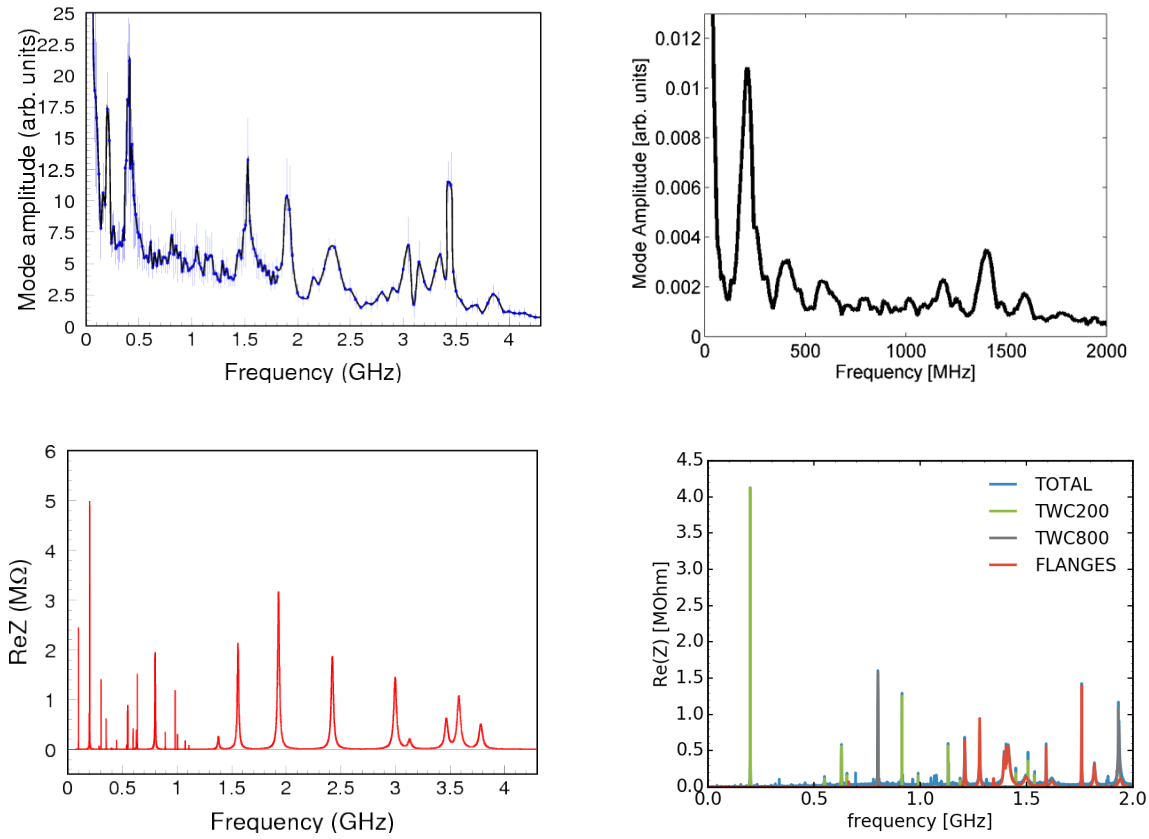


Fig. 7: Top: spectral distributions of long (25 ns) unstable bunches on the SPS flat bottom with RF off measured before the first SPS impedance reduction with $N_\mathrm{b} = 2 \times 10^{10}$ (left-hand side) and recently with $N_\mathrm{b} = 1 \times 10^{11}$. Bottom: corresponding longitudinal impedance models of the SPS; left-hand side: all resonant peaks above 1.4 GHz are due to intermagnet pumping ports shielded in 1999–2000 shutdown; right-hand side: impedance of vacuum flanges (blue) which will be reduced in the 2019–2020 shutdown.

small momentum spread also helps.

This method helped to find the impedances driving the microwave instability on the 26 GeV/c flat bottom in the CERN SPS almost 20 years ago, see Fig. 4 (left-hand side). Around 900 of the intermagnet pumping ports were shielded during the shutdown of 1999–2000, providing stability of the nominal LHC beam. Recently, this method (but with higher-intensity bunches) was used again in preparation of the SPS for its role as an injector of the high-luminosity LHC, where intensities that are twice as high as those achieved so far are required. The second impedance reduction programme, which includes shielding of $\sim$200 vacuum flanges, found responsible for a longitudinal instability during the SPS ramp (with minimum threshold on the flat top) [26, 27], is planned during the long shutdown in 2019–2020.

In the SPS measurements, bunch profiles were acquired during the first 100 ms after injection at regular time intervals. Each of these profiles was Fourier analysed and the maximum amplitude of the signal at each frequency was plotted (projection of mountain range). The spectral distributions measured with long bunches are shown in Fig. 7 (top), together with a corresponding impedance model at the time of measurement (bottom). Note the peak at 400 MHz, which disappeared after the first impedance reduction (shielding of pumping ports with resonant peaks in the range 1.5–3.5 GHz) and removal of the lepton equipment (SPS as LEP injector), although the particular impedance source was not identified.

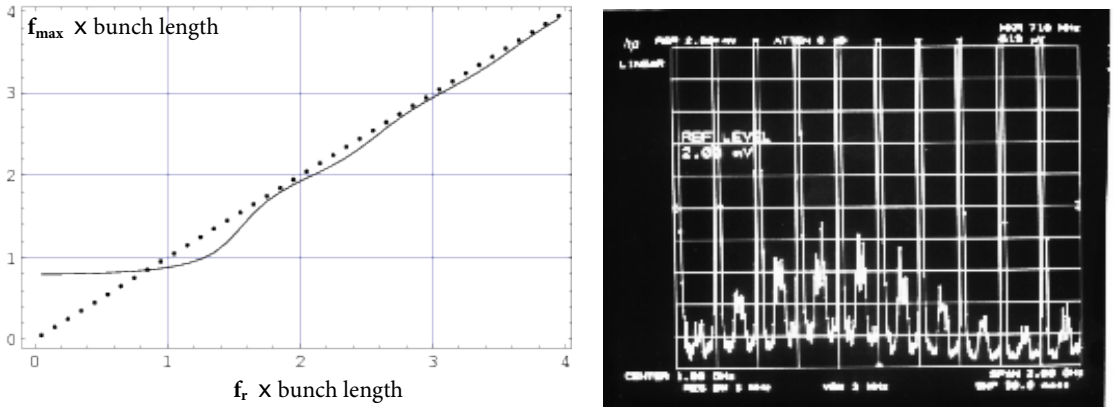


Fig. 8: Left-hand side: position of the maximum $f_{\max}\tau$ in the beam spectrum envelope as a function of $f_{\rm r}\tau$ for a parabolic line density and dipole mode $m = 1$. For higher multipoles m, the dependence is similar. Right-hand side: unstable beam spectrum from 0 to 2 GHz at the end of the fixed-target proton cycle in the CERN SPS for low-intensity beam (4×10^{12}) with 5 ns bunch spacing (strong lines at the 200 MHz harmonics).

3.2 Unstable beam spectra with RF on: multibunch case

Let us consider a coupled-bunch instability driven by some narrow-band impedance (e.g., high-order mode in the RF cavities) at resonant frequency $\omega_{\rm r} = 2\pi f_{\rm r} = \omega_0 p_{\rm r}$. The spectrum of the unstable multibunch beam has components at frequencies

$$\omega = (n + lM)\omega_0 + m\omega_{\rm s}, \quad l = 0, \pm 1, ..., \tag{27}$$

where M is the number of the equidistant bunches in the ring, $n = 0, 1, ..., M - 1$ is the coupled-bunch mode number describing the phase shift $2\pi n/M$ between adjacent bunches, and $m = 0, 1...$ is the multipole number related to the interbunch motion ($m = 1$, dipole; 2, quadrupole, and so on). On the spectrum analyser, the negative frequencies for a given value of n appear at $[(l + 1)M - n]\omega_0 - m\omega_{\rm s}$, so the spectrum is mirrored at $lM + n$ and $(l + 1)M - n$. With high-frequency resolution, one can use the fact that, above transition, internal synchrotron sidebands around revolution lines n and $M - n$ correspond to impedances at a higher frequency and external sidebands correspond to impedances at a lower frequency (the opposite is true below transition) [28]; however, the value of lM is still unknown. Measuring n for different (and, in particular, large) numbers of bunches M (with $M_1 \neq kM_2$) can also help to determine the resonant frequency $\omega_{\rm r}$. During development of the instability, the signals can also be directly observed from the coupling devices in the cavities.

Nevertheless, available data are often insufficient; then measurements of the envelope spectra can give additional information about the resonant frequency. The analysis of unstable beam spectra for different particle distribution functions has shown that if the measured position of the absolute maximum in the unstable bunch spectra $f_{\max} > 1/\tau$, then it practically coincides with the resonant frequency $f_{\rm r}$ [29]. The uncertainty does not exceed $\pm 0.2/\tau$. If the measured $f_{\max} < 1/\tau$, one can only say that $f_{\rm r} < 1.2/\tau$. The expected position of the maximum in the beam spectrum envelope as a function of resonant frequency is shown in Fig. 8 for a parabolic line density and $m = 1$ together with the spectrum envelope observed during development of a coupled-bunch instability in the CERN SPS for 4000 bunches spaced at 5 ns. From the measured value of $nf_0 = 113$ MHz, a few candidates were possible. Measurements were made at the end of the cycle for a bunch length of 2 ns. The spectrum has a maximum around 700 MHz so that the parameter $f_{\max}\tau = 1.4$. Then the driving impedance can be a high-order mode in the 200 MHz RF system with $f_{\rm r} = 912$ MHz.

Measurements of the instability growth rates and instability thresholds can be used to give an estimation of $R_{\rm sh}$ of the narrow-band impedance.

4 Discussion and summary

There are many other methods, for both transverse and longitudinal planes, which were not discussed here, see, for example, Refs. [20, 30].

Thanks to careful initial design, the coupling impedance of circular accelerators becomes smaller and therefore more elaborate methods are required to measure it with beams, even in proton machines (e.g., the LHC has $\mathrm{Im}Z/n = 0.09\ \Omega$). Numerical simulations of various collective effects also become more advanced and most of them can be also used for beam tests of impedance.

Measurements with stable beams are mainly used to verify existing impedance models. Conversely, measurements with unstable beams can provide important information about parameters of the dominant impedances driving the instabilities.

Acknowledgements

I am very grateful to all colleagues participating in beam impedance measurements in the CERN SPS and LHC, and in particular to T. Argyropoulos, T. Bohl, H. Damerau, A. Lasheen, T. Linnecar, J. Esteban Muller, and H. Timko.

References

[1] B.W. Zotter and S.A. Kheifets, *Impedances and Wakes in High-Energy Particle Accelerators* (World Scientific, Singapore, 1998). https://doi.org/10.1142/3068

[2] A. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (J. Wiley & Sons, New York, 1993).

[3] U. Niedermayer, Bench measurements and simulations of beam coupling impedance, these proceedings.

[4] R. Wanzerberg, Wakefields and impedances, these proceedings.

[5] J.M. Byrd *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **455** (2000) 271. https://doi.org/10.1016/S0168-9002(00)00504-0

[6] J.M. Byrd, in *Handbook of Accelerator Physics and Engineering*, 2nd ed., Eds. A. Chao *et al.* (World Scientific, Singapore, 2013), p.750.

[7] B. Zotter, Potential-well bunch lengthening, CERN Report SPS-81-14-(DI) (1981).

[8] J.L. Laclare, Bunched beam coherent instabilities, Proc. CERN Accelerator School (CAS) 1985, CERN 87-03 (1987) p. 264.

[9] J. Haissinski, Exact longitudinal equilibrium distribution in the presence of self-fields, *Nuovo Cimento B* **18** (1973) 72.

[10] D. Boussard, Schottky noise and beam transfer function diagnostics, Proc. CERN Accelerator School (CAS) 1985, CERN 87-03 (1987) p. 416.

[11] M. Blaskiewicz *et al.*, Longitudinal impedance of RHIC, Proc. IPAC2015, Richmond, VA, USA, 2015, p. 746.

[12] E. Shaposhnikova *et al.*, Longitudinal peak detected Schottky spectrum, Proc. HB2010, Morschach, Switzerland, 2010, p. 363, TUO1C04.

[13] J. Esteban Muller, Ph.D. thesis, EPFL, Lausanne, Switzerland, 2016.

[14] E. Shaposhnikova *et al.*, Reference measurements of the longitudinal impedance in the CERN SPS, Proc. PAC09, Vancouver, Canada, 2009.

[15] A. Lasheen *et al.*, Single bunch longitudinal instability in the CERN SPS, Proc. IPAC2016, TUPOR009, Busan, Korea, 2016, p.1670.

[16] H. Burkhardt *et al.*, Coherent beam oscillations and transverse impedance in the SPS, Proc. EPAC02, Paris, France, 2002.

[17] E. Shaposhnikova and T. Linnecar, Another method to measure the low-frequency machine impedance, Proc. EPAC96, Sitges, Spain, 1996.

[18] E. Shaposhnikova, Nonlinear bunch motion in an accelerator with reactive impedance, Proc. EPAC96, Sitges, Spain, 1996.

[19] M. A. Allen *et al.*, *IEEE Trans. Nucl. Sci.* **NS-22** (1975) 1838. https://doi.org/10.1109/TNS.1975.4328005

[20] A. Hofmann, Impedance measurements, computations and their interpretation, Proc. EPAC96, Sitges, Spain, 1996, p. 143.

[21] N. S. Sereno *et al.*, A potpourri of impedance measurements at the Advanced Photon Source Storage Ring, Proc. PAC97, 1997, p. 1700.

[22] E. Shaposhnikova *et al.*, Energy loss of a single bunch in the CERN SPS, Proc. EPAC04, Lucerne, Switzerland, 2004, p.1909.

[23] G. Rumolo, Electron cloud, these proceedings.

[24] T. Bohl *et al.*, *Phys. Rev. Lett.* **78** (1997) 3109. https://doi.org/10.1103/PhysRevLett.78.3109

[25] E. Shaposhnikova, Methods of observing the microwave instability above and below transition, Proc. PAC01, Chicago, USA, 2001, p. 385. https://doi.org/10.1109/pac.2001.987521

[26] T. Argyropoulos *et al.*, Identification of the SPS impedance at 1.4 GHz, Proc. IPAC2013, Shanghai, China, 2014, p. 1793.

[27] E. Shaposhnikova *et al.*, Identification of high-frequency resonant impedance in the CERN SPS, Proc. IPAC2014, Dresden, Germany, 2014, p. 1416.

[28] F. Sacherer and F. Pedersen, *IEEE Trans. Nucl. Sci.* **NS-24** (1977) 1393. https://doi.org/10.1109/TNS.1977.4328955

[29] E. Shaposhnikova, Analysis of coupled bunch instability spectra, Proc. Workshop on Instabilities of High Intensity Hadron Beams in Rings, Upton, New York, USA, **496** (1999) 256. https://doi.org/10.1063/1.1301890

[30] A. Chao *et al. Handbook of Accelerator Physics and Engineering*, 2nd ed. (World Scientific, Singapore, 2013). https://doi.org/10.1142/8543

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam Instabilities in Linear Machines: Space Charge Effects

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 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.

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.

https://doi.org/10.23730/CYRSP-2017-003.121

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, $p_{x,y} \ll p_z \approx p$. As a consequence, it is convenient 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.

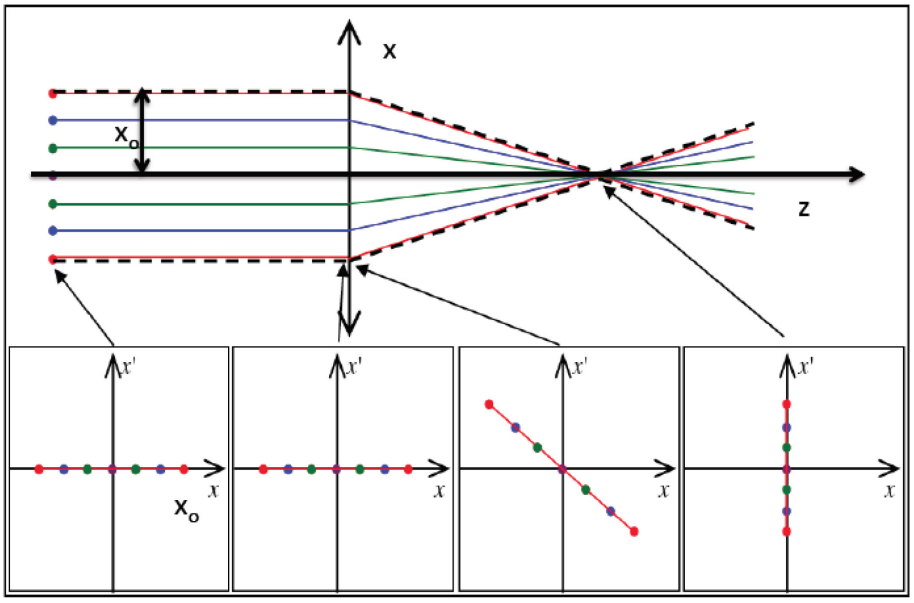


Fig. 1: Particle trajectories and phase space evolution of a laminar beam [7]

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, such 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.

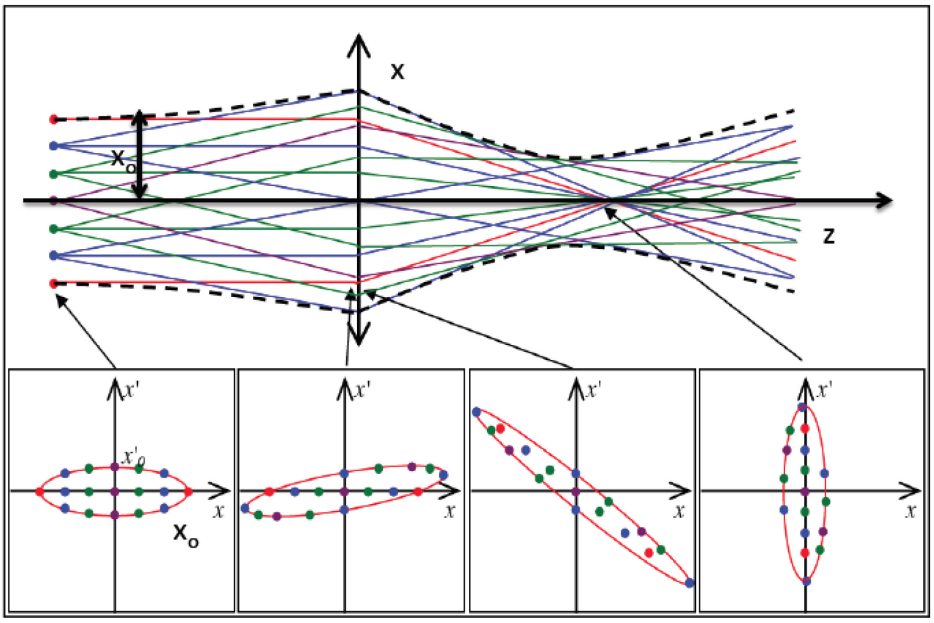


Fig. 2: Particle trajectories and phase space evolution of a non-laminar beam [7]

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 millimetres milliradians or, more often, in micrometres.

The ellipse equation is written as

$$\gamma_x x^2 + 2\alpha_x x x' + \beta_x x'^2 = \varepsilon_x \ , \tag{1}$$

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:

$$\begin{cases} X_{\max} = \sqrt{\beta_x \varepsilon_x} \\ \left(X_{\max}\right)' = -\alpha\sqrt{\dfrac{\varepsilon}{\beta}} \\ X'_{\max} = \sqrt{\gamma_x \varepsilon_x} \end{cases} . \tag{3}$$

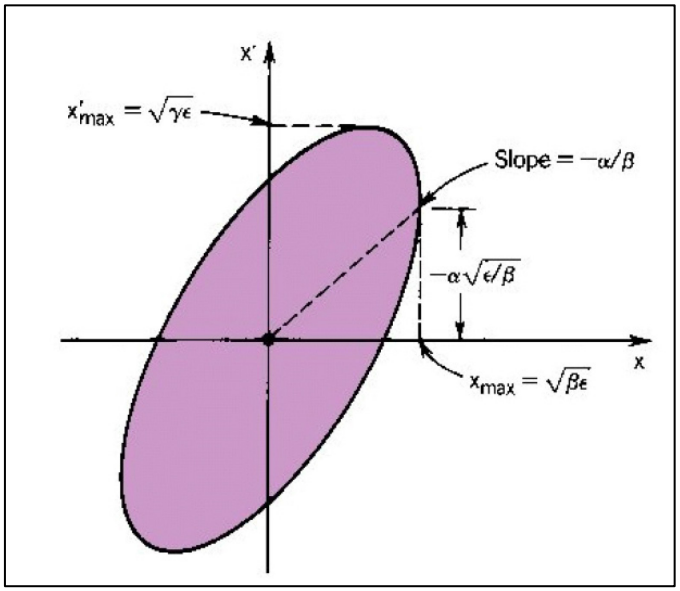


Fig. 3: Phase space distribution in a skewed elliptical boundary, showing the relationship of Twiss parameters to the ellipse geometry [6].

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.

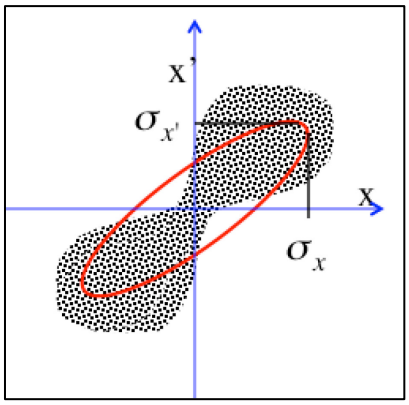


Fig. 4: Typical evolution of phase space distribution (black dots) under the effects of non-linear forces with the equivalent ellipse superimposed (red line).

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

$$\begin{cases} \sigma_x^2(z) = \langle x^2 \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \\ \sigma_{x'}^2(z) = \langle x'^2 \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x'^2 f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \end{cases} \tag{6}$$

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

$$\sigma_{xx'}(z) = \langle xx' \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xx' f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \ . \tag{7}$$

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.

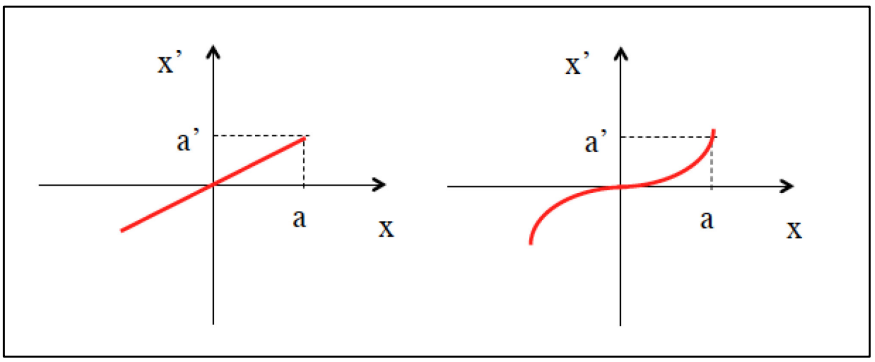
By substituting the Twiss parameter defined by Eqs. (5) and (8) into condition 2 we obtain [5]

$$\frac{\sigma_{x'}^2}{\varepsilon_{x,\mathrm{rms}}}\frac{\sigma_x^2}{\varepsilon_{x,\mathrm{rms}}} - \left(\frac{\sigma_{xx'}}{\varepsilon_{x,\mathrm{rms}}}\right) = 1 \, . \tag{9}$$

Reordering the terms in Eq. (8) we obtain the definition of *rms emittance* in terms of the second moments of the distribution:

$$\varepsilon_{\mathrm{rms}} = \sqrt{\sigma_x^2 \sigma_{x'}^2 - \sigma_{xx'}^2} = \sqrt{\left(\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2\right)} \ , \tag{10}$$

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

$$\varepsilon_{\text{rms}}^2 = C\sqrt{\left\langle x^2\right\rangle\left\langle x^{2n}\right\rangle - \left\langle x^{n+1}\right\rangle^2} \quad \begin{cases} n=1 \Rightarrow \ \varepsilon_{\text{rms}} = 0 \\ n>1 \Rightarrow \ \varepsilon_{\text{rms}} \neq 0 \end{cases} . \tag{11}$$

When $n = 1$, the line is straight and the rms emittance is $\varepsilon_{\text{rms}} = 0$. 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:

$$\varepsilon_{n,\text{rms}} = \frac{1}{m_0 c}\sqrt{\sigma_x^2\sigma_{p_x}^2 - \sigma_{xp_x}^{\ 2}} = \frac{1}{m_0 c}\sqrt{\left(\left\langle x^2\right\rangle\left\langle p_x^2\right\rangle - \left\langle xp_x\right\rangle^2\right)} = \sqrt{\left(\left\langle x^2\right\rangle\left\langle \left(\beta\gamma x'\right)^2\right\rangle - \left\langle x\beta\gamma x'\right\rangle^2\right)} . \tag{12}$$

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. Assuming a small energy spread within the beam, the normalized and un-normalized emittances can be related by the approximated relation $\langle\beta\gamma\rangle\varepsilon_{\text{rms}}$. 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

$$\varepsilon_{n,\text{rms}}^2 = \left\langle\beta^2\gamma^2\right\rangle\left\langle x^2\right\rangle\left\langle x'^2\right\rangle - \left\langle\beta\gamma\right\rangle^2\left\langle xx'\right\rangle^2 . \tag{13}$$

Consider now the definition of relative energy spread

$$\sigma_\gamma^2 = \frac{\left\langle\beta^2\gamma^2\right\rangle - \left\langle\beta\gamma\right\rangle^2}{\left\langle\beta\gamma\right\rangle^2},$$

which can be inserted into Eq. (13) to give

$$\varepsilon_{n,\text{rms}}^2 = \left\langle\beta^2\gamma^2\right\rangle\sigma_\gamma^2\left\langle x^2\right\rangle\left\langle x'^2\right\rangle + \left\langle\beta\gamma\right\rangle^2\left(\left\langle x^2\right\rangle\left\langle x'^2\right\rangle - \left\langle xx'\right\rangle^2\right) . \tag{14}$$

Assuming relativistic particles ($\beta = 1$), we get

$$\varepsilon_{n,\mathrm{rms}}^2 = \left\langle \gamma^2 \right\rangle \left(\sigma_\gamma^2 \sigma_x^2 \sigma_{x'}^2 + \varepsilon_{\mathrm{rms}}^2 \right). \tag{15}$$

If the first term in the parentheses is negligible, we find the conventional approximation of the normalized emittance, as $\left\langle \gamma \right\rangle \varepsilon_{\mathrm{rms}}$. 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 (σ_γ > 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 particle distribution during beam transport and acceleration. One can use the first collective variable defined in Eq. (6), the second moment of the distribution termed 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]:

$$\begin{aligned} \frac{\mathrm{d}\sigma_x}{\mathrm{d}z} &= \frac{\mathrm{d}}{\mathrm{d}z}\sqrt{\left\langle x^2 \right\rangle} = \frac{1}{2\sigma_x}\frac{\mathrm{d}}{\mathrm{d}z}\left\langle x^2 \right\rangle = \frac{1}{2\sigma_x} 2\left\langle xx' \right\rangle = \frac{\sigma_{xx'}}{\sigma_x} \\ \frac{\mathrm{d}^2\sigma_x}{\mathrm{d}z^2} &= \frac{\mathrm{d}}{\mathrm{d}z}\frac{\sigma_{xx'}}{\sigma_x} = \frac{1}{\sigma_x}\frac{\mathrm{d}\sigma_{xx'}}{\mathrm{d}z} - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{1}{\sigma_x}\left(\left\langle x'^2 \right\rangle + \left\langle xx' \right\rangle\right) - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{\sigma_{x'}^2 + \left\langle xx'' \right\rangle}{\sigma_x} - \frac{\sigma_{xx'}^2}{\sigma_x^3} \end{aligned} . \tag{16}$$

Rearranging the second derivative in Eq. (16), we obtain a second-order non-linear differential equation for the beam envelope evolution,

$$\sigma_x'' = \frac{\sigma_x^2 \sigma_{x'}^2 - \sigma_{xx'}^2}{\sigma_x^3} + \frac{\left\langle xx'' \right\rangle}{\sigma_x} , \tag{17}$$

or, in a more convenient form, using the rms emittance definition Eq. (10),

$$\sigma_x'' - \frac{1}{\sigma_x}\left\langle xx'' \right\rangle = \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_x^3} . \tag{18}$$

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's now consider, for example, the simple case with $\left\langle xx'' \right\rangle = 0$, 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:

$$\sigma(z) = \sqrt{\left(\sigma_0 + \sigma_0'\left(z - z_0\right)\right)^2 + \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_0^2}\left(z - z_0\right)^2} . \tag{20}$$

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

$$\sigma(z) = \sigma_0 \sqrt{1 + \left(\frac{z - z_0}{\beta_w} \right)^2} \quad . \tag{21}$$

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 $\beta_w = \sigma_0^2 / \varepsilon_{\text{rms}}$, as shown in Fig 6.

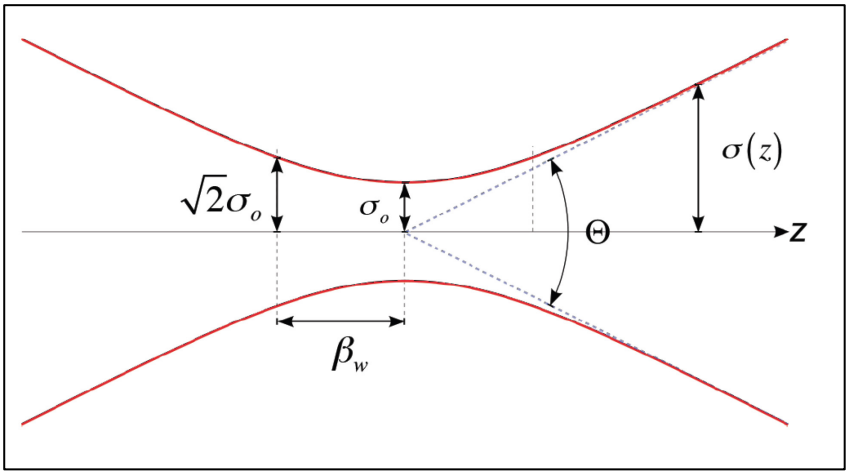


Fig. 6: Schematic representation of the beam envelope behaviour near the beam waist

At the waist, the relation $\varepsilon^2_{\text{rms}} = \sigma^2_{0,x}\sigma^2_{0,x'}$ 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

$$\varepsilon^2_{n,\text{rms}}(z) = \langle \gamma^2 \rangle \left(\sigma^2_\gamma \sigma^4_{x'} (z - z_0)^2 + \varepsilon^2_{\text{rms}} \right)$$

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 $w = 2\sigma_{\text{ph}}$ increases away from its minimum value at the waist w_0 with characteristic length $Z_R = \pi w_0^2 / \lambda$ (Rayleigh length) [4]. This analogy suggests that we can identify an effective emittance of a photon beam as $\varepsilon_{\text{ph}} = \lambda / 4\pi$.

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:

$$\frac{dp_x}{dt} = F_x \quad . \tag{22}$$

Under the paraxial approximation $p_x \ll p = \beta\gamma mc$, the transverse momentum p_x can be written as $p_x \sim = \sim p x' = \beta\gamma m_0 c x'$, so that

$$\frac{dp_x}{dt} = \frac{d}{dt}(px') = \beta c \frac{d}{dz}(px') = F_x \quad , \tag{23}$$

and the transverse acceleration results in

$$x'' = -\frac{p'}{p}x' + \frac{F_x}{\beta cp} \quad . \tag{24}$$

It follows that

$$\langle xx'' \rangle = -\frac{p'}{p}\langle xx' \rangle + \frac{\langle xF_x \rangle}{\beta cp} = \frac{p'}{p}\sigma_{xx'} + \frac{\langle xF_x \rangle}{\beta cp} \quad . \tag{25}$$

Inserting Eq. (25) into Eq. (18) and recalling Eq. (16), $\sigma_x' = \sigma_{xx'}/\sigma_x$, the complete rms envelope equation is

$$\sigma_x'' + \frac{p'}{p}\sigma_x' - \frac{1}{\sigma_x}\frac{\langle xF_x \rangle}{\beta cp} = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} \quad , \tag{26}$$

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

$$\langle xF_x \rangle = \mp k\langle x^2 \rangle = \mp k\sigma_x^2 \tag{27}$$

and the envelope equation becomes

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' \mp k_{\mathrm{ext}}^2\sigma_x = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} \quad , \tag{28}$$

where we have explicitly used the momentum definition $p = \gamma mc$ for a relativistic particle with $\beta \approx 1$ and defined the wavenumber

$$k_{\mathrm{ext}}^2 = \frac{k}{\gamma m_0 c^2} \quad .$$

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

$$k_{\mathrm{quad}} = \frac{qB_0'}{\gamma m_0 c} = k_{\mathrm{ext}}^2 \ .$$

In a solenoid the focusing strength is given by

$$k_{\mathrm{sol}} = \left(\frac{qB_0}{2\gamma m_0 c} \right)^2 = k_{\mathrm{ext}}^2 \ .$$

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

$$\sigma_x'' + k_{\mathrm{ext}}^2 \sigma_x = \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_x^3} \ . \tag{30}$$

By substituting $\sigma_x = \sqrt{\beta_x \varepsilon_{\mathrm{rms}}}$ into Eq. (30) one obtains an equation for the 'betatron function' $\beta_x(z)$ that is independent of the emittance term:

$$\beta_x'' + 2k_{\mathrm{ext}}^2 \beta_x = \frac{2}{\beta_x} + \frac{\beta_x'^2}{2\beta_x} \ . \tag{31}$$

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 describe the transport line characteristic by itself. 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

$$\beta_{\mathrm{eq}} = \frac{1}{k_{\mathrm{ext}}} = \frac{\lambda_\beta}{2\pi} \ ,$$

as can be easily verified. This result shows that the matched β_x function is simply the inverse of the focusing wave number or, equivalently, is proportional to the 'betatron wavelength' λ_β.

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*,

$$\lambda_{\mathrm{D}} = \sqrt{\frac{\varepsilon_0 \gamma^2 k_{\mathrm{B}} T_{\mathrm{b}}}{e^2 n}} \ , \tag{32}$$

where the transverse beam temperature T_{b} is defined as $k_{\mathrm{B}} T_{\mathrm{b}} = \gamma m_0 \left\langle v_{\perp}^2 \right\rangle$, and k_{B} is the Boltzmann constant. 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]

$$E_z(\zeta) = \frac{\hat{I} L}{2\pi\varepsilon_0 R^2 \beta c} h(\zeta) \ , \tag{33}$$

$$E_{\mathrm{r}}(r,\zeta) = \frac{\hat{I} r}{2\pi\varepsilon_0 R^2 \beta c} g(\zeta) \ . \tag{34}$$

The field form factor is described by the functions

$$h(\zeta) = \sqrt{A + (1-\zeta)^2} - \sqrt{A + \zeta^2 + (2\zeta - 1)} \ , \tag{35}$$

$$g(\zeta) = \frac{(1-\zeta)}{2\sqrt{A^2 + (1-\zeta)^2}} + \frac{\zeta}{2\sqrt{A^2 + \zeta}} \ , \tag{36}$$

where $\zeta = z/L$ is the normalized longitudinal coordinate along the bunch and $A = R/\gamma L$ is the beam aspect ratio. The field form factors account for the variation of the fields along the bunch. 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

$$B_g = \frac{\beta}{c} E_r \, .$$

Thus, the Lorentz force acting on each single particle is given by

$$F_r = e(E_r - \beta c B_g) = e(1-\beta^2) E_r = \frac{e E_r}{\gamma^2} \, . \tag{37}$$

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:

$$F_x = \frac{e\hat{I}x}{2\pi\gamma^2\varepsilon_0\sigma_x^2\beta c}g(\zeta) \ . \tag{38}$$

Then, computing the moment of the force, we need

$$x'' = \frac{F_x}{\beta cp} = \frac{eIx}{2\pi\varepsilon_0\gamma^3 m_0\beta^3 c^3\sigma_x^2} = \frac{k_{\mathrm{sc}}(\zeta)}{(\beta\gamma)^3\sigma_x^2} \tag{39}$$

where we have introduced the generalized beam perveance,

$$k_{\mathrm{sc}}(\zeta) = \frac{2\hat{I}}{I_{\mathrm{A}}}g(\zeta) \ , \tag{40}$$

normalized to the Alfven current $I_{\mathrm{A}} = 4\pi\varepsilon_0 m_0 c^3/e$ = 17 kA 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,

$$\langle xx'' \rangle = \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x^2}\langle x^2 \rangle = \frac{k_{\mathrm{sc}}}{\gamma^3} \ , \tag{41}$$

leading to the complete envelope equation

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' + k_{\mathrm{ext}}^2\sigma_x = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} + \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x} \ . \tag{42}$$

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 $x' \ll 1$. 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:

$$\rho = \frac{\hat{I}}{2I_{\mathrm{A}}\gamma}\frac{\sigma^2}{\varepsilon_n^2} \ . \tag{43}$$

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

$$\gamma_{\mathrm{tr}} = \frac{\hat{I}}{2I_{\mathrm{A}}}\frac{\sigma^2}{\varepsilon_n^2} \ . \tag{44}$$

For example, a beam with $\hat{I}$ = 100 A, ε_n = 1 μm, and σ = 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 kiloamperes 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 also dependent on the bunch coordinate 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:

$$\varepsilon_{rms,cor} = \sqrt{\left\langle \sigma_s^2 \right\rangle \left\langle \sigma_s'^2 \right\rangle - \left\langle \sigma_s \sigma_s' \right\rangle^2} \quad , \tag{45}$$

where the average is performed over the entire slice ensemble. In the simplest case of a two-slice model, the previous definition reduces to

$$\varepsilon_{rms,cor} = \left| \sigma_1 \sigma_2' - \sigma_2 \sigma_1' \right| \quad , \tag{46}$$

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

$$\varepsilon_{rms,cor} = \left\langle \gamma \right\rangle \sqrt{\varepsilon_{rms}^2 + \varepsilon_{rms,cor}^2} \quad . \tag{47}$$

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 ($k_{ext} = k_{sol}$), as illustrated in Fig. 7. To simplify our computations, we can neglect acceleration, as in the case of a simple beam transport line. The envelope equation for each slice, indicated as σ_s, reduces to

$$\sigma_s'' + k_{ext}^2 \sigma_s = \frac{k_{sc,s}}{\gamma^3 \sigma_s} \quad . \tag{48}$$

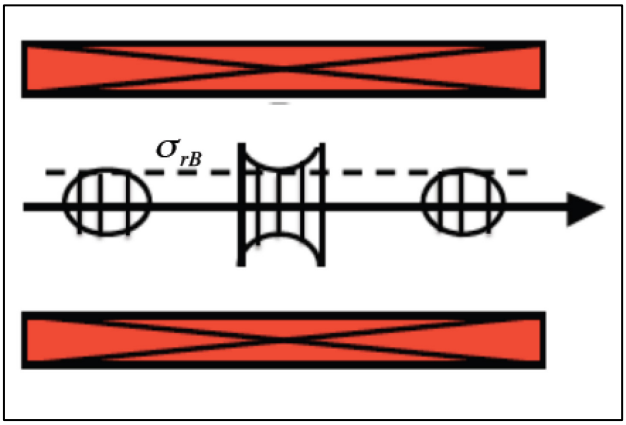


Fig. 7: Schematic representation of a nearly matched beam in a long solenoid. The dashed line represents the reference slice envelope fully matched to the Brillouin flow condition. The other slice envelopes are oscillating around the equilibrium solution.

A stationary solution, called the *Brillouin flow*, is given by

$$\sigma_{s,B} = \frac{1}{k_{ext}^2}\sqrt{\frac{\hat{I}g(\zeta)}{2\gamma^3 I_A}} \ , \tag{49}$$

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 cannot be achieved at the same time for all of the bunch slices. Assuming that there is a reference slice perfectly matched with an envelope $\sigma_{r,B}$, 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 small perturbation δ_s from the equilibrium 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 $\delta_0 = \sigma - \sigma_{sB}$ 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

$$\varepsilon_{\mathrm{rms,cor}} = \frac{1}{4} k_{\mathrm{sol}} \sigma_{\mathrm{rB}} \left| \frac{\Delta I}{\hat{I}} \delta_0 \sin\left(\sqrt{2} k_{\mathrm{ext}} z\right) \right| , \tag{55}$$

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 the expense of the longitudinal emittance.

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.

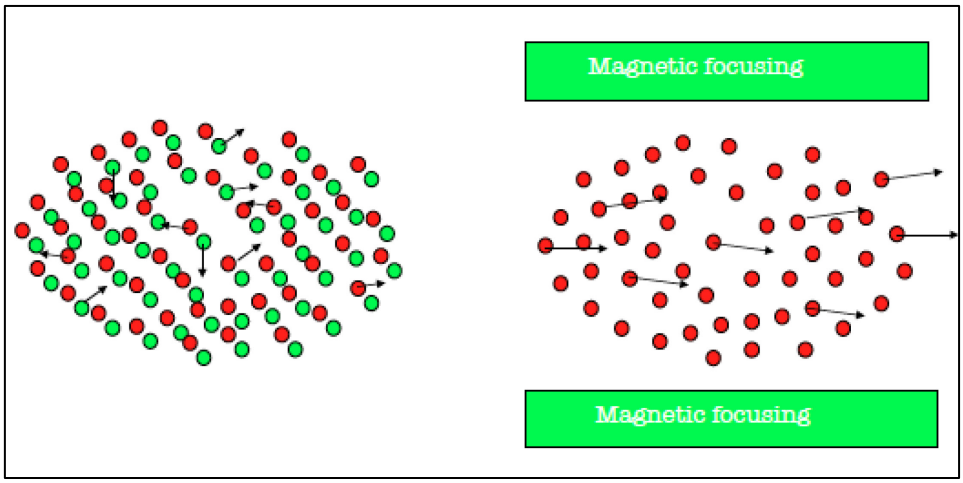


Fig. 8: The restoring force produced by the ions (green dots) in a plasma may cause electron (red dots) oscillations around the equilibrium distribution. In a similar way, the restoring force produced by a magnetic field may cause beam envelope oscillations around the matched envelope equilibrium.

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.

Acknowledgements

I wish to thank A. Cianchi, P. Muggli, J.B. Rosenzweig, A.R. Rossi 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.
http://dx.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). http://dx.doi.org/10.1002/9783527617623

[4] J.B. Rosenzweig, *Fundamentals of Beam Physics* (Oxford University Press, Oxford, 2003). http://dx.doi.org/10.1093/acprof:oso/9780198525547.001.0001

[5] T. Wangler, *Principles of RF Linear Accelerators* (Wiley, New York, 1998). http://dx.doi.org/10.1002/9783527618408

[6] S. Humphries, *Charged Particle Beams* (Wiley, New York, 2002).

[7] N. Pichoff, in Proceedings of the CERN Accelerator School: Small Accelerators, Zeegse, the Netherlands, 24 May – 2 June 2005, edited by D. Brandt, CERN-2006-012 (CERN, Geneva, 2006), pp. 145-177. http://dx/doi.org/10.5170/CERN-2006-012.145 http://dx.doi.org/10.5170/CERN-2006-012

[8] M. Migliorati *et al*., *Phys. Rev. ST Accel. Beams* **16** (2013) 011302. http://dx.doi.org/10.1103/PhysRevSTAB.16.011302

[9] M. Ferrario *et al*., *Nucl. Instrum. Methods Phys. Res. B* **309** (2013) 183. http://dx.doi.org/10.1016/j.nimb.2013.03.049

[10] A. Cianchi *et al*., *Nucl. Instrum. Methods Phys. Res. A* **720** (2013) 153. http://dx.doi.org/10.1016/j.nima.2012.12.012

[11] F.J. Sacherer, *IEEE Trans. Nucl. Sci*. **NS-18** (1971) 1105. http://dx.doi.org/10.1109/TNS.1971.4326293

[12] K. Floettmann, *Phys. Rev. ST Accel. Beams* **6** (2003) 034202. http://dx.doi.org/10.1103/PhysRevSTAB.6.034202

[13] M. Ferrario *et al*., *Int. J. Mod. Phys. A* **22** (2007) 4214. http://dx.doi.org/10.1142/S0217751X07037779

[14] J. Buon, in Proceedings of the CERN Accelerator School: 5th General Accelerator Physics Course, Jyvaskyla, Finland, 7 – 18 September 1992, edited by S. Turner, CERN-1994-001 (CERN, Geneva, 1994), pp. 89-116. http://dx/doi.org/10.5170/CERN-1994-001.89

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Introduction to Landau Damping*

W. Herr
CERN, Geneva, Switzerland

Abstract

The mechanism of Landau damping is observed in various systems from plasma oscillations to accelerators. Despite its widespread use, some confusion has been created, partly because of the different mechanisms producing the damping but also due to the mathematical subtleties treating the effects. In this article the origin of Landau damping is demonstrated for the damping of plasma oscillations. In the second part it is applied to the damping of coherent oscillations in particle accelerators. The physical origin, the mathematical treatment leading to the concept of stability diagrams and the applications are discussed.

Keywords

Landau Damping, oscillations, dispersion integral, stability

1 Introduction and history

Landau damping is referred to as the damping of a collective mode of oscillations in plasmas without collisions of charged particles. These Langmuir [1] oscillations consist of particles with long-range interactions and cannot be treated with a simple picture involving collisions between charged particles. The damping of such collisionless oscillations was predicted by Landau [2]. Landau deduced this effect from a mathematical study without reference to a physical explanation. Although correct, this derivation is not rigorous from the mathematical point of view and resulted in conceptual problems. Many publications and lectures have been devoted to this subject [3–5]. In particular, the search for stationary solutions led to severe problems. This was solved by Case [6] and Van Kampen [7] using normal-mode expansions. For many theorists Landau's result is counter-intuitive and the mathematical treatment in many publications led to some controversy and is still debated. This often makes it difficult to connect mathematical structures to reality. It took almost 20 years before dedicated experiments were carried out [8] to demonstrate successfully the reality of Landau damping. In practice, Landau damping plays a very significant role in plasma physics and can be applied to study and control the stability of charged beams in particle accelerators [9, 10].

It is the main purpose of this article to present a physical picture together with some basic mathematical derivations, without touching on some of the subtle problems related to this phenomenon.

The plan of this article is the following. First, the Landau damping in plasmas is derived and the physical picture behind the damping is shown. In the second part it is shown how the concepts are used to study the stability of particle beams. Some emphasis is put on the derivation of stability diagrams and beam transfer functions (BTFs) and their use to determine the stability. In an accelerator the decoherence or filamentation of an oscillating beam due to non-linear fields is often mistaken for Landau damping and significant confusion in the community of accelerator physicists still persists today. Nevertheless, it became a standard tool to stabilize particle beams of hadrons. In this article it is not possible to treat all the possible applications nor the mathematical subtleties and the references should be consulted.

*Previously published in Proceedings of the CAS-CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 18-29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), DOI: http://dx.doi.org/10.5170/CERN-2014-009

https://doi.org/10.23730/CYRSP-2017-003.137

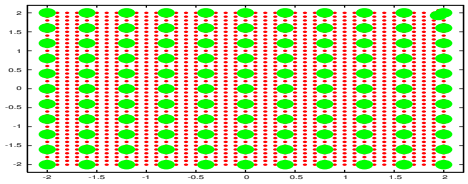


Fig. 1: Plasma without disturbance (schematic)

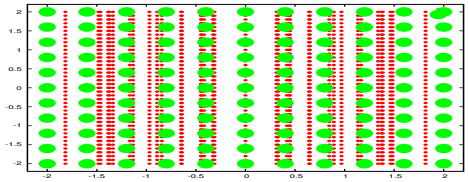


Fig. 2: Plasma with disturbance (schematic)

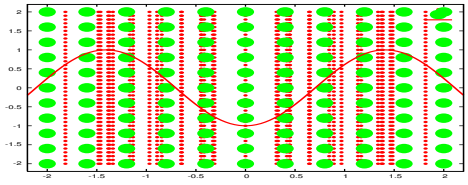


Fig. 3: Plasma with disturbance and restoring field (schematic)

2 Landau damping in plasmas

Initially, Landau damping was derived for the damping of oscillations in plasmas. In the next section, we shall follow the steps of this derivation in some detail.

2.1 Plasma oscillations

We consider an electrically quasi-neutral plasma in equilibrium, consisting of positively charged ions and negatively charged electrons (Fig. 1). For a small displacement of the electrons with respect to the ions (Fig. 2), the electric fields act on the electrons as a restoring force (Fig. 3). Due to the restoring force, standing density waves are possible with a fixed frequency [1] $\omega_\mathrm{p}^2 = \frac{ne^2}{m\epsilon_0}$, where n is the density of electrons, e the electric charge, m the effective mass of an electron and ϵ_0 the permittivity of free space. The individual motion of the electrons is neglected for this standing wave. In what follows, we allow for a random motion of the electrons with a velocity distribution for the equilibrium state and evaluate under which condition waves with a wave vector $\mathbf{k}$ and a frequency ω are possible.

The oscillating electrons produce fields (modes) of the form

$$E(x,t) = E_0 \sin(kx - \omega t) \tag{1}$$

or, rewritten,

$$E(x,t) = E_0 \mathrm{e}^{\mathrm{i}(kx-\omega t)}. \tag{2}$$

The corresponding wave (phase) velocity is then $v = \frac{\omega}{k}$. In Fig. 4, we show the electron distribution together with the produced field. The positive ions are omitted in this figure and are assumed to produce a stationary, uniform background field. This assumption is valid when we consider the ions to have

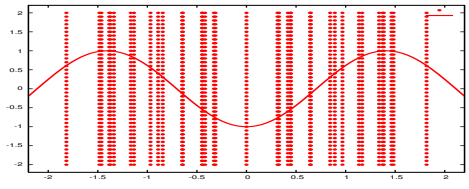


Fig. 4: Plasma with disturbance

infinite mass, which is a good approximation since the ion mass is much larger than the mass of the oscillating electrons.

2.2 Particle interaction with modes

The oscillating electrons now interact with the field they produce, i.e. individual particles interact with the field produced by all particles. This in turn changes the behaviour of the particles, which changes the field producing the forces. Furthermore, the particles may have different velocities. Therefore, a self-consistent treatment is necessary. If we allow ω to be complex ($\omega = \omega_r + \mathrm{i}\omega_i$), we separate the real and imaginary parts of the frequency ω and rewrite the fields:

$$E(x,t) = E_0 \mathrm{e}^{\mathrm{i}(kx-\omega t)} \Rightarrow E(x,t) = E_0 \mathrm{e}^{\mathrm{i}(kx-\omega_\mathrm{r} t)} \cdot \mathrm{e}^{\omega_\mathrm{i} t} \tag{3}$$

and we have a *damped* oscillation for $\omega_\mathrm{i} < 0$.

If we remember that particles may have different velocities, we can consider a much simplified picture as follows.

i) If more particles are moving more slowly than the wave:
   Net absorption of energy *from* the wave $\rightarrow$ wave is damped.
ii) If more particles are moving faster than the wave:
   Net absorption of energy *by* the wave $\rightarrow$ wave is antidamped.

We therefore have to assume that the *slope* of the particle distribution at the wave velocity is important. Although this picture is not completely correct, one can imagine a surfer on a wave in the sea, getting the energy from the wave (antidamping). Particles with very different velocities do not interact with the mode and cannot contribute to the damping or antidamping.

2.3 Liouville theorem

The basis for the self-consistent treatment of distribution functions is the Liouville theorem. It states that the phase space distribution function is constant along the trajectories of the system, i.e. the density of system points in the vicinity of a given system point travelling through phase space is constant with time, i.e. the density is always conserved. If the density distribution function is described by $\psi(\vec{q},\vec{p},t)$, then the probability to find the system in the phase space volume $\mathrm{d}q^n\mathrm{d}p^n$ is defined by $\psi(\vec{q},\vec{p})\mathrm{d}q^n\mathrm{d}p^n$ with $\int \psi(\vec{q},\vec{p},t)\mathrm{d}q^n\mathrm{d}p^n = N$. We have used canonical coordinates $q_i,\ i = 1,\ldots,n$ and momenta $p_i,\ i = 1,\ldots,n$ since it is defined for a Hamiltonian system. The evolution in time is described by the Liouville equation:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \sum_{i=1}^{n}\left(\frac{\partial\psi}{\partial q_i}\dot{q}_i + \frac{\partial\psi}{\partial p_i}\dot{p}_i\right) = 0. \tag{4}$$

If the distribution function is *stationary* (i.e. does not depend on q and t), then $\psi(\vec{q},\vec{p},t)$ becomes $\psi(\vec{p})$. Under the influence of non-linear fields the phase space distribution is changing. The shape of the

distribution function is distorted by the non-linearity, but the *local* phase space density is conserved. However, the *global* density may change, i.e. the (projected) measured beam size. The Liouville equation will lead us to the Boltzmann and Vlasov equations. We move again to Cartesian coordinates x and v.

2.4 Boltzmann and Vlasov equations

Time evolution of $\psi(\vec{x}, \vec{v}, t)$ is described by the Boltzmann equation:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \underbrace{\frac{\partial\psi}{\partial t}}_{\text{time change}} + \underbrace{\vec{v}\cdot\frac{\partial\psi}{\partial\vec{x}}}_{\text{space change}} + \underbrace{\frac{1}{m}\vec{F}(\vec{x},t)\cdot\frac{\partial\psi}{\partial\vec{v}}}_{v \text{ change, force } F} + \underbrace{\Omega(\psi)}_{\text{collision}} . \tag{5}$$

This equation contains a term which describes mutual collisions of charged particles in the distribution $\Omega(\psi)$. To study Landau damping, we ignore collisions and the collisionless Boltzmann equation becomes the Vlasov equation:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \vec{v}\cdot\frac{\partial\psi}{\partial\vec{x}} + \frac{1}{m}\vec{F}(\vec{x},t)\cdot\frac{\partial\psi}{\partial\vec{v}} = 0. \tag{6}$$

Here $\vec{F}(\vec{x}, t)$ is the force of the field (mode) on the particles.

Why is the Vlasov equation useful?

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \vec{v}\cdot\frac{\partial\psi}{\partial\vec{x}} + \frac{1}{m}\vec{F}(\vec{x},t)\cdot\frac{\partial\psi}{\partial\vec{v}} = 0. \tag{7}$$

Here $\vec{F}(\vec{x}, t)$ can be a force caused by impedances, beam–beam effects etc. From the solution one can determine whether a disturbance is growing (instability, negative imaginary part of frequency) or decaying (stability, positive imaginary part of frequency). It is a standard tool to study collective effects.

2.4.1 Vlasov equation for plasma oscillations

For our problem we need the force $\vec{F}$ (depending on field $\vec{E}$):

$$\vec{F} = e\cdot\vec{E} \tag{8}$$

and the field $\vec{E}$ (depending on potential Φ):

$$\vec{E} = -\nabla\Phi \tag{9}$$

for the potential Φ (depending on distribution ψ):

$$\Delta\Phi = -\frac{\rho}{\epsilon_0} = -\frac{e}{\epsilon_0}\int\psi\,\mathrm{d}v. \tag{10}$$

Therefore,

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \vec{v}\cdot\frac{\partial\psi}{\partial\vec{x}} + \frac{e}{m}\vec{E}(\vec{x},t)\cdot\frac{\partial\psi}{\partial\vec{v}} = 0. \tag{11}$$

We have obtained a set of coupled equations: the perturbation produces a field which acts back on the perturbation.

Can we find a solution? Assume a small *non-stationary* perturbation ψ_1 of the *stationary* distribution $\psi_0(\vec{v})$:

$$\psi(\vec{x},\vec{v},t) = \psi_0(\vec{v}) + \psi_1(\vec{x},\vec{v},t). \tag{12}$$

Then we get for the Vlasov equation:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi_1}{\partial t} + \vec{v}\cdot\frac{\partial\psi_1}{\partial\vec{x}} + \frac{e}{m}\vec{E}(\vec{x},t)\cdot\frac{\partial\psi_0}{\partial\vec{v}} = 0 \tag{13}$$

and

$$\Delta\Phi = -\frac{\rho}{\epsilon_0} = -\frac{e}{\epsilon_0}\int\psi_1\,\mathrm{d}v. \tag{14}$$

The density perturbation produces electric fields which act back and change the density perturbation, which therefore changes with time.

$$\psi_1(\vec{x},\vec{v},t) \Longrightarrow \vec{E}(\vec{x},t) \Longrightarrow \psi_1'(\vec{x},\vec{v},t) \Longrightarrow \cdots . \tag{15}$$

How can one treat this quantitatively and find a solution for ψ_1? We find two different approaches, one due to Vlasov and the other due to Landau.

2.4.2 Vlasov solution and dispersion relation

The Vlasov equation is a partial differential equation and we can try to apply standard techniques. Vlasov expanded the distribution and the potential as a double Fourier transform [11]:

$$\psi_1(\vec{x},\vec{v},t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\tilde{\psi_1}(k,\vec{v},\omega)\mathrm{e}^{\mathrm{i}(kx-\omega t)}\,\mathrm{d}k\,\mathrm{d}\omega, \tag{16}$$

$$\Phi(\vec{x},\vec{v},t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\tilde{\Phi}(k,\vec{v},\omega)\mathrm{e}^{\mathrm{i}(kx-\omega t)}\,\mathrm{d}k\,\mathrm{d}\omega \tag{17}$$

and applied these to the Vlasov equation. Since we assumed the field (mode) of the form $E(x,t) = E_0\mathrm{e}^{\mathrm{i}(kx-\omega t)}$, we obtain the condition (after some algebra)

$$1 + \frac{e^2}{\epsilon_0 mk}\int\frac{\partial\psi_0/\partial v}{(\omega-kv)}\,\mathrm{d}v = 0 \tag{18}$$

or, rewritten using the plasma frequency ω_p,

$$1 + \frac{\omega_\mathrm{p}^2}{k}\int\frac{\partial\psi_0/\partial v}{(\omega-kv)}\,\mathrm{d}v = 0 \tag{19}$$

or, slightly re-arranged for later use,

$$1 + \frac{\omega_\mathrm{p}^2}{k^2}\int\frac{\partial\psi_0/\partial v}{(\omega/k-v)}\,\mathrm{d}v = 0. \tag{20}$$

This is the dispersion relation for plasma waves, i.e. it relates the frequency ω with the wave vector $\mathbf{k}$. For this relation, waves with frequency ω and wave vector $\mathbf{k}$ are possible, answering the previous question.

Looking at this relation, we find the following properties.

i) It depends on the (velocity) distribution ψ.
ii) It depends on the slope of the distribution $\partial\psi_0/\partial v$.
iii) The effect is strongest for velocities close to the wave velocity, i.e. $v \approx \omega/k$.
iv) There seems to be a complication (singularity) at $v \equiv \omega/k$.

Can we deal with this singularity? Some proposals have been made in the past:

i) Vlasov's hand-waving argument [11]: in practice ω is never real (collisions).
ii) Optimistic argument [3,4]: $\partial\psi_0/\partial v = 0$, where $v \equiv \frac{\omega}{k}$.
iii) Alternative approach [6,7]:
   a) search for stationary solutions (normal-mode expansion);
   b) results in continuous versus discrete modes.
iv) Landau's argument [2]:
   a) Initial-value problem with perturbation $\psi_1(\vec{x},\vec{v},t)$ at $t = 0$ (time-dependent solution with complex ω).
   b) Solution: in time domain use Laplace transformation; in space domain use Fourier transformation.

2.4.3 Landau's solution and dispersion relation

Landau recognized the problem as an initial-value problem (in particular for the initial values $x = 0, v' = 0$) and accordingly used a different approach, i.e. he used a Fourier transform in the space domain:

$$\tilde{\psi_1}(k,\vec{v},t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \psi_1(\vec{x},\vec{v},t)\mathrm{e}^{\mathrm{i}(kx)}\,\mathrm{d}x, \tag{21}$$

$$\tilde{E}(k,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} E(\vec{x},t)\mathrm{e}^{\mathrm{i}(kx)}\,\mathrm{d}x \tag{22}$$

and a Laplace transform in the time domain:

$$\Psi_1(k,\vec{v},p) = \int_{0}^{+\infty} \tilde{\psi_1}(k,\vec{v},t)\mathrm{e}^{(-pt)}\,\mathrm{d}t \tag{23}$$

$$\mathcal{E}(k,p) = \int_{0}^{+\infty} \tilde{E}(k,t)\mathrm{e}^{(-pt)}\,\mathrm{d}t. \tag{24}$$

Inserted into the Vlasov equation and after some algebra (see references), this leads to the modified dispersion relation:

$$1 + \frac{e^2}{\epsilon_0 mk}\left[\mathrm{P.V.}\int \frac{\partial\psi_0/\partial v}{(\omega - kv)}\,\mathrm{d}v - \frac{\mathrm{i}\pi}{k}\left(\frac{\partial\psi_0}{\partial v}\right)_{v=\omega/k}\right] = 0 \tag{25}$$

or, rewritten using the plasma frequency ω_p,

$$1 + \frac{\omega_\mathrm{p}^2}{k}\left[\mathrm{P.V.}\int \frac{\partial\psi_0/\partial v}{(\omega - kv)}\,\mathrm{d}v - \frac{\mathrm{i}\pi}{k}\left(\frac{\partial\psi_0}{\partial v}\right)_{v=\omega/k}\right] = 0. \tag{26}$$

Here P.V. refers to 'Cauchy principal value'.

It must be noted that the second term appears only in Landau's treatment as a consequence of the initial conditions and is responsible for damping. The treatment by Vlasov failed to find this term and therefore did not lead to a damping of the plasma. Evaluating the term

$$-\frac{\mathrm{i}\pi}{k}\left(\frac{\partial\psi}{\partial v}\right)_{v=\omega/k}, \tag{27}$$

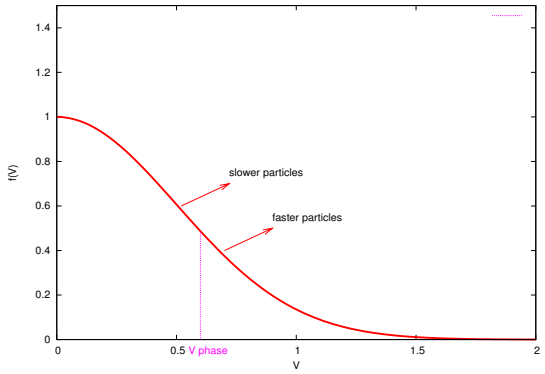


Fig. 5: Velocity distribution and damping conditions

We find damping of the oscillations if $\left(\frac{\partial\psi}{\partial v}\right)_{v=\omega/k} < 0$. This is the condition for Landau damping. How the dispersion relations for both cases are evaluated is demonstrated in Appendix A. The Maxwellian velocity distribution is used for this calculation.

2.5 Damping mechanism in plasmas

Based on these findings, we can give a simplified picture of this condition. In Fig. 5, we show a velocity distribution (e.g. a Maxwellian velocity distribution). As mentioned earlier, the damping depends on the number of particles below and above the phase velocity.

i) More 'slower' than 'faster' particles $\Longrightarrow$ damping.
ii) More 'faster' than 'slower' particles $\Longrightarrow$ antidamping.

This intuitive picture reflects the damping condition derived above.

3 Landau damping in accelerators

How to apply it to accelerators? Here we do not have a velocity distribution, but a frequency distribution $\rho(\omega)$ (in the transverse plane the tune). It should be mentioned here that $\rho(\omega)$ is the distribution of *external focusing frequencies*. Since we deal with a distribution, we can introduce a frequency spread of the distribution and call it $\Delta\omega$. The problem can be formally solved using the Vlasov equation, but the physical interpretation is very fuzzy (and still debated). Here we follow a different (more intuitive) treatment (following [5, 12, 13]). Although, if not taken with the necessary care, it can lead to a wrong physical picture, it delivers very useful concepts and tools for the design and operation of an accelerator. We consider now the following issues.

i) Beam response to excitation.
ii) BTF and stability diagrams.
iii) Phase mixing.
iv) Conditions and tools for stabilization and the related problems.

This treatment will lead again to a dispersion relation. However, the stability of a beam is in general not studied by directly solving this equation, but by introducing the concept of stability diagrams, which allow us more directly to evaluate the stability of a beam during the design or operation of an accelerator.

3.1 Beam response to excitation

How does a beam respond to an external excitation?

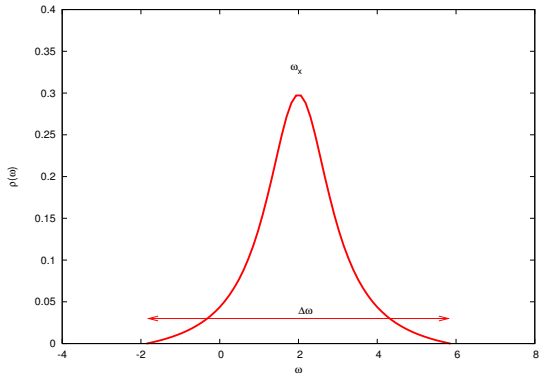


Fig. 6: Frequency distribution in a beam and frequency spread (very schematic)

To study the dynamics in accelerators, we replace the velocity v by $\dot{x}$ to be consistent with the standard literature. Consider a harmonic, linear oscillator with frequency ω driven by an external sinusoidal force $f(t)$ with frequency Ω. The equation of motion is

$$\ddot{x} + \omega^2 x = A\cos\Omega t = f(t). \tag{28}$$

For initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$, the solution is

$$x(t) = -\frac{A}{(\Omega^2 - \omega^2)}(\cos\Omega t \underbrace{-\cos\omega t}_{x(0)=0,\dot{x}(0)=0}). \tag{29}$$

The term $-\cos\omega t$ is needed to satisfy the initial conditions. Its importance will become clear later. The beam consists of an ensemble of oscillators with different frequencies ω with a distribution $\rho(\omega)$ and a spread $\Delta\omega$, schematically shown in Fig. 6.

Recall that for a transverse (betatron) motion ω_x is the tune. The number of particles per frequency band is $\rho(\omega) = \frac{1}{N}\mathrm{d}N/\mathrm{d}\omega$ with $\int_{-\infty}^{\infty}\rho(\omega)\mathrm{d}\omega = 1$. The *average* beam response (centre of mass) is then

$$\langle x(t)\rangle = \int_{-\infty}^{\infty} x(t)\rho(\omega)\,\mathrm{d}\omega, \tag{30}$$

and re-written using (29)

$$\langle x(t)\rangle = -\int_{-\infty}^{\infty}\left[\frac{A}{(\Omega^2 - \omega^2)}(\cos\Omega t - \cos\omega t)\right]\rho(\omega)\,\mathrm{d}\omega. \tag{31}$$

For a narrow beam spectrum around a frequency ω_x (tune) and the driving force near this frequency ($\Omega \approx \omega_x$),

$$\langle x(t)\rangle = -\frac{A}{2\omega_x}\int_{-\infty}^{\infty}\left[\frac{1}{(\Omega - \omega)}(\cos\Omega t - \cos\omega t)\right]\rho(\omega)\mathrm{d}\omega. \tag{32}$$

For the further evaluation, we transform variables from ω to $u = \omega - \Omega$ (see Ref. [12]) and assume that Ω is complex: $\Omega = \Omega_\mathrm{r} + \mathrm{i}\Omega_\mathrm{i}$ where

$$\langle x(t)\rangle = -\frac{A}{2\omega_x}\cos(\Omega t)\int_{-\infty}^{\infty}\mathrm{d}u\rho(u+\Omega)\frac{1-\cos(ut)}{u} \tag{33}$$

$$+\frac{A}{2\omega_x}\sin(\Omega t)\int_{-\infty}^{\infty}\mathrm{d}u\rho(u+\Omega)\frac{\sin(ut)}{u}. \tag{34}$$

This avoids singularities for $u = 0$.

We are interested in long-term behaviour, i.e. $t \to \infty$, so we use

$$\lim_{t\to\infty} \frac{\sin(ut)}{u} = \pi\delta(u), \tag{35}$$

$$\lim_{t\to\infty} \frac{1-\cos(ut)}{u} = \text{P.V.}\left(\frac{1}{u}\right), \tag{36}$$

$$\langle x(t)\rangle = \frac{A}{2\omega_x}\left[\pi\rho(\Omega)\sin(\Omega t) + \cos(\Omega t)\text{P.V.}\int_{-\infty}^{\infty} \mathrm{d}\omega \frac{\rho(\omega)}{(\omega-\Omega)}\right]. \tag{37}$$

This response or BTF has two parts as follows.

i) Resistive part, the first term in the expression is in phase with the excitation: absorbs energy from oscillation $\Longrightarrow$ damping (would not be there without the term $-\cos\omega t$ in (29)).
ii) Reactive part, the second term in the expression is out of phase with the excitation.

Assuming Ω is complex, we integrate around the pole and obtain a P.V. and a 'residuum' (Sokhotski–Plemelj formula), a standard technique in complex analysis.

We can discuss this expression and find

i) the 'damping' part only appeared because of the initial conditions;
ii) with other initial conditions, we get additional terms in the beam response;
iii) that is, for $x(0) \neq 0$ and $\dot{x}(0) \neq 0$ we may add

$$x(0)\int \mathrm{d}\omega\, \rho(\omega)\cos(\omega t) + \dot{x}(0)\int \mathrm{d}\omega\, \rho(\omega)\frac{\sin(\omega t)}{\omega}. \tag{38}$$

With these initial conditions, we do not obtain Landau damping and the dynamics is very different. We have again:

i) oscillation of particles with different frequencies (tunes);
ii) now with different initial conditions, $x(0) \neq 0$ and $\dot{x}(0) = 0$ or $x(0) = 0$ and $\dot{x}(0) \neq 0$;
iii) again we average over particles to obtain the centre of mass motion.

We obtain Figs. 8–10. Figure 8 shows the oscillation of individual particles where all particles have the initial conditions $x(0) = 0$ and $\dot{x}(0) \neq 0$. In Fig. 8, we plot again Fig. 7 but add the average beam response. We observe that although the individual particles continue their oscillations, the average is 'damped' to zero. The equivalent for the initial conditions $x(0) \neq 0$ and $\dot{x}(0) = 0$ is shown in Fig. 9. With a frequency (tune) spread the average motion, which can be detected by a position monitor, damps out. However, this is *not* Landau damping, rather filamentation or decoherence. Contrary to Landau damping, it leads to emittance growth.

3.2 Interpretation of Landau damping

Compared with the previous case of Landau damping in a plasma, the interpretation of the mechanism is quite different. The initial conditions of the beams are important and also the spread of external focusing frequencies. For the initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$, the beam is quiet and a spread of frequencies $\rho(\omega)$ is present. When an excitation is applied:

i) particles cannot organize into collective response (phase mixing);

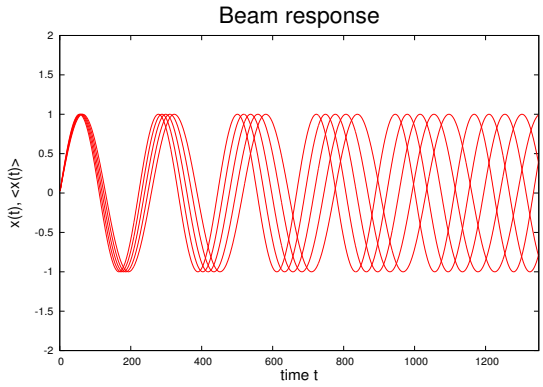


Fig. 7: Motion of particles with frequency spread. Initial conditions: $x(0) = 0$ and $\dot{x}(0) \neq 0$

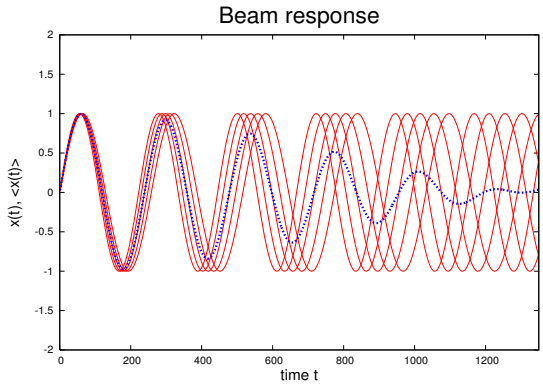


Fig. 8: Motion of particles with frequency spread and total beam response. Initial conditions: $x(0) = 0$ and $\dot{x}(0) \neq 0$.

ii) average response is zero;

iii) the beam is kept stable, i.e. stabilized.

In the case of accelerators the mechanism is therefore not a dissipative damping but a mechanism for stabilization. Landau damping should be considered as an 'absence of instability'. In the next step this is discussed quantitatively and the dispersion relations are derived.

3.3 Dispersion relations

We rewrite (simplify) the response in complex notation:

$$\langle x(t)\rangle = \frac{A}{2\omega_x}\left[\pi\rho(\Omega)\sin(\Omega t) + \cos(\Omega t)\mathrm{P.V.}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \frac{\rho(\omega)}{(\omega-\Omega)}\right] \tag{39}$$

becomes

$$\langle x(t)\rangle = \frac{A}{2\omega_x}\mathrm{e}^{-\mathrm{i}\Omega t}\left[\mathrm{P.V.}\int \mathrm{d}\omega\, \frac{\rho(\omega)}{(\omega-\Omega)} + \mathrm{i}\pi\rho(\Omega)\right]. \tag{40}$$

The first part describes an oscillation with complex frequency Ω.

Since we know that the collective motion is described as $\mathrm{e}^{(-\mathrm{i}\Omega t)}$, an oscillating solution Ω must fulfil the relation

$$1 + \frac{1}{2\omega_x}\left[\mathrm{P.V.}\int \mathrm{d}\omega \frac{\rho(\omega)}{(\omega-\Omega)} + \mathrm{i}\pi\rho(\Omega)\right] = 0. \tag{41}$$

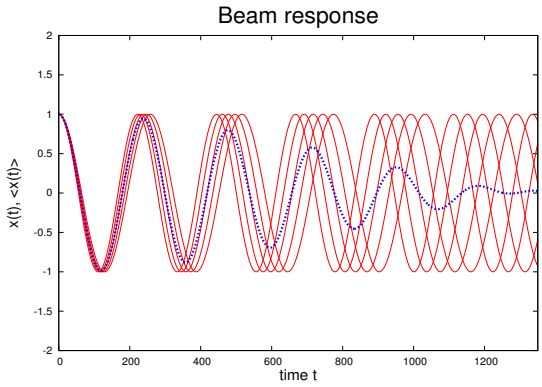


Fig. 9: Motion of particles with frequency spread and total beam response. Initial conditions $x(0) \neq 0$ and $\dot{x}(0) = 0$.

This is again a dispersion relation, i.e. the condition for the oscillating solution. What do we do with that? We could look where $\Omega_\mathrm{i} < 0$ provides damping. Note that no contribution to damping is possible when Ω is outside the spectrum. In the following sections, we introduce BTFs and stability diagrams which allow us to determine the stability of a beam.

3.4 Normalized parametrization and BTFs

We can simplify the calculations by moving to normalized parametrization. Following Chao's proposal [12], in the expression

$$\langle x(t)\rangle = \frac{A}{2\omega_x}\mathrm{e}^{-\mathrm{i}\Omega t}\left[\mathrm{P.V.}\int \mathrm{d}\omega \frac{\rho(\omega)}{(\omega-\Omega)} + \mathrm{i}\pi\rho(\Omega)\right] \tag{42}$$

we use again u, but normalized to frequency spread $\Delta\omega$. We have

$$u = (\omega_x - \Omega) \Longrightarrow u = \frac{(\omega_x - \Omega)}{\Delta\omega} \tag{43}$$

and introduce two functions $f(u)$ and $g(u)$:

$$f(u) = \Delta\omega \mathrm{P.V.}\int \mathrm{d}\omega\, \frac{\rho(\omega)}{\omega-\Omega}, \tag{44}$$

$$g(u) = \pi\Delta\omega\rho(\omega_x - u\Delta\omega) = \pi\Delta\omega\rho(\Omega). \tag{45}$$

The response with the driving force discussed above is now

$$\langle x(t)\rangle = \frac{A}{2\omega_x\Delta\omega}\mathrm{e}^{-\mathrm{i}\Omega t}\left[f(u) + \mathrm{i}\cdot g(u)\right], \tag{46}$$

where $\Delta\omega$ is the frequency spread of the distribution. The expression $f(u)+\mathrm{i}\cdot g(u)$ is the BTF. With this, it is easier to evaluate the different cases and examples. For important distributions $\rho(\omega)$ the analytical functions $f(u)$ and $g(u)$ exist (see e.g. [14]) and will lead us to stability diagrams.

3.5 Example: response in the presence of wake fields

The driving force comes from the displacement of the beam as a whole, i.e. $\langle x\rangle = X_0$, for example driven by a wake field or impedance (see Fig. 10). The equation of motion for a particle is then something like

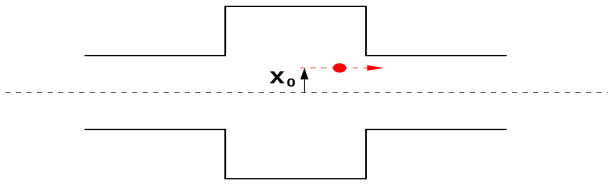


Fig. 10: Bunch with offset in a cavity-like object

$$\ddot{x} + \omega^2 x = f(t) = K \cdot \langle x \rangle, \tag{47}$$

where K is a 'coupling coefficient'. The coupling coefficient K depends on the nature of the wake field.

i) If K is purely real: the force is in phase with the displacement, e.g. image space charge in perfect conductor.
ii) For purely imaginary K: the force is in phase with the velocity and out of phase with the displacement.
iii) In practice, we have both and we can write

$$K = 2\omega_x(U - \mathrm{i}V). \tag{48}$$

Interpretation:

a) a beam travelling off centre through an impedance induces transverse fields;
b) transverse fields act back on all particles in the beam, via

$$\ddot{x} + \omega^2 x = f(t) = K \cdot \langle x \rangle; \tag{49}$$

c) if the beam moves as a whole (in phase, collectively), this can grow for $V > 0$;
d) the coherent frequency Ω becomes complex and is shifted by $(\Omega - \omega_x)$.

3.5.1 Beam without frequency spread

For a beam *without frequency spread* (i.e. $\rho(\omega) = \delta(\omega - \omega_x)$), we can easily sum over all particles and for the centre of mass motion $\langle x \rangle$ we get

$$\ddot{\langle x \rangle} + \Omega^2 \langle x \rangle = f(t) = -2\omega_x(U - \mathrm{i}V) \cdot \langle x \rangle. \tag{50}$$

For the original coherent motion with frequency Ω, this means that:

i) in-phase component U changes the frequency;
ii) out-of-phase component V creates growth ($V > 0$) or damping ($V < 0$).

For any $V > 0$, the beam is unstable (even if very small).

3.5.2 Beam with frequency spread

What happens for a beam *with a frequency spread*?

The response (and therefore the driving force) was

$$\langle x(t) \rangle = \frac{A}{2\omega_x \Delta\omega} \mathrm{e}^{-\mathrm{i}\Omega t} \left[f(u) + \mathrm{i} \cdot g(u) \right]. \tag{51}$$

The (complex) frequency Ω is now determined by the condition

$$-\frac{(\Omega-\omega_x)}{\Delta\omega}=\frac{1}{(f(u)+\mathrm{i}g(u))}. \tag{52}$$

All information about the stability is contained in this relation.

i) The (complex) frequency difference $(\Omega-\omega_x)$ contains the effects of impedance, intensity, γ etc. (see the article by G. Rumolo [15]).
ii) The right-hand side contains information about the frequency spectrum (see definitions for $f(u)$ and $g(u)$ in (44) and (45)).

Without Landau damping (no frequency spread):

i) if $\mathrm{Im}(\Omega-\omega_x)<0$, the beam is stable;
ii) if $\mathrm{Im}(\Omega-\omega_x)>0$, the beam is unstable (growth rate τ^{-1}).

With a frequency spread, we have a condition for stability (Landau damping)

$$-\frac{(\Omega-\omega_x)}{\Delta\omega}=\frac{1}{(f(u)+\mathrm{i}g(u))}. \tag{53}$$

How do we proceed to find the limits?

We could try to find the complex Ω at the edge of stability ($\tau^{-1}=0$). In the next section, we develop a more powerful tool to assess the stability of a dynamic system.

4 Stability diagrams

To study the stability of a particle beam, it is necessary to develop easy to use tools to relate the condition for stability with the complex tune shift due to, e.g., impedances. We consider the right-hand side first and call it D_1. Both D_1 and the tune shift are complex and should be analysed in the complex plane.

Using the (real) parameter u in

$$D_1=\frac{1}{(f(u)+\mathrm{i}g(u))}, \tag{54}$$

if we know $f(u)$ and $g(u)$ we can scan u from $-\infty$ to $+\infty$ and plot the real and imaginary parts of D_1 in a complex plane.

Why is this formulation interesting? The expression

$$(f(u)+\mathrm{i}g(u)) \tag{55}$$

is actually the BTF, i.e. it can be measured. With its knowledge (more precisely: its inverse), we have conditions on $(\Omega-\omega_x)$ for stability and the limits for intensities and impedances.

4.1 Examples for bunched beams

As an example, we use a rectangular distribution function for the frequencies (tunes), i.e.

$$\rho(\omega)=\begin{cases}\frac{1}{2\Delta\omega} & \mathrm{for}|\omega-\omega_x|\leq\Delta\omega,\\ 0 & \mathrm{otherwise.}\end{cases} \tag{56}$$

We now follow some standard steps.

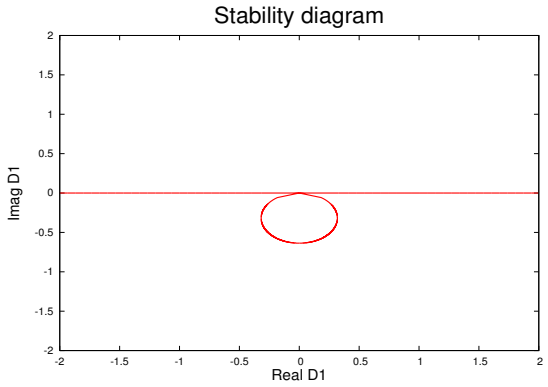


Fig. 11: Stability diagram for rectangular frequency distribution

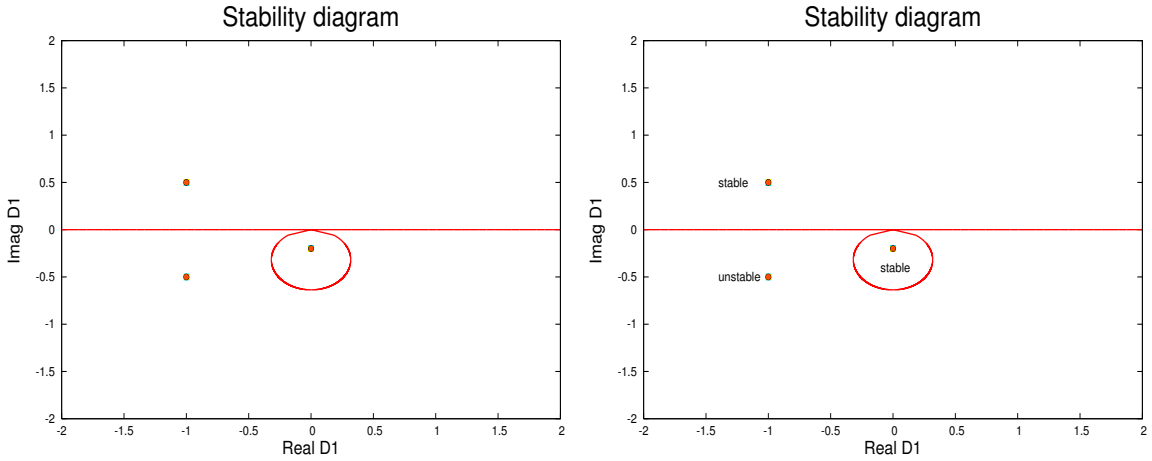


Fig. 12: Stability diagrams for rectangular distribution together with examples for complex tune shifts (left); stable and unstable points are indicated (right).

Step 1. Compute $f(u)$ and $g(u)$ (or look it up, e.g. [14]): we get for the rectangular distribution function

$$f(u) = \frac{1}{2}\ln\left|\frac{u+1}{u-1}\right|, \quad g(u) = \frac{\pi}{2}\cdot H(1-|u|). \tag{57}$$

Step 2. Plot the real and imaginary parts of D_1.

The result of this procedure is shown in Fig. 11 and the interpretation is as follows.

- We plot $\mathrm{Re}(D_1)$ versus $\mathrm{Im}(D_1)$ for a rectangular distribution $\rho(\omega)$.
- This is a stability boundary diagram.
- It separates stable from unstable regions (stability limit).

Now we plot the complex expression of $-\frac{(\Omega-\omega_x)}{\Delta\omega}$ in the same plane as a point (this point depends on impedances, intensities, etc.). In Fig. 12, we show the same stability diagram together with examples of complex tune shifts. The stable and unstable points in the stability diagram are indicated in the right-hand side of Fig. 12.

We can use other types of frequency distributions, for example a bi-Lorentz distribution $\rho(\omega)$. We follow the same procedure as above and the result is shown in Fig. 13. It can be shown that in all cases half of the complex plane is stable without Landau damping, as indicated in Fig. 13.

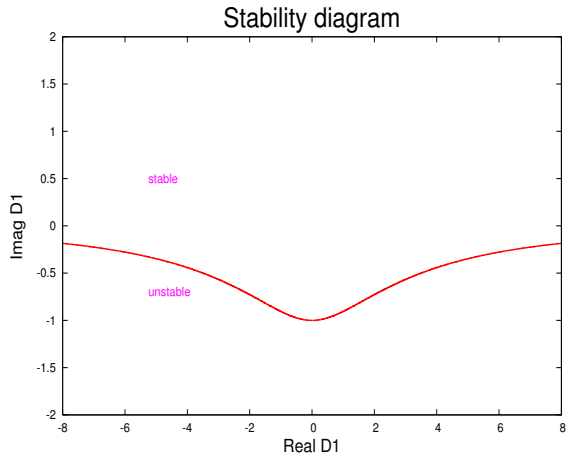


Fig. 13: Stability diagram for bi-Lorentz frequency distribution. Stable and unstable regions are indicated

4.2 Examples for unbunched beams

A similar treatment can be applied to unbunched beams, although some care has to be taken, in particular in the case of longitudinal stability.

4.2.1 Transverse unbunched beams

The technique applies directly. Frequency (tune) spread is from:

i) change of revolution frequency with energy spread (momentum compaction);
ii) change of betatron frequency with energy spread (chromaticity).

The oscillation depends on the mode number n (number of oscillations around the circumference C):

$$\propto \exp(-\mathrm{i}\Omega t + \mathrm{i}n(s/C)) \tag{58}$$

and the variable u should be written

$$u = (\omega_x + n \cdot \omega_0 - \Omega)/\Delta\omega. \tag{59}$$

The rest has the same treatment. Transverse collective modes in an unbunched beam for mode numbers 4 and 6 are shown in Fig. 14.

4.2.2 Longitudinal unbunched beams

What about longitudinal instability of unbunched beams? This is a special case since there is no external focusing, therefore also no spread $\Delta\omega$ of focusing frequencies. However, we have a spread in revolution frequency, which is directly related to energy, and energy excitations directly affect the frequency spread:

$$\frac{\Delta\omega_{\mathrm{rev}}}{\omega_0} = -\frac{\eta}{\beta^2}\frac{\Delta E}{E_0}. \tag{60}$$

The frequency distribution is described by

$$\rho(\omega_{\mathrm{rev}}) \quad \text{and} \quad \Delta\omega_{\mathrm{rev}}. \tag{61}$$

Since there is no external focusing ($\omega_x = 0$), we have to modify the definition of our parameters:

$$u = \frac{(\omega_x + n \cdot \omega_0 - \Omega)}{\Delta\omega} \Longrightarrow u = \frac{(n \cdot \omega_0 - \Omega)}{n \cdot \Delta\omega} \tag{62}$$

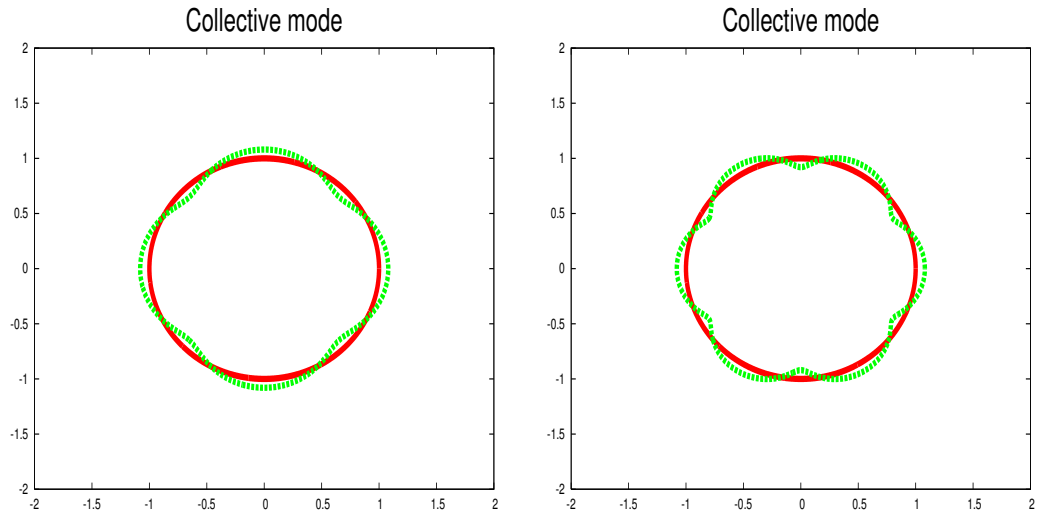


Fig. 14: Transverse collective mode with mode index $n = 4$ and $n = 6$

and introduce two new functions $F(u)$ and $G(u)$: the variable n is the mode number.

$$F(u) = n \cdot \Delta\omega^2 \mathrm{P.V.} \int \mathrm{d}\omega_0 \, \frac{\rho'(\omega_0)}{n \cdot \omega_0 - \Omega}, \tag{63}$$

$$G(u) = \pi \Delta\omega^2 \rho'(\Omega/n) \tag{64}$$

to obtain

$$-\frac{(\Omega - n \cdot \omega_0)^2}{n^2 \Delta\omega^2} = \frac{1}{(F(u) + \mathrm{i}G(u))} = D_1. \tag{65}$$

As an important consequence, the impedance is now related to the square of the complex frequency shift $(\Omega - n \cdot \omega_0)^2$. This has rather severe implications.

i) Consequence: no more stable region in one half of the plane.
ii) Landau damping is always required to ensure stability.

As an illustration, we show some stability diagrams derived from the new D_1. The stability diagram for unbunched beams, for the longitudinal motion, and without spread, is shown in Fig. 15. The stable region is just an infinitely narrow line for $\mathrm{Im}(D_1) = 0$ and positive $\mathrm{Re}(D_1)$.

Introducing a frequency spread for a parabolic distribution, we have the stability diagram shown in Fig. 16. The locus of the diagram is now the stable region. As in the previous example, we treat again a Lorentz distribution $\rho(\omega)$ and we show both in Fig. 17. The stable region from the Lorentz distribution is significantly larger. We can investigate this by studying the distributions. Both distributions (normalized) are displayed in Fig. 18. With a little imagination, we can assume that the difference is due to the very different populations of the tails of the distribution.

A particular problem we encounter is when we do not know the shape of the distribution. In Fig. 19, we show the stability diagrams again together with a circle inscribed inside both distributions. This is a simplified (and pessimistic) criterion for the longitudinal stability of unbunched beams [17]. Since D_1 is directly related to the machine impedance Z, we obtain a criterion which can be readily applied for estimating the beam stability. For longitudinal stability/instability, we have the condition

$$\frac{|Z_\parallel|}{n} \leq F \frac{\beta^2 E_0 |\eta_\mathrm{c}|}{qI} \left(\frac{\Delta p}{p}\right)^2. \tag{66}$$

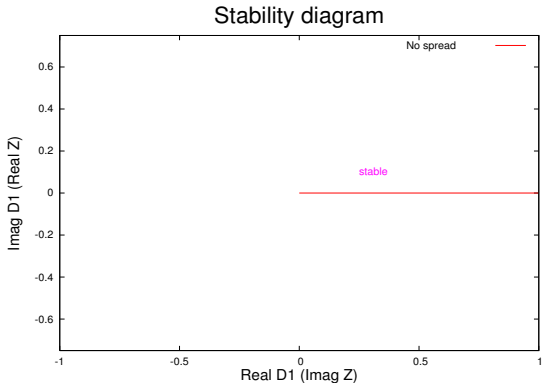


Fig. 15: Re(D_1) versus Im(D_1) for unbunched beam without spread

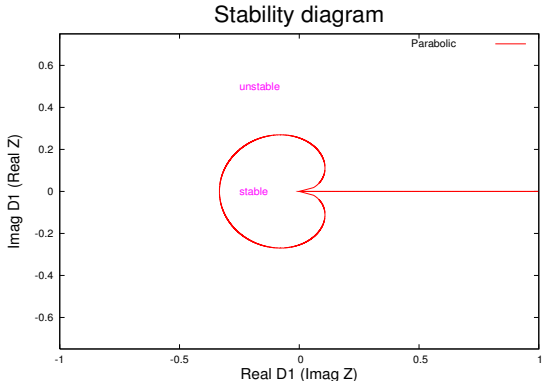


Fig. 16: Re(D_1) versus Im(D_1) for parabolic distribution $\rho(\omega)$ and unbunched beam

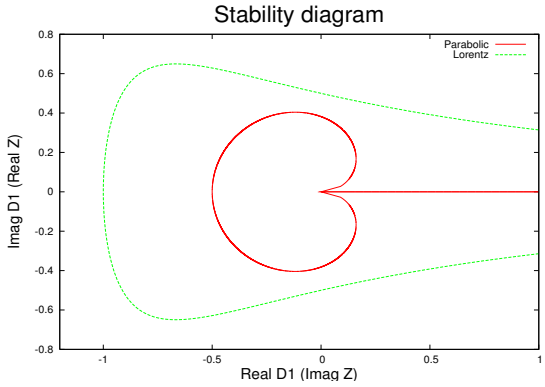


Fig. 17: Re(D_1) versus Im(D_1) for parabolic and Lorentz distributions $\rho(\omega)$ and unbunched beam

This is the Keil–Schnell criterion [17] applicable to other types of distributions and frequently used for an estimate of the maximum allowed impedance. We have the frequency spread from momentum spread and momentum compaction η_c related to the machine impedance. For given beam parameters, this defines the maximum allowed impedance $\frac{|Z_{\|}|}{n}$.

5 Effect of simplifications

We have used a few simplifications in the derivation.

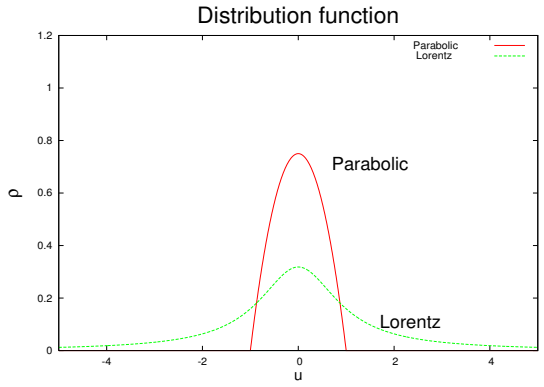


Fig. 18: Frequency distribution for parabolic and Lorentz distributions $\rho(\omega)$ and unbunched beam

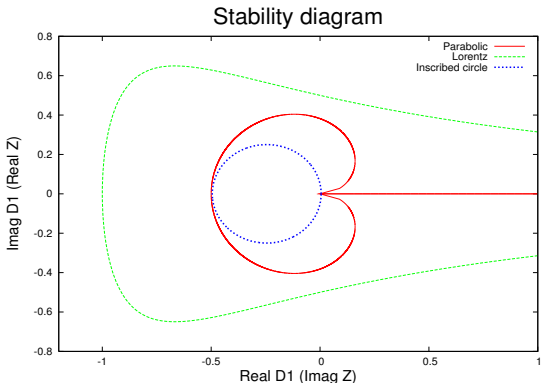


Fig. 19: Stability diagrams for parabolic and Lorentz distributions, a circle inscribed

i) Oscillators are linear.
ii) Movement of the beam is rigid (i.e. beam shape and size do not change).

What if we consider the 'real' cases, i.e. non-linear oscillators? Consider now a bunched beam; because of the synchrotron oscillation, revolution frequency and betatron spread (from chromaticity) average out. As sources for the frequency spread we have non-linear forces, such as the following examples.

i) Longitudinal: sinusoidal RF wave.
ii) Transverse: octupolar or high multipolar field components.

Can we use the same considerations as for an ensemble of *linear* oscillators?

The excited betatron oscillation will change the frequency distribution $\rho(\omega)$ (frequency depends now on the oscillation amplitude). An oscillating bunch changes the tune (and $\rho(\omega)$) due to the detuning in the non-linear fields.

A complete derivation can be done through the Vlasov equation [5], but this is well beyond the scope of this article. The equation

$$\langle x(t)\rangle = \frac{A}{2\omega_x} \mathrm{e}^{-\mathrm{i}\Omega t} \left[\mathrm{P.V.} \int \mathrm{d}\omega \, \frac{\rho(\omega)}{(\omega - \Omega)} + \mathrm{i}\pi\rho(\Omega) \right] \tag{67}$$

becomes

$$\langle x(t)\rangle = \frac{A}{2\omega_x} \mathrm{e}^{-\mathrm{i}\Omega t} \left[\mathrm{P.V.} \int \mathrm{d}\omega \, \frac{\partial\rho(\omega)/\partial\omega}{(\omega - \Omega)} + \mathrm{i}\pi\partial\rho(\Omega)/\partial\Omega \right]. \tag{68}$$

The distribution function $\rho(\omega)$ has to be replaced by the derivative $\partial\rho(\omega)/\partial\omega$. We evaluate now this configuration for instabilities in the transverse plane.
Since the frequency ω depends now on the particle's amplitudes J_x and J_y (see [16]), the expression

$$\omega_x(J_x, J_y) = \frac{\partial H}{\partial J_x} \tag{69}$$

is the amplitude-dependent betatron tune (similar for ω_y).

We then have to write

$$\rho(\omega) \Longrightarrow \rho(J_x, J_y). \tag{70}$$

Assuming a periodic force in the horizontal (x) plane and using now the tune (normalized frequency) $Q = \omega/\omega_0$,

$$F_x = A \cdot \exp(-\mathrm{i}\omega_0 Q t), \tag{71}$$

the dispersion integral can be written as (see also [25] and references therein)

$$1 = -\Delta Q_{\mathrm{coh}} \int_0^\infty \mathrm{d}J_x \int_0^\infty \mathrm{d}J_y \frac{J_x \frac{\partial\rho(J_x,J_y)}{\partial J_x}}{Q - Q_x(J_x, J_y)}. \tag{72}$$

Then we proceed as before to get the stability diagram.

If the particle distribution changes (often as a function of time), obviously the frequency distribution $\rho(\omega)$ changes as well.

i) Examples: higher-order modes; coherent beam–beam modes.
ii) Treatment requires solving the Vlasov equation (perturbation theory or numerical integration).
iii) Instead, we use a pragmatic approach: *unperturbed stability region* and *perturbed complex tune shift*.

6 Landau damping as a cure

If the boundary of

$$D_1 = \frac{1}{(f(u) + \mathrm{i}g(u))} \tag{73}$$

determines the stability, can we increase the stable region by either of the following methods?

i) Modifying the frequency distribution $Q_x(\omega)$, i.e. the distribution of the amplitudes $Q_x(J_x, J_y)$ (see definition of $f(u)$ and $g(u)$)?
ii) Introducing tune spread artificially (octupoles, other high-order fields)?

The tune dependence of an octupole (k_3) can be written as [16]

$$Q_x(J_x, J_y) = Q_0 + a \cdot k_3 \cdot J_x + b \cdot k_3 \cdot J_y. \tag{74}$$

Other sources to introduce tune spread are for example:

i) space charge;
ii) chromaticity;
iii) high-order multipole fields;
iv) beam–beam effects (colliders only).

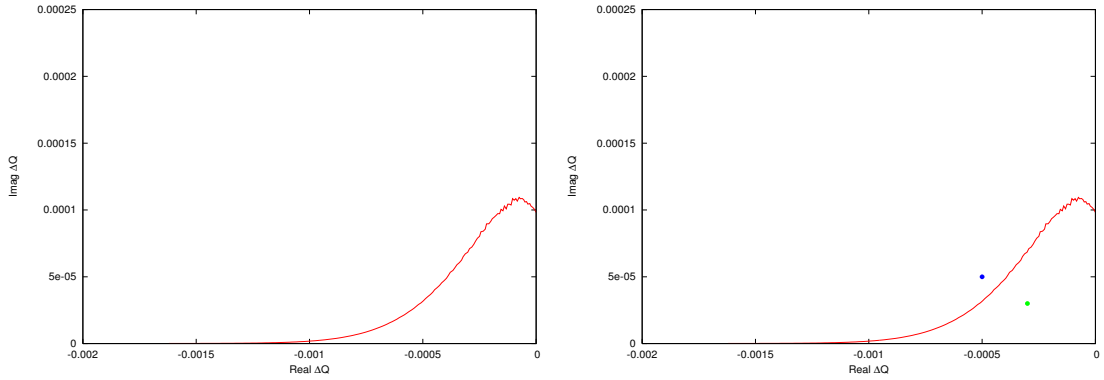


Fig. 20: Stability diagrams for an octupolar field (left) and together with the complex tune shifts for a stable and an unstable mode (right).

6.1 Landau damping in the presence of non-linear fields

The recipe for 'generating' Landau damping is:

i) for a multipole field, compute detuning $Q(J_x, J_y)$;
ii) given the distribution $\rho(J_x, J_y)$;
iii) compute the stability diagram by scanning frequency, i.e. the amplitudes.

6.2 Stabilization with octupoles

Figure 21 (left) shows the stability diagram we can get for an octupole. In Fig. 21 (right), we have added to complex tune shift for a stable and an unstable mode. The point is inside the region and the beam is stable. An unstable mode (e.g. after increase of intensity or impedance) is shown in Fig. 20 (right); it lies outside the stable region.

Theoretically, we can increase the octupolar strength to increase the stable region. However, can we increase the octupole strength as we like? There are some downsides as follows.

i) Octupoles introduce strong non-linearities at large amplitudes and may lead to reduced dynamic aperture and bad lifetime.
ii) We do not have many particles at large amplitudes; this requires large strengths of the octupoles.
iii) Can cause reduction of dynamic aperture and lifetime.
iv) They change the chromaticity via feed-down effects!
v) The lesson: use them if you have no other choice.

We see [16] that the contribution from octupoles to the stability region mainly comes from large-amplitude particles, i.e. the population of tails. This has some frightening consequences: changes in the distribution of the tails can significantly change the stable region. It is rather doubtful whether tails can be easily reproduced. Furthermore, in accelerators requiring small losses and long lifetimes (e.g. colliders) the tail particles are unwanted and often caught by the collimation system to protect the machine. A reliable calculation of the stability becomes a gamble. Proposals to artificially increase tails in a hadron collider are a bad choice although it had been proposed for fast-cycling machines since they have no need for a long beam lifetime.

6.3 Example: head–tail modes

We now apply the scheme to the explicit example of head–tail modes. They can be due to short-range wake fields or broadband impedance, etc.

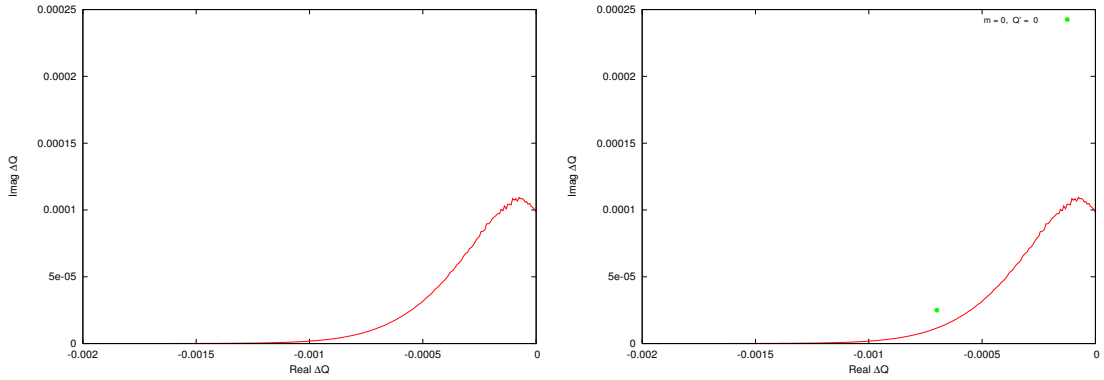


Fig. 21: Stability diagrams for an octupolar field and $m = 0$ head–tail mode with $Q' = 0$

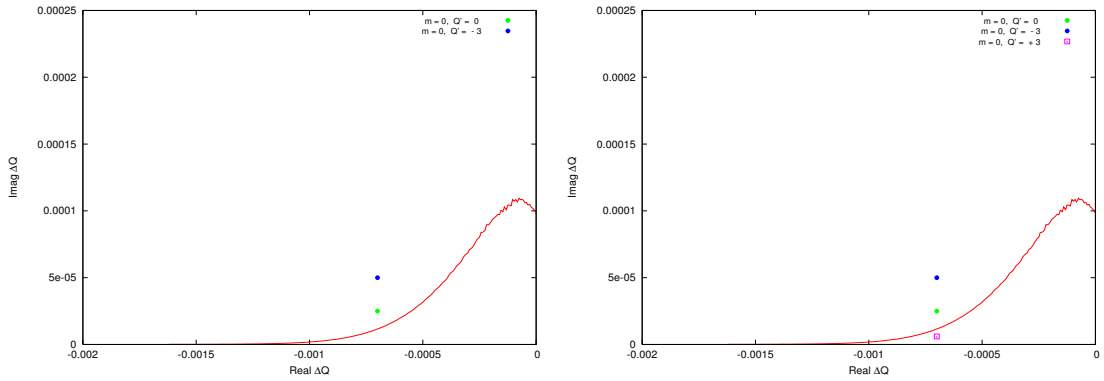


Fig. 22: Stability diagrams for an octupolar field and $m = 0$ head–tail mode with negative chromaticity $Q' = -3$ (left) and in the right-hand figure as well a head–tail mode with positive chromaticity $Q' = +3$.

The growth and damping times of head–tail modes are usually controlled with chromaticity Q':

i) some modes need positive Q';
ii) some modes need negative Q';
iii) some modes can be damped by feedback ($m = 0$).

In Figs. 22–24, the stability diagrams are shown for different head–tail modes and different values of the chromaticity Q'. For a zero chromaticity $Q' = 0$, the $m = 0$ mode is unstable (Fig. 21). The positive chromaticity stabilizes the $m = 0$ mode by shifting it into the stable area Fig. 22. A negative chromaticity makes the beam more unstable by shifting the tune shift further into the unstable region. However, too large positive chromaticity moves higher order head–tail modes to larger imaginary values, until they may become unstable (Fig. 23). Increasing the chromaticity further than in Fig. 23, the mode $m = -1$ becomes unstable. Landau damping is required in this case, but with the necessary care to avoid detrimental effects if this is achieved with octupoles.

6.4 Stabilization with other non-linear elements

Can we always stabilize the beam with strong octupole fields?

i) Would need very large octupole strength for stabilization.
ii) The known problems:
   a) can cause reduction of dynamic aperture and lifetime;

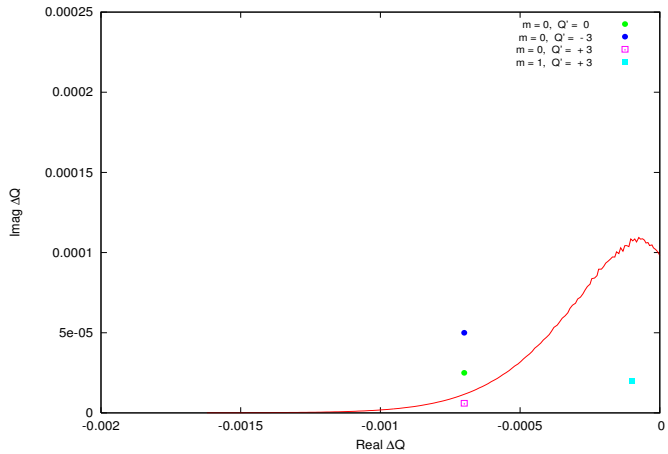


Fig. 23: Stability diagrams for an octupolar field and $m = 0$ and $m = -1$ head–tail modes with positive chromaticity $Q' = +3$.

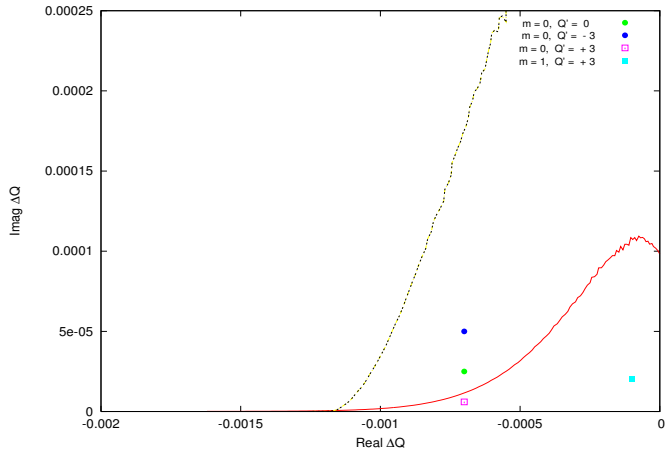


Fig. 24: Stability diagrams for an octupolar field and head–tail modes (Fig. 23) and stability diagram from beam–beam effects.

b) lifetime important when beam stays in the machine for a long time;

c) colliders, lifetime more than 10–20 h needed.

iii) Is there another option?

In the case of colliders, the beam–beam effects provide a large tune spread that can be used for Landau damping (Fig. 24). Even when the tune spread is comparable, the stability region from a head-on beam–beam interaction is much increased [18]. We observe a large difference between the two stability diagrams. Where does this difference come from? The tune dependence of an octupole can be written as

$$Q_x(J_x, J_y) = Q_0 + aJ_x + bJ_y \tag{75}$$

and is linear in the action J (for coefficients, see Appendix B).

The tune dependence of a head-on beam–beam collision can be written [19, 20] as follows: with $\alpha = x/\sigma^*$, we get $\Delta Q/\xi = (4/\alpha^2)\left[1 - I_0(\alpha^2/4)\cdot \mathrm{e}^{-\alpha^2/4}\right]$. As a demonstration, we show in Fig. 25 the tune spread from octupoles, long-range beam–beam effects [19] and head-on beam–beam effects. The parameters are adjusted such that the overall tune spread is always the same for the three cases.

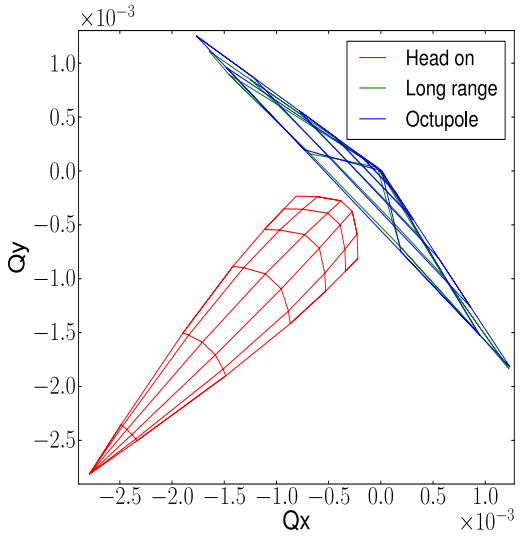


Fig. 25: Tune spread in two dimensions for octupoles, long-range beam–beam and head-on beam–beam. Tune spread adjusted to be the same for all cases.

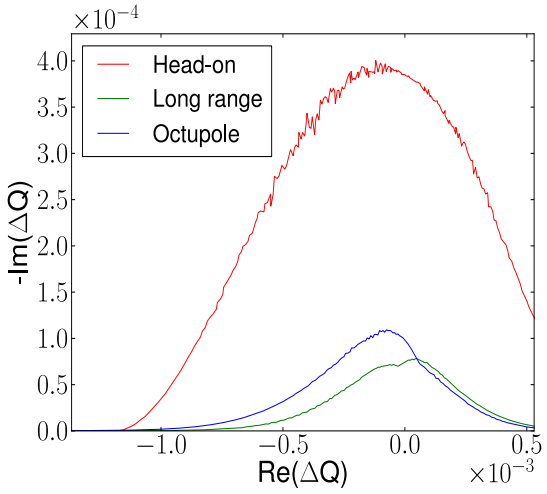


Fig. 26: Stability diagrams for octupoles, long-range beam–beam and head-on beam–beam. Tune spread adjusted to be the same for all cases.

However, the head-on beam–beam tune spread has the largest effect for *small* amplitudes, i.e. where the particle density is largest and the contribution to the stability is strong. For the tune spreads shown in Fig. 25, we have computed the corresponding stability diagrams [21]. Although the tune spread is the same, the stability region shown in Fig. 26 is significantly larger. As another and most important advantage, since the stability is ensured by small-amplitude particles, the stable region is very insensitive to the presence of tails or the exact distribution function. Special cases such as collisions of 'hollow beams' are not noteworthy in this overview.

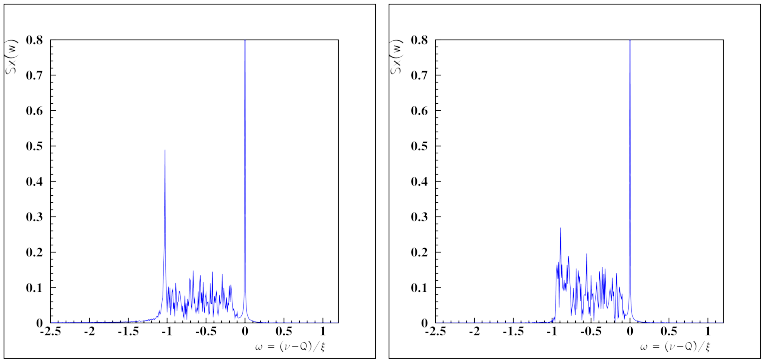


Fig. 27: Coherent beam–beam modes with π-mode inside and outside the incoherent tune spread due to beam–beam effects. Decoupling is due to intensity difference between the two beams.

6.4.1 Conditions for Landau damping

A tune spread is not sufficient to ensure Landau damping. To summarize, we need:

i) presence of incoherent frequency (tune) spread;
ii) coherent mode must be inside this spread;
iii) the same particles must be involved in the oscillation and the spread.

The last item requires some explanation. When the tune spread is not provided by the particles actually involved in the coherent motion, it is not sufficient and results in no damping. For example, this is the case for coherent beam–beam modes where the eigenmodes of coherent beam–beam modes driven either by head-on effects or long range interactions are very different [22–24]. For modes driven by the head-on interactions, mainly particles at small amplitudes are participating in the oscillation. The tune spread produced by long range interactions produces little damping in this case.

Another option to damp coherent beam–beam modes is shown in Fig. 27. It shows the results of a simulation for the main coherent beam–beam modes (0-mode and π-mode) for slightly different intensities.

The difference is sufficient to decouple the two beams and they cannot maintain the correct phase. We have:

i) coherent mode inside or outside the incoherent spectrum, depending on the intensity difference;
ii) Landau damping restored when the symmetry is sufficiently broken.

6.5 Landau damping with non-linear fields: are there any side effects?

Introducing non-linear fields into an accelerator is not always a desirable procedure. It may have implications for the beam dynamics and we separate them into three categories.

i) Good:
   a) stability region increased.
ii) Bad:
   a) non-linear fields introduced (resonances);
   b) changes optical properties, e.g. chromaticity (feed-down).
iii) Special cases:

a) non-linear effects for large amplitudes (octupoles);
b) much better: head-on beam–beam (but only in colliders).

Landau damping with non-linear fields is a very powerful tool, but the side effects and implications have to be taken into account.

7 Summary

The collisionless damping of coherent oscillations as predicted by Landau in 1946 is an important concept in plasma physics as well as in other applications such as hydrodynamics, astrophysics and biophysics to mention a few. It is used extensively in particle accelerators to avoid coherent oscillations of the beams and the instabilities. Despite its intensive use, it is not a simple phenomenon and the interpretation of the physics behind the mathematical structures is a challenge. Even after many decades after its discovery, there is (increasing) interest in this phenomenon and work on the theory (and extensions of the theory) continues.

References

[1] I. Langmuir and L. Tonks, *Phys. Rev.* **33** (1929) 195.
[2] L.D. Landau, *J. Phys. USSR* **10** (1946) 26.
[3] D. Bohm and E. Gross, *Phys. Rev.* **75** (1949) 1851.
[4] D. Bohm and E. Gross, *Phys. Rev.* **75** (1949) 1864.
[5] D. Sagan, On the physics of Landau damping, CLNS 93/1185 (1993).
[6] K. Case, *Ann. Phys.* **7** (1959) 349.
[7] N.G. Van Kampen, *Physica* **21** (1955) 949.
[8] J. Malmberg and C. Wharton, *Phys. Rev. Lett.* **13** (1964) 184.
[9] V. Neil and A. Sessler, *Rev. Sci. Instrum.* **36** (1965) 429.
[10] L. Laslett, V. Neil and A. Sessler, *Rev. Sci. Instrum.* **36** (1965) 436.
[11] A.A. Vlasov, *J. Phys. USSR* **9** (1945) 25.
[12] A. Chao, *Theory of Collective Beam Instabilities in Accelerators* (Wiley, New York, 1993).
[13] A. Hofmann, Landau damping, Proc. CERN Accelerator School (2009).
[14] A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1998).
[15] G. Rumolo, Beam instabilities, these proceedings, CERN Accelerator School (2013).
[16] W. Herr, Mathematical and numerical methods for non-linear beam dynamics, these proceedings, CERN Accelerator School (2013).
[17] E. Keil and W. Schnell, Concerning longitudinal stability in the ISR, CERN-ISR-TH-RH/69-48 (1969).
[18] W. Herr and L. Vos, Tune distributions and effective tune spread from beam–beam interactions and the consequences for Landau damping in the LHC, LHC Project Note 316 (2003).
[19] W. Herr, Beam–beam effects, Proc. CERN Accelerator School (2003).
[20] T. Pieloni, Beam–beam effects, these proceedings, CERN Accelerator School (2013).
[21] X. Buffat, Consequences of missing collisions – beam stability and Landau damping, Proc. ICFA Beam–Beam Workshop 2013, CERN (2013).
[22] Y. Alexahin, W. Herr *et al.*, Coherent beam–beam effects, Proc. HEACC 2001, Tsukuba, Japan, 2001.
[23] Y. Alexahin, A study of the coherent beam–beam effect in the framework of the Vlasov perturbation theory, LHC Project Report 461 (2001).

[24] Y. Alexahin, *Nucl. Instrum. Methods* **A480** (2002) 235.

[25] J. Scott Berg and F. Ruggiero, Landau damping with two-dimensional tune spread, CERN SL-AP-96-71 (AP) (1996).

Appendices

A Solving the dispersion relation

A.1 Dispersion relation from Vlasov's calculation

Starting with the dispersion relation derived using Vlasov's approach (20):

$$1 + \frac{\omega_\mathrm{p}^2}{k^2} \int \frac{\partial\psi_0/\partial v}{(\frac{\omega}{k} - v)} \,\mathrm{d}v = 0, \tag{A.1}$$

we have to make a few assumptions. We assume that we can restrict ourselves to waves with $\omega/k \gg v$ or $\omega/k \ll v$. The latter case cannot occur in Langmuir waves and we can assume the case with $\omega/k \gg v$. Then we can integrate the integral by parts and get

$$1 + \frac{\omega_\mathrm{p}^2}{k^2} \int \frac{\psi_0}{(\omega/k - v)^2} \,\mathrm{d}v = 0 \tag{A.2}$$

or, rewritten for the next step,

$$1 + \frac{\omega_\mathrm{p}^2}{\omega^2} \int \frac{\psi_0}{(1 - vk/\omega)^2} \,\mathrm{d}v = 0. \tag{A.3}$$

With the assumption $\omega/k \gg v$, we can expand the denominator in series of $\frac{vk}{\omega}$ and obtain

$$1 + \frac{\omega_\mathrm{p}^2}{\omega^2} \int \psi_0 \mathrm{d}v \cdot \left(1 + 2 \cdot \left(\frac{vk}{\omega}\right) + 3 \cdot \left(\frac{vk}{\omega}\right)^2\right) = 0. \tag{A.4}$$

For the next step as an explicit example we use a Maxwellian velocity distribution, i.e.

$$\psi(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{v_\mathrm{p}} \mathrm{e}^{-v^2/2v_\mathrm{p}^2}. \tag{A.5}$$

The individual integrals are then

$$\int \psi(v) \,\mathrm{d}v = 1, \quad \int \psi(v) \cdot v \,\mathrm{d}v = 0, \quad \int \psi(v) \cdot v^2 \,\mathrm{d}v = v_\mathrm{p}^2 = \frac{\omega_\mathrm{p}^2}{k^2} \tag{A.6}$$

to obtain finally

$$1 - \frac{\omega_\mathrm{p}^2}{\omega^2} - \frac{3k^2 v_\mathrm{p}^2 \omega_\mathrm{p}^2}{\omega^4} = 0. \tag{A.7}$$

This dispersion relation can now be solved for ω and we get two solutions:

$$\omega^2 = \frac{1}{2}\omega_\mathrm{p}^2 \pm \frac{1}{2}\omega_\mathrm{p}^2 \left(1 + \frac{12k^2 v_\mathrm{p}^2}{\omega_\mathrm{p}^2}\right)^{1/2}. \tag{A.8}$$

Rewritten, we obtain the well-known dispersion relation for Langmuir waves:

$$\omega^2 = \omega_\mathrm{p}^2 \left(1 + 3k^2\lambda^2\right) \quad (\lambda = v_\mathrm{p}/\omega_\mathrm{p}). \tag{A.9}$$

The frequency is real and there is no damping.

A.2 Dispersion relation from Landau's approach

Here we solve the dispersion relation using the result obtained by Landau (26):

$$1+\frac{\omega_{\mathrm{p}}^2}{k}\left[\mathrm{P.V.}\int\frac{\partial\psi_0/\partial v}{(\omega-kv)}\,\mathrm{d}v-\frac{\mathrm{i}\pi}{k}\left(\frac{\partial\psi_0}{\partial v}\right)_{v=\omega/k}\right]=0. \tag{A.10}$$

This should lead to a damping.

Integration by parts leads now to

$$1-\frac{\omega_{\mathrm{p}}^2}{\omega^2}-\frac{3k^2v_{\mathrm{p}}^2\omega_{\mathrm{p}}^2}{\omega^4}-\frac{\mathrm{i}\pi}{k}\left(\frac{\partial\psi_0}{\partial v}\right)_{v=\omega/k}=0. \tag{A.11}$$

For the real part, we can use the same reasoning as for Vlasov's calculation and find again

$$1-\frac{\omega_{\mathrm{p}}^2}{\omega_{\mathrm{r}}^2}-\frac{3k^2v_{\mathrm{p}}^2\omega_{\mathrm{p}}^2}{\omega_{\mathrm{r}}^4}=0. \tag{A.12}$$

Now we have used ω_{r} to indicate that we computed the real part of the complex frequency

$$\omega=\omega_{\mathrm{r}}+\mathrm{i}\cdot\omega_{\mathrm{i}}. \tag{A.13}$$

We assume 'weak damping', i.e. $\omega_{\mathrm{i}}\ll\omega_{\mathrm{r}}$. This leads us to $\omega^2\simeq\omega_{\mathrm{r}}+2\mathrm{i}\omega_{\mathrm{r}}\omega_{\mathrm{i}}$. With the solution (A.9) and $k\lambda\ll1$, we find for the imaginary part ω_{i}

$$\omega_{\mathrm{i}}=\frac{\pi\cdot\omega_{\mathrm{p}}^3}{2\cdot k^2}\cdot\left(\frac{\partial\psi_0}{\partial v}\right)_{v=\omega/k}. \tag{A.14}$$

Using the Maxwell velocity distribution as an example:

$$\psi(v)=\frac{1}{\sqrt{2\pi}}\frac{1}{v_{\mathrm{p}}}\mathrm{e}^{-v^2/2v_{\mathrm{p}}^2}, \tag{A.15}$$

we get for the derivative

$$\frac{\partial\psi(v)}{\partial v}=-\frac{1}{\sqrt{2\pi}}\frac{1}{v_{\mathrm{p}}^3}\mathrm{e}^{-v^2/2v_{\mathrm{p}}^2}. \tag{A.16}$$

Using again $\lambda=v_{\mathrm{p}}/\omega_{\mathrm{p}}$ to simplify the expression, we can expand in a series (because $\omega/k\gg v$) and arrive at

$$\omega_{\mathrm{i}}=-\omega_{\mathrm{p}}\cdot\frac{1}{k^3\lambda^3}\sqrt{\frac{\pi}{2}}\cdot\exp\left(-\frac{1}{2k^2\lambda^2}-\frac{3}{2}\right). \tag{A.17}$$

This is the damping term we obtain using Landau's result for plasma oscillations.

B Detuning with octupoles

The tune dependence of an octupole can be written as [16]

$$Q_x(J_x,J_y)=Q_0+aJ_x+bJ_y \tag{B.1}$$

for the coefficients:

$$\Delta Q_x=\left[\frac{3}{8\pi}\int\beta_x^2\frac{K_3}{B\rho}\,\mathrm{d}s\right]J_x-\left[\frac{3}{8\pi}\int2\beta_x\beta_y\frac{K_3}{B\rho}\,\mathrm{d}s\right]J_y \tag{B.2}$$

and

$$\Delta Q_y=\left[\frac{3}{8\pi}\int\beta_y^2\frac{K_3}{B\rho}\,\mathrm{d}s\right]J_y-\left[\frac{3}{8\pi}\int2\beta_x\beta_y\frac{K_3}{B\rho}\,\mathrm{d}s\right]J_x. \tag{B.3}$$

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam Instabilities in Linear Machines: Wakefields Effects[1]

M. Ferrario, M. Migliorati, and L. Palumbo
INFN-LNF and Università di Roma 'La Sapienza'

Abstract
When a charged particle travels across the vacuum chamber of an accelerator, it induces electromagnetic fields, which are left mainly behind the generating particle. These electromagnetic fields act back on the beam and influence its motion. Such an interaction of the beam with its surroundings results in beam energy losses, alters the shape of the bunches, and shifts the betatron and synchrotron frequencies. At high beam current the fields can even lead to instabilities, thus limiting the performance of the accelerator in terms of beam quality and current intensity. We discuss in this lecture the general features of the electromagnetic fields, introducing the concepts of wakefields and giving a few simple examples in cylindrical geometry. We then show the effect of the wakefields on the dynamics of a beam in a linac, dealing in particular with the beam breakup instability and how to cure it.

Keywords
Wakefields; instabilities; Linac; beam breakup; BNS damping; coupling impedance.

1 Introduction

Self-induced electromagnetic (e.m.) forces in an accelerator are generated by a charged particle beam which interacts with all the components of the vacuum chamber. These components may have a complex geometry: kickers, bellows, RF cavities, diagnostics components, special devices, etc. The study of the fields generally requires the solution of Maxwell's equations in a given structure, taking the beam current as the source of the fields. This could be a quite complicated task, and therefore several dedicated computer codes, used to study and design accelerator devices, which solve the e.m. problem in the frequency or time domain, have been developed. These include, for example, CST Studio Suite [1], GDFIDL [2], ACE3P [3], ABCI [4], and others.

In this lecture, we discuss some general features of the self-induced e.m. forces and introduce the concepts of wakefields and coupling impedances [5–12], and we present some simple examples in cylindrical geometry. Although the space charge forces have been studied separately [13], they can be seen as a particular case of wakefields.

In the second part of the lecture we study the effects of the wakefields on the dynamics of a beam in a linac, such as energy loss and energy spread. Finally, we deal with the beam breakup (BBU) instability [14], and the way to cure it [15].

[1] This article has already been published in the Proceedings of the CAS-CERN Accelerator School on Advanced Accelerator Physics, CERN-2014-009 (CERN, Geneva, 2014). https://doi.org/10.5170/CERN-2014-009.357

https://doi.org/10.23730/CYRSP-2017-003.165

2 Wake fields and potentials

2.1 Longitudinal and transverse wakefields

The self-induced e.m. fields acting on a particle inside a beam depend on the whole charge distribution. However, if we know the fields in a given structure that are created by a single charge (i.e. we obtain the Green function of the structure), by using the superposition principle we can easily reconstruct the fields produced by any charge distribution.

The e.m. fields created by a point charge act back on the charge itself, and on any other charge of the beam. Referring to the coordinate system of Fig. 1, let us call $q_0(s_0, \boldsymbol{r}_0)$ the source charge, travelling with constant longitudinal velocity $v = c$ (ultrarelativistic limit) along a trajectory parallel to the axis of a given accelerator structure. Let us consider a test charge q, in a position $(s = s_0 - z, \boldsymbol{r})$, which is moving with the same constant velocity on a parallel trajectory inside the structure.

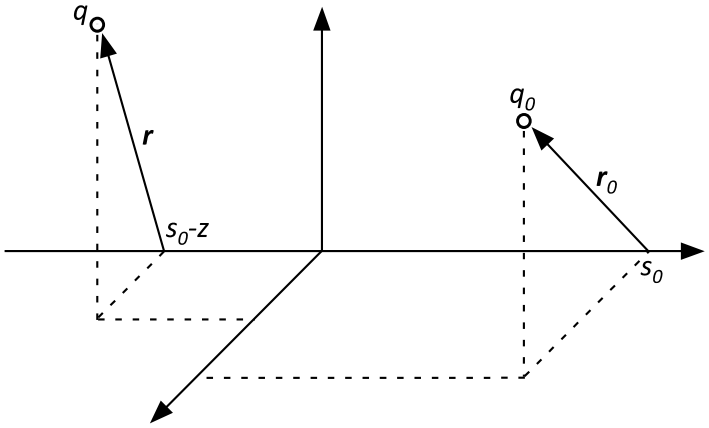


Fig. 1: Reference coordinate system

Let $\boldsymbol{E}$ and $\boldsymbol{B}$ be the electric and magnetic fields generated by q_0 inside the structure. Since the velocity of both charges is along z, the Lorentz force acting on q has the following components:

$$\boldsymbol{F} = q\left[E_z\hat{z} + \left(E_x - vB_y\right)\hat{x} + \left(E_y + vB_x\right)\hat{y}\right] \equiv \boldsymbol{F}_{/\!/} + \boldsymbol{F}_{\perp}\,. \tag{1}$$

From Eq. (1) we see that there can be two effects on the test charge: a longitudinal force, which changes its energy, and a transverse force, which deflects its trajectory. If we consider a device of length L, the energy change (in joule) of q due to this force is given by

$$U\left(\boldsymbol{r}, \boldsymbol{r}_0, z\right) = \int_0^L F_{/\!/}\ \mathrm{d}s\,, \tag{2}$$

and the transverse deflecting kick (expressed in Newton metres), is given by

$$\boldsymbol{M}\left(\boldsymbol{r}, \boldsymbol{r}_0, z\right) = \int_0^L \boldsymbol{F}_{\perp}\ \mathrm{d}s\,. \tag{3}$$

Note that the integration is performed over a given path of the trajectory. The quantities given by Eqs. (2) and (3), normalized to the two charges q_0 and q, are called the longitudinal and transverse *wakefields*, respectively. In many cases, we deal with structures having particular symmetric shapes, generally cylindrical. It is possible to demonstrate that, with a multipole expansion of the wakefields, the dominant term in the longitudinal wakefield depends only on the distance z between the two charges, while the dominant one in the transverse wakefield, although still a function of the distance z, is also linear with the transverse position of the source charge $\boldsymbol{r}_0$. If we then divide the transverse wakefield by

r_0, we obtain the transverse dipole wakefield, that is the transverse wake per unit of transverse displacement, depending only on z.

Longitudinal wakefield [V/C]: $$w_{/\!/}(z) = -\frac{U}{q_0 q} \; ; \tag{4}$$

Transverse dipole wakefield [V/C·m]: $$\boldsymbol{w}_{\perp}(z) = \frac{1}{r_0}\frac{\boldsymbol{M}}{q_0 q} . \tag{5}$$

The minus sign in the definition of the longitudinal wakefield means that the test charge loses energy when the wake is positive. A positive transverse wake means that the transverse force is defocusing. The wakefields are properties of the vacuum chamber and the beam environment, but they are independent of the beam parameters (bunch size, bunch length, etc.).

To study the effect of wakefields on the beam dynamics, it is convenient to distinguish between short-range wakefields, which are generated by the particles at the head of the bunch affecting trailing particles in the same bunch, and those that influence the multibunch (or multiturn) beam dynamics, which are generally resonant modes trapped inside a structure, and are called long-range wakefields.

As a first example of wakefields, let us consider the longitudinal wakefield of 'space charge'. Even if in the ultrarelativistic limit with $\gamma \to \infty$, and there is no space charge effect, we can still define a wakefield by considering a moderately relativistic beam with $\gamma \gg 1$ but not infinite. It turns out that the space charge forces can fit into the definition of a wakefield, and when that is done we find that the wake depends on the beam properties such as the transverse beam radius a and the beam energy γ. In Appendix A we show an example of such an interpretation. Let us consider here a relativistic beam with cylindrical symmetry and uniform transverse distribution of radius a. The longitudinal force acting on a charge q of the beam travelling inside a cylindrical pipe of radius b is given by [13]

$$F_{/\!/}(r,z) = \frac{-q}{4\pi\varepsilon_0\gamma^2}\left(1 - \frac{r^2}{a^2} + 2\ln\frac{b}{a}\right)\frac{\partial\lambda(z)}{\partial z} \quad , \tag{6}$$

with $\lambda(z)$ the longitudinal distribution ($z > 0$ at the bunch head). Note that, since the space charge forces move together with the beam, they are constant along the accelerator if the beam pipe cross-section remains constant. We can therefore define the longitudinal wakefield per unit length [V/C·m]. To obtain the longitudinal wakefield of a piece of pipe, we just multiply by the pipe length. Assuming $r \to 0$ (particle on-axis), and a charge line density given by $\lambda(z) = q_0\delta(z)$, we obtain

$$\frac{\mathrm{d}w_{/\!/}(z)}{\mathrm{d}s} = \frac{1}{4\pi\varepsilon_0\gamma^2}\left(1 + 2\ln\frac{b}{a}\right)\frac{\partial}{\partial z}\delta(z) , \tag{7}$$

which has the peculiarity of being also dependent on the beam size a.

Another interesting case is the longitudinal wake potential of a resonant higher-order mode (HOM) in an RF cavity, which is an example of long-range wakefield. When a charge crosses a resonant structure, as an RF cavity, it excites the fundamental mode and HOMs. Each mode can be treated as an electric RLC circuit loaded by an impulsive current, as shown in Fig. 2.

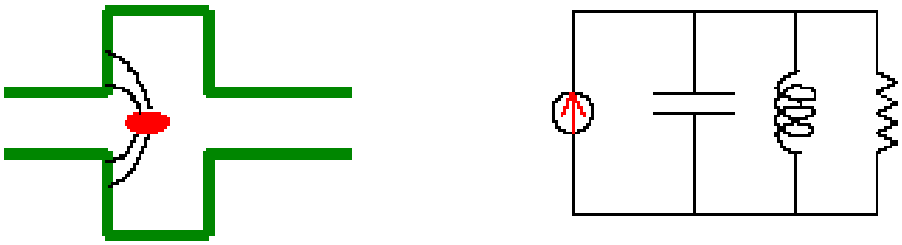


Fig. 2: RF cavity and the equivalent RLC parallel circuit model driven by a current generator

Just after the charge passage, the capacitor is charged with a voltage $V(0) = q_0 / C$, and the longitudinal electric field is $E_z = V / l$, where l is the length of the cavity. The time evolution of the electric field is then governed by the same differential equation of the voltage, which can be written as follows:

$$\ddot{V} + \frac{1}{RC}\dot{V} + \frac{1}{LC}V = \frac{1}{C}\dot{I}. \tag{8}$$

The passage of the impulsive current charges only the capacitor, which changes its potential by an amount $V_c(0)$. This potential will oscillate and decay, producing a current flow in the resistor and inductance. After the charge passage, for $t > 0$ the potential satisfies the following equation and boundary conditions:

$$\begin{aligned} &\ddot{V} + \frac{1}{RC}\dot{V} + \frac{1}{LC}V = 0, \\ &V(t=0^+) = \frac{q_0}{C} = V_0, \\ &\dot{V}(t=0^+) = \frac{\dot{q}}{C} = -\frac{I(0^+)}{C} = -\frac{V_0}{RC}, \end{aligned} \tag{9}$$

which has the following solution:

$$\begin{aligned} &V(t) = V_0 \mathrm{e}^{-\Gamma t}\left[\cos(\bar{\omega}t) - \frac{\Gamma}{\bar{\omega}}\sin(\bar{\omega}t)\right], \\ &\bar{\omega}^2 = \omega_r^2 - \Gamma^2, \end{aligned} \tag{10}$$

where $\omega_r^2 = 1/LC$ and $\Gamma = 1/2RC$. For the HOM it is also convenient to define the quality factor $Q = \omega_r / 2\Gamma$, from which we can write $C = Q/R\omega_r$.

Putting $z = ct$ (z is positive behind the source charge), we obtain the longitudinal wakefield shown in Fig. 3:

$$w_{//}(z) = \frac{V(z)}{q_0} = \frac{R\omega_r}{Q}\mathrm{e}^{-\Gamma z/c}\left[\cos(\bar{\omega}z/c) - \frac{\Gamma}{\bar{\omega}}\sin(\bar{\omega}z/c)\right]. \tag{11}$$

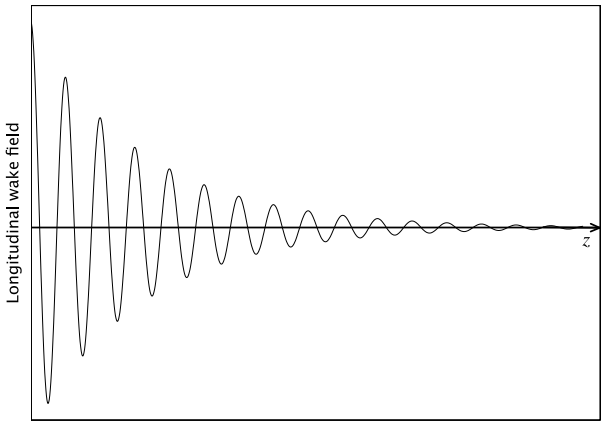


Fig. 3: Qualitative behaviour of a resonant mode wakefield

In an analogous way, it is possible to obtain the transverse wakefield of a HOM,

$$w_{\perp}(z) = \frac{R_{\perp}\omega_{\mathrm{r}}}{Q}\mathrm{e}^{-\Gamma z/c}\sin(\bar{\omega}z/c)\,, \tag{12}$$

with $R_{\perp}$ expressed in ohm per metre.

We conclude this section by giving the longitudinal and transverse short-range wakefields of a rectangular cell, as shown in Fig. 4, under the hypothesis that the bunch length is much smaller than the pipe radius *b*. Its expression can be useful in studying the effects of the short-range wakefields of an accelerating structure in a linac.

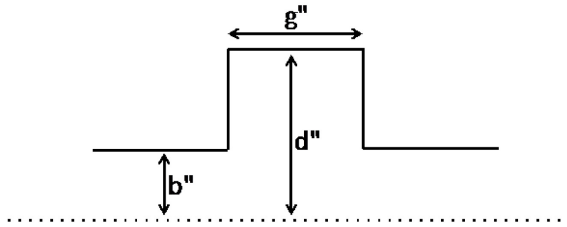


Fig. 4: Geometry of a single cell of a linac accelerating structure

The model considers each cell as a pillbox cavity. When a bunch reaches the edge of the cavity, the e.m. field it creates is simply the one that would occur when a plane wave passes through a hole; using this hypothesis, it is possible to use the classical diffraction theory of optics to calculate the fields [7]. If the condition $g < (d-b)^2/(2\sigma)$ is satisfied, where *g* is the cell gap, *d* is the cell radius, and σ is the rms bunch length of a Gaussian bunch, the longitudinal and transverse wakefields can be written, respectively, as:

$$\begin{aligned} w_{/\!/}(z) &= \frac{Z_0 c}{\sqrt{2}\pi^2 b}\sqrt{\frac{g}{z}}, \\ w_{\perp}(z) &= \frac{2^{3/2} Z_0 c}{\pi^2 b^3}\sqrt{gz}. \end{aligned} \tag{13}$$

For a collection of cavities, Eqs. (13) cannot be used because the wakefields along the cells do not sum in phase and the result would be an overestimation of the effects. An asymptotic wakefield for a periodic collection of cavities of period *p*, obtained numerically at SLAC [16] and then fitted to a simple function, is used instead. Such wakefields are thus valid after a certain number of cavities given by

$$N_{\mathrm{cr}} = \frac{b^2}{2g\left(\sigma + \dfrac{2b}{\gamma}\right)}. \tag{14}$$

Under these assumptions, the wakefields of Eqs. (13) are modified to

$$\begin{aligned} w_{/\!/}(z) &= \frac{Z_0 cp}{\pi b^2} \mathrm{e}^{-\sqrt{z/s_1}}, \\ w_{\perp}(z) &= \frac{4 Z_0 c p s_2}{\pi b^4}\left[1 - \left(1 + \sqrt{\frac{z}{s_2}}\right)\mathrm{e}^{-\sqrt{z/s_2}}\right], \end{aligned} \tag{15}$$

with

$$\begin{aligned} s_1 &= 0.41 \frac{b^{1.8} g^{1.6}}{p^{2.4}}, \\ s_2 &= 0.17 \frac{b^{1.79} g^{0.38}}{p^{1.17}}. \end{aligned} \tag{16}$$

2.2 Loss factor and beam loading theorem

A useful quantity for the effects of the longitudinal wakefield on the beam dynamics is the loss factor, defined as the normalized energy lost by the source charge q_0:

$$k = -\frac{U(z=0)}{q_0^2}. \tag{17}$$

For charges travelling at the velocity of light, there is the problem that the longitudinal wakefield is discontinuous at $z = 0$, as shown in Fig. 5, leading to an ambiguity in the evaluation of the loss factor. Indeed, when the source charge travels at the velocity of light, it leaves the e.m. fields in its wake, hence the reason why we call these fields 'wakefields'. Any e.m. perturbation produced by the charge cannot overtake the charge itself. This means that the longitudinal wakefield vanishes in the region $z < 0$. This property is a consequence of the *causality principle*. It is the causality that requires that the longitudinal wakefield of a charge travelling at the velocity of light to be discontinuous at the origin.

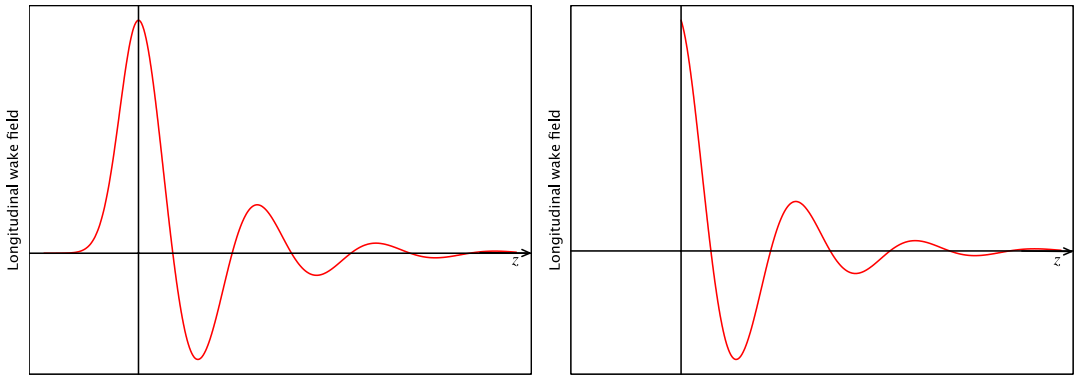


Fig. 5: Examples of longitudinal wakefields: left $\beta < 1$, right $\beta = 1$

The exact relationship between k and $w_{/\!/}(z \to 0)$ is, in this case, given by the beam loading theorem [17], which states that

$$k = \frac{w_{//}(z \to 0)}{2}. \tag{18}$$

As an example of verification of the beam loading theorem, let us consider the wakefield of the resonant mode given by Eq. (11). The energy lost by the charge q_0 loading the capacitor is $U = CV_0^2/2 = q_0^2/2C$, giving $k = 1/2C$, compared with $w_{//}(z \to 0) = 1/C$.

2.3 Relationship between transverse and longitudinal forces

Another important feature worth mentioning here is the differential relationship existing between longitudinal and transverse forces and the corresponding wakefields: the transverse gradient of the longitudinal force/wake is equal to the longitudinal gradient of the transverse force/wake, that is

$$\begin{aligned} \nabla_\perp F_{//} &= \frac{\partial}{\partial z}\boldsymbol{F}_\perp, \\ \nabla_\perp w_{//} &= \frac{\partial}{\partial z}\boldsymbol{w}_\perp. \end{aligned} \tag{19}$$

The above relations are known as the *Panofsky–Wenzel theorem* [18].

2.4 Coupling impedance

The wakefields are generally used to study the beam dynamics in the time domain. If we take the equations of motion in the frequency domain, we need the Fourier transform of the wakefields. Since these quantities have units of ohm, they are called *coupling impedances*.
Longitudinal impedance [Ω]:

$$Z_{//}(\omega) = \frac{1}{v}\int_{-\infty}^{\infty} w_{//}(z)\mathrm{e}^{\mathrm{i}\frac{\omega z}{v}}\,\mathrm{d}z\,; \tag{20}$$

Transverse dipole impedance [Ω/m]:

$$\boldsymbol{Z}_\perp(\omega) = -\frac{i}{v}\int_{-\infty}^{\infty} \boldsymbol{w}_\perp(z)\mathrm{e}^{\mathrm{i}\frac{\omega z}{v}}\,\mathrm{d}z\,. \tag{21}$$

The longitudinal coupling impedance of the space charge wake given by Eq. (7) [Ω/m] is given by

$$\frac{\partial Z_{//}(\omega)}{\partial s} = \frac{1}{v}\int_{-\infty}^{\infty}\frac{\partial w_{//}(z)}{\partial s}\mathrm{e}^{\mathrm{i}\frac{\omega z}{v}}\,\mathrm{d}z = \frac{1+2\ln(b/a)}{v4\pi\varepsilon_0\gamma^2}\int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}z}\delta(z)\mathrm{e}^{\mathrm{i}\frac{\omega z}{v}}\,\mathrm{d}z\,; \tag{22}$$

since $\int_{-\infty}^{\infty}\delta'(z)f(z)\mathrm{d}z = f'(0)$, we get

$$\frac{\partial Z_{//}(\omega)}{\partial s} = \frac{i\omega Z_0}{4\pi c\beta^2\gamma^2}\left(1+2\ln\frac{b}{a}\right). \tag{23}$$

The longitudinal coupling impedance of a resonant HOM, corresponding to the Fourier transform of Eq. (11), is given by

$$Z_{/\!/}(\omega)=\frac{R}{1+\mathrm{i}Q\left(\frac{\omega_\mathrm{r}}{\omega}-\frac{\omega}{\omega_\mathrm{r}}\right)}, \tag{24}$$

where R is also called the shunt impedance of the longitudinal HOM. Note that the loss factor can be written as $k=\omega_\mathrm{r}R/2Q$.

The transverse impedance obtained from Eq. (12) is given by

$$Z_{\perp}(\omega)=\frac{\bar{\omega}}{\omega}\frac{R_{\perp}}{1+\mathrm{i}Q\left(\frac{\omega_\mathrm{r}}{\omega}-\frac{\omega}{\omega_\mathrm{r}}\right)}, \tag{25}$$

where $R_\perp$ is called the transverse shunt impedance.

2.5 Wake potential and energy loss of a bunched distribution

When we have a bunch with total charge q_0 and longitudinal distribution $\lambda(z)$, such that $q_0=\int_{-\infty}^{\infty}\lambda(z')\mathrm{d}z'$, we can obtain the amount of energy lost or gained by a single charge q in the beam by using the superposition principle.

To this end, we calculate the effect on the charge by the whole bunch, as shown in Fig. 6, with the superposition principle, which gives the convolution integral:

$$U(z)=-q\int_{-\infty}^{\infty}w_{/\!/}\left(z'-z\right)\lambda(z')\mathrm{d}z'. \tag{26}$$

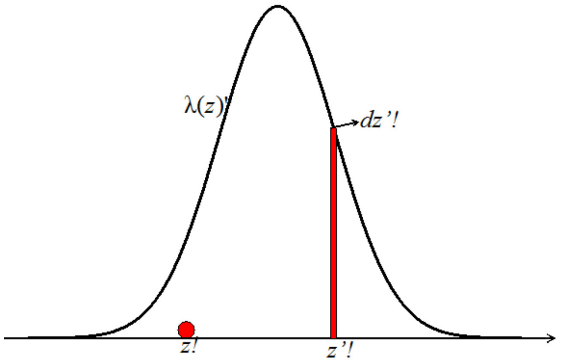


Fig. 6: Convolution integral for a charge distribution to obtain the energy loss of a particle due to the whole bunch

Equation (26) allows us to define the longitudinal wake potential of a distribution:

$$W_{/\!/}(z)=-\frac{U(z)}{qq_0}=\frac{1}{q_0}\int_{-\infty}^{\infty}w_{/\!/}\left(z'-z\right)\lambda(z')\mathrm{d}z'. \tag{27}$$

The total energy lost by the bunch is computed by summing the energy loss of all the particles:

$$U_{\mathrm{bunch}}=\frac{1}{q}\int_{-\infty}^{\infty}U\left(z'\right)\lambda\left(z'\right)\mathrm{d}z'=-q_0\int_{-\infty}^{\infty}W_{/\!/}\left(z'\right)\lambda\left(z'\right)\mathrm{d}z'. \tag{28}$$

3 Wakefield effects in linear accelerators

3.1 Energy spread

The longitudinal wake forces change the energy of individual particles depending on their position in the beam, as given by Eq. (26). As a consequence, the short-range wakefield can induce an energy spread in the beam.

For example, the energy spread induced by the space charge force in a Gaussian bunch is given by

$$\frac{\mathrm{d}U(z)}{\mathrm{d}s} = -q\int_{-\infty}^{\infty}\frac{\mathrm{d}w_{/\!/}\left(z'-z\right)}{\mathrm{d}s}\lambda(z')\mathrm{d}z' = \frac{qq_0}{4\pi\varepsilon_0\gamma^2\sqrt{2\pi}\sigma_z^3}\left(1+2\ln\frac{b}{a}\right)z\ \mathrm{e}^{-(z^2/2\sigma_z^2)}. \tag{29}$$

The bunch head gains energy ($z > 0$), and the tail loses energy. The total energy lost by the bunch, U_{bunch}, is zero.

In a similar way, one can show that the energy loss induced by a resonant HOM inside a rectangular uniform bunch of length l_0 when $\Gamma << \bar{\omega}$ is given by

$$U(z) = \frac{-qq_0R\omega_{\mathrm{r}}}{2Q}\frac{\sin\left[\dfrac{\omega_{\mathrm{r}}}{c}\left(\dfrac{l_0}{2}-z\right)\right]}{\left(\dfrac{\omega_{\mathrm{r}}l_0}{2c}\right)}, \tag{30}$$

and the total energy loss obtained with Eq. (28) is

$$U_{\mathrm{bunch}} = -\frac{2q_0^2Rc^2}{\omega_{\mathrm{r}}l_0^2Q}\sin^2\left(\frac{\omega_{\mathrm{r}}l_0}{2c}\right). \tag{31}$$

3.2 Single-bunch beam breakup: two-particle model

A beam injected off-centre in a linac, for example due to misalignments of the focusing quadrupoles, executes betatron oscillations. The bunch displacement produces a transverse wakefield in all the devices crossed during the flight, which deflects the trailing charges (single-bunch BBU), or other bunches following the first one in a multibunch regime (multibunch BBU). The first observation of BBU was made at SLAC in 1966 [19].

To understand the effect, we consider, as a first example, a simple model with only two charges, $q_1 = q_0/2$ (leading = half bunch) and $q_2 = q$ (trailing = single charge), travelling with $\beta = 1$.

The leading charge executes free betatron oscillations of the following kind:

$$y_1(s) = \hat{y}_1\cos\left(\frac{\omega_y}{c}s\right). \tag{32}$$

The trailing charge, a distance z behind, experiences over a length L_{w} an average deflecting force that is proportional to the displacement y_1 and dependent on the distance z. From the definition of the transverse dipole wakefield, this force is given by

$$\left\langle F_y(z,y_1)\right\rangle = \frac{qq_0}{2L_{\mathrm{w}}}w_{\perp}(z)y_1(s). \tag{33}$$

Note that L_w is the length of the device for which the transverse wake has been computed. For example, in the case of a cavity cell, L_w is the length of the cell. This force drives the motion of the trailing charge:

$$y_2' + \left(\frac{\omega_y}{c}\right)^2 y_2 = \frac{qq_0 w_\perp(z)}{2E_o L_w}\hat{y}_1 \cos\left(\frac{\omega_y}{c}s\right). \tag{34}$$

This is a typical equation for a resonator driven at the resonant frequency.

The solution is given by the superposition of the 'free' and the 'forced' oscillations, which, being driven at the resonant frequency, grow linearly with s, as shown in Fig. 7:

$$y_2(s) = \hat{y}_2 \cos\left(\frac{\omega_y}{c}s\right) + y_2^{\text{forced}}, \tag{35}$$

$$y_2^{\text{forced}} = \frac{cqq_0 w_\perp(z)s}{4\omega_y E_o L_w}\hat{y}_1 \sin\left(\frac{\omega_y}{c}s\right). \tag{36}$$

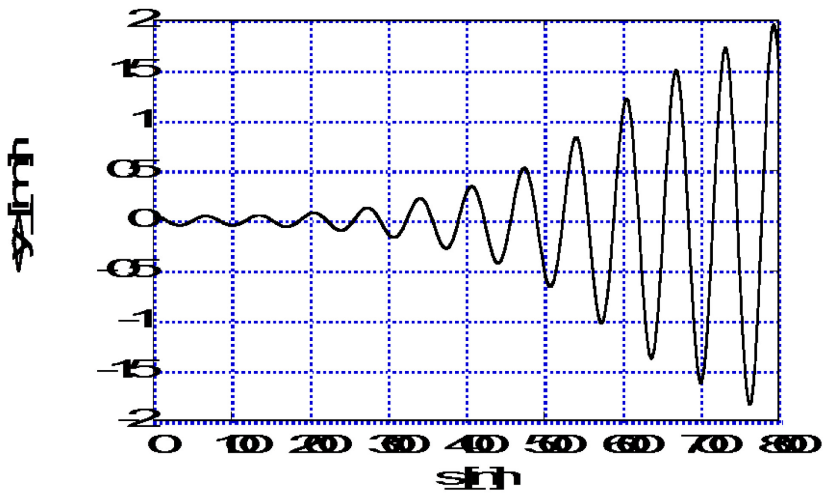


Fig. 7: HOMDYN [20] simulation of a typical BBU instability, 50 μm initial offset, no energy spread

At the end of a linac of length L_L, the oscillation amplitude will have grown by $(\hat{y}_1 = \hat{y}_2)$

$$\left(\frac{\Delta\hat{y}_2}{\hat{y}_2}\right)_{\max} = \frac{cNew_\perp(z)L_L}{4\omega_y(E_o/e)L_w}. \tag{37}$$

If the transverse wake is given per cell, the relative displacement of the tail with respect to the head of the bunch depends on the number of cells. Of course, it also depends on the focusing strength through the betatron frequency ω_y.

3.3 BNS damping

The BBU instability is quite harmful and hard to control even at high energy with strong focusing, and after careful injection and steering. A simple cure has been proposed after observing that the strong

oscillation amplitude of the bunch tail is mainly due to the 'resonant' driving head. If the tail and the head move with different frequencies, this effect can be significantly reduced [15].

Let us assume that the tail oscillates with a frequency $\omega_y + \Delta\omega_y$, so that Eq. (34) becomes

$$y_2' + \left(\frac{\omega_y + \Delta\omega_y}{c}\right)^2 y_2 = \frac{Ne^2 w_\perp(z)}{2E_o L_w}\hat{y}_1 \cos\left(\frac{\omega_y}{c}s\right), \tag{38}$$

the solution of which is given by

$$y_2(s) = \hat{y}_2 \cos\left(\frac{\omega_y + \Delta\omega_y}{c}s\right) - \frac{c^2 Ne^2 w_\perp(z)}{4\omega_y \Delta\omega_y E_o L_w}\hat{y}_1\left[\cos\left(\frac{\omega_y + \Delta\omega_y}{c}s\right) - \cos\left(\frac{\omega_y}{c}s\right)\right]. \tag{39}$$

In this case, we observe that the amplitude of the oscillation is limited and no longer grows linearly with s. Furthermore, by making a suitable choice for $\Delta\omega_y$, it is possible to suppress fully the oscillations of the tail. Indeed, by setting

$$\Delta\omega_y = \frac{c^2 Ne^2 w_\perp(z)}{4\omega_y E_o L_w} \tag{40}$$

if $\hat{y}_2 = \hat{y}_1$, from Eq. (39) we obtain

$$y_2(s) = \hat{y}_1 \cos\left(\frac{\omega_y}{c}s\right); \tag{41}$$

that is, the tail oscillates with the same amplitude as the head and with the same betatron frequency. This method of curing the single-bunch BBU instability is called BNS damping, after the names of the authors Balakin, Novokhatsky, and Smirnov who proposed it [15].

To have BNS damping, Eq. (40) imposes an extra focusing at the tail, which must have a higher betatron frequency than the head. This extra focusing can be obtained by: (1) using a Radiofrequency Quadrupole (RFQ), where the head and tail see a different focusing strength, (2) creating a correlated energy spread across the bunch, which, because of the chromaticity, induces a spread in the betatron frequency. An energy spread correlated with position is attainable with the external accelerating voltage or with wakefields.

In Fig. 8, we show the betatron oscillation corresponding to Fig. 7, but with a 2% energy spread.

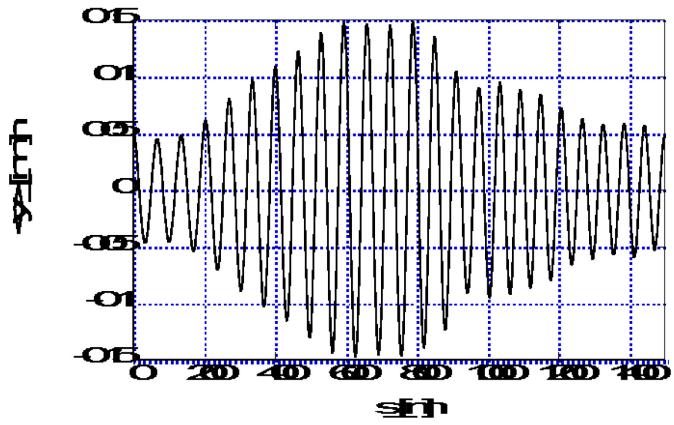


Fig. 8: HOMDYN simulation of a typical BNS damping, 50 μm initial offset, 2% energy spread

3.4 Single-bunch beam breakup: general distribution

To extend the analysis carried out in Section 3.2 to a particle distribution, we write the transverse equation of motion of a single charge q with the inclusion of the transverse wakefield effects as follows [14]:

$$\frac{\partial}{\partial s}\left[\gamma(s)\frac{\partial y(z,s)}{\partial s}\right]+k_y^2(s)\gamma(s)y(z,s)=\frac{q}{m_0c^2L_{\mathrm{w}}}\int_z^\infty y(s,z')w_\perp(z'-z)\lambda(z')\mathrm{d}z', \tag{42}$$

where $\gamma(s)$ is the relativistic parameter, which varies along the linac, and $1/k_y(s)$ is the betatron function. We recall that the integral of the longitudinal distribution function $\lambda(z)$ is the total charge of the bunch, q_0.

The solution of the equation in the general case is unknown. We can, however, apply a perturbation method to obtain the solution at any order in the wakefield intensity. Indeed, we write

$$y(z,s)=\sum_n y^{(n)}(z,s), \tag{43}$$

where n represents the nth-order solution. The first-order solution is found without the wakefield effect from the following equation:

$$\frac{\partial}{\partial s}\left[\gamma(s)\frac{\partial y^{(0)}(z,s)}{\partial s}\right]+k_y^2(s)\gamma(s)y^{(0)}(z,s)=0. \tag{44}$$

It is important to note that the above equation no longer depends on z. This means that the bunch distribution remains constant along the structure.

If the s-dependence of $\gamma(s)$ and $k_y^2(s)\gamma(s)$ is moderate, we can use the Wentzel–Kramers–Brillouin (WKB) approximation [5], and the solution of the above equation with the starting conditions $y(0)=\hat{y}$, $y'(0)=0$ is

$$y^{(0)}(s)=\sqrt{\frac{\gamma_0 k_{y0}}{\gamma(s)k_y(s)}}\hat{y}\cos\left[\psi(s)\right], \tag{45}$$

where

$$\psi(s)=\int_0^s k_y(s')\mathrm{d}s'. \tag{46}$$

Equation (45) represents the unperturbed transverse motion of the bunch in a linac.

The differential equation of the second-order solution is obtained by substituting the first-order solution (45) into the right-hand side of Eq. (42), yielding

$$\frac{\partial}{\partial s}\left[\gamma(s)\frac{\partial y^{(1)}(z,s)}{\partial s}\right]+k_y^2(s)\gamma(s)y^{(1)}(z,s)=\frac{q}{m_0c^2L_{\mathrm{w}}}y^{(0)}(s)\int_z^\infty w_\perp(z'-z)\lambda(z')\mathrm{d}z'. \tag{47}$$

We are interested in the forced solution of the above equation that can be written in the form

$$y^{(1)}(z,s)=\hat{y}\frac{q}{m_0c^2L_{\mathrm{w}}}\sqrt{\frac{\gamma_0 k_{y0}}{\gamma(s)k_y(s)}}G(s)\int_z^\infty w_\perp(z'-z)\lambda(z')\mathrm{d}z', \tag{48}$$

where

$$G(s) = \int_0^s \frac{1}{\gamma(s')k_y(s')} \sin\left[\psi(s) - \psi(s')\right] \cos\left[\psi(s')\right] \mathrm{d}s' = \frac{1}{2}\int_0^s \frac{\sin\left[\psi(s) - 2\psi(s')\right]}{\gamma(s')k_y(s')} \mathrm{d}s' + \frac{1}{2}\sin\left[\psi(s)\right] \int_0^s \frac{1}{\gamma(s')k_y(s')}\, \mathrm{d}s'. \tag{49}$$

The first integral undergoes several oscillations with s, and if $\gamma(s)$ and $k_y(s)$ do not vary much it is negligible, so that we can finally write

$$y^{(1)}(z,s) = \hat{y}\frac{q}{2m_0c^2L_\mathrm{w}}\sqrt{\frac{\gamma_0 k_{y0}}{\gamma(s)k_y(s)}}\sin\left[\psi(s)\right]\int_0^s \frac{\mathrm{d}s'}{\gamma(s')k_y(s')}\int_z^\infty w_\perp(z'-z)\lambda(z')\mathrm{d}z'. \tag{50}$$

Note that the last integral in Eq. (50) is proportional to the transverse wake potential produced by the whole bunch, defined in a similar way to Eq. (27). This solution can then be substituted again into the right-hand side of Eq. (42) to obtain a third-order solution, and so on. If we consider constant $\gamma(s)$ and $k_y(s)$, Eq. (50) gives the same result as for the two-particle model of Eq. (36) when we substitute $\lambda(z)$ with $q_0/2$ representing the leading half bunch affecting a trailing charge q.

If the BBU effect is strong, it is necessary to include higher-order terms in the perturbation expansion. Under the following assumptions:

i) rectangular bunch distribution $\lambda(z) = q_0 / l_0$, $-l_0/2 < z < l_0/2$, where l_0 is the bunch length;

ii) monoenergetic beam;

iii) constant acceleration gradient, $\mathrm{d}E_0/\mathrm{d}s$ = const.;

iv) constant beta function;

v) linear wake function inside the bunch, $w_\perp(z) = w_{\perp 0} z / l_0$,

the sum of Eq. (43) can be written in terms of powers of the adimensional parameter η, also called the BBU strength,

$$\eta = \frac{qq_0}{k_y\left(\mathrm{d}E_0/\mathrm{d}s\right)}\frac{w_{\perp 0}}{L_\mathrm{w}}\ln\left(\frac{\gamma_\mathrm{f}}{\gamma_\mathrm{i}}\right), \tag{51}$$

where γ_i and γ_f are the initial and final relativistic parameter, respectively.

By using the method of steepest descent [8], it is possible to obtain the asymptotic expression of $y(z,s)$ thus finding, at the end of the linac,

$$y(L_\mathrm{L}) = y_\mathrm{m}\sqrt{\frac{\gamma_\mathrm{i}}{6\pi\gamma_\mathrm{f}}}\eta^{-1/6}\exp\left[\frac{3\sqrt{3}}{4}\eta^{1/3}\right]\cos\left[k_y L_\mathrm{L} - \frac{3}{4}\eta^{1/3} + \frac{\pi}{12}\right], \tag{52}$$

which, unlike the two-particle model and that from the first-order solution, gives a tail displacement growing exponentially with η.

3.5 Multibunch beam breakup

We have seen in the previous sections that when a bunch passes off-axis (due, for example, to betatron oscillations) in an axis-symmetric accelerating structure, it excites transverse wakefields which may cause the tail of the bunch to oscillate with increasing amplitude as the bunch goes along the linac. In the same way, the whole bunch may excite deflecting trapped modes in the RF cavities of the linac that

may cause trailing bunches to be deflected, whether they are on-axis or not. These angular deflections are transformed into transverse displacements through the transfer matrices of the focusing system, and the displaced bunches will themselves create similar wakefields in the downstream accelerating structures of a linac. The subsequent bunches will be further deflected leading to beam blow-up. Due to the long-range wakefields, there is a coupling in the motion of the bunches that are increasingly deflected as they proceed along the linac in a process called multibunch BBU. Even if the bunches are not lost, the transverse beam emittance [illegible] can be greatly increased, leading to a significant luminosity reduction.

We summarize here the analytical study of multibunch BBU performed with the formalism used in [14]. All the bunches are considered to be rigid macro-particles, such as delta-functions, separated by period T, and we assume all bunches to be injected with the same initial offset x_0. We consider the transverse equation of motion of a bunch as a whole, ignoring internal structures; the beam is therefore made of a train of bunches with the same charge (Q_b) evenly spaced by period T, which is an integer number of the RF period of the accelerating mode.

We also consider all the cells of the linac accelerating structure to be identical and with the same dipole trapped mode in each cell of length L_w. Rigorously, the analytical approach requires that many betatron oscillations are performed in the linac, and the BBU remains moderate within a betatron oscillation. Moreover, the theory is valid if the beam energy does not change too much in a betatron wavelength. This last hypothesis is also called adiabatic acceleration.

The transverse wakefield force experienced by the kth bunch, spaced kT from the first bunch, depends on the transverse wakefield generated by the preceding bunches (and thus by their transverse displacement). The dipole long-range wakefield is produced by a high-order deflecting mode, identical in all the cavities of the structure, and it is described in terms of its resonant frequency ω_r, the quality factor Q, and the dipole shunt resistance $R_\perp$ (expressed in ohm per meter).

The equations of motion are then written in terms of the Z-transform (see, e.g., Ref. [21]) since the displacement $x(kT, s)$ of the kth bunch at the position s is a discrete function of time. The solution can be retrieved with a perturbation method, which considers its expansion as a series of the driving wakefield force.

The zeroth-order solution is given for a vanishing driving force, i.e. a pure betatron oscillation (unperturbed motion). It represents the motion of the first bunch, which is not affected by any wakefield because of the causality principle (the wakefield cannot travel ahead of the bunch itself). The nth-order solution is driven by the wakefield excited by the solution of the order $n-1$. Thus the first-order solution is computed from the motion of the first bunch, and it affects all the bunches, except the first one; it means that the nth-order solution affects only bunches of index larger then n. Therefore the summation of the series can be stopped at the Mth order of a train of M bunches. The nth-order solution in the Z-domain can be written as follows [14]:

$$x_n(z,s) = \sqrt{\frac{\gamma_0 k_{y0}}{\gamma(s) k_y(s)}}\, x_0 \mathrm{e}^{\mathrm{i}\psi(s)} \frac{a^n(s)}{\mathrm{i}^n n!} G_n(z), \tag{53}$$

where $a(s)$ is the so-called dimensionless BBU strength given, in case of constant $k_y(s)$, by

$$a(s) = \frac{Q_\mathrm{b}}{2k_{y0}G}\omega_\mathrm{r}\frac{R_\perp}{L_\mathrm{w}Q}\ln\left[\frac{\gamma(s)}{\gamma_0}\right], \tag{54}$$

where G is the accelerating gradient [V/m], and

$$G_n(z) = \frac{z}{z-1}\tilde{w}_\perp^n(z) \tag{55}$$

with

$$\tilde{w}_{\perp}\left(z\right)=\frac{1}{2\mathrm{i}}\left(\frac{z}{z-z_1}-\frac{z}{z-z_2}\right) \tag{56}$$

and

$$z_{1,2}=\mathrm{e}^{-\frac{T\omega_{\mathrm{r}}}{2Q}}\mathrm{e}^{\pm\mathrm{i}\omega_{\mathrm{r}}T}. \tag{57}$$

The inverse Z-transform of $x_n(z,s)$, i.e. $x_n(kT,s)$, can then be summed to get the transverse displacement of the kth bunch as

$$x\left(kT,s\right)=\sum_{n=0}^{\infty}x_n\left(kT,s\right). \tag{58}$$

We recall that the sum can be stopped at the Mth term for a beam containing M bunches.

For $a(s)\ll 1$ the series expansion can be stopped at the first-order term, although, if the BBU strength parameter a is moderate, it is sufficient to keep only a few terms of the summation.

In the Z-domain, the nth-order solution, given by Eq. (53), has been determined analytically, and the same is possible with its infinite sum, but its inverse Z-transform, Eq. (58), is, in general, not possible to write in a closed analytical form. It is, however, possible to compute the exact solution for the nth bunch as a sum of n terms if the BBU instability is moderate in a betatron period. Moreover, it is possible to use an asymptotic technique, valid when the blow-up is strong, to have an expression of the transverse displacement that puts in evidence the main parameters playing an important role in the instability.

The asymptotic transverse displacement of the kth bunch, expressed in terms of the oscillation amplitude only, is [14]

$$x\left(kT,s\right)=x_{\infty}\left(s\right)+\sqrt{\frac{\gamma_0 k_{y0}}{\gamma\left(s\right)k_y\left(s\right)}}\frac{x_0\mathrm{e}^{-k\frac{T\omega_{\mathrm{r}}}{2Q}}}{\sqrt{2\pi}\left(2e^{\frac{T\omega_{\mathrm{r}}}{2Q}}\right)^{1/2}\left[\cosh\left(\frac{T\omega_{\mathrm{r}}}{2Q}\right)-\cos\left(\omega_{\mathrm{r}}T\right)\right]^{1/2}}\frac{\left[2a(s)k\right]^{1/4}}{2k}\mathrm{e}^{\left[2a(s)k\right]^{1/2}}, \tag{59}$$

where $x_{\infty}(s)$ is the steady state solution that is reached when long (rigorously infinite) trains of bunches are accelerated.

In Fig. 9, we show a comparison between the analytical solution obtained by numerically solving Eq. (58) and a simple tracking code that considers the bunches in the train as rigid macro-particles, but which can also take into account the contribution of several resonant modes, along with different initial offsets and displacements of the bunches. The parameters used for the calculations are given in Table 1. They refer to a C-band linac with the BBU effect produced by a HOM. In the vertical axis the normalized transverse position, evaluated at the exit of the linac, is defined as follows:

$$\frac{x\left(kT,s\right)}{x_0}\sqrt{\frac{\gamma\left(s\right)k_y\left(s\right)}{\gamma_0 k_{y0}}}. \tag{60}$$

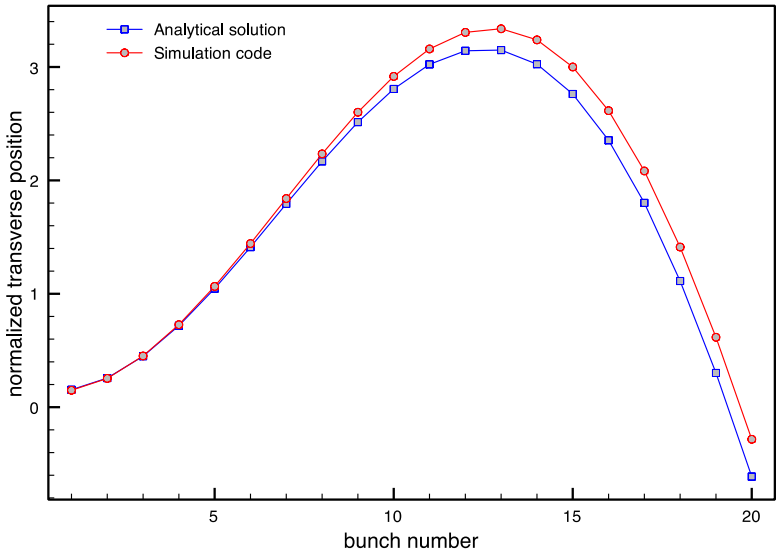


Fig. 9: Normalized transverse position as a function of the bunch number: comparison between the analytical solution and a tracking code.

Table 1: Beam parameters used for comparing the analytical solution of multibunch BBU with the results of a tracking code.

Linac length	30 m
Initial energy	80 MeV
Energy gradient	30 MeV/m
Betatron function, $1/k_y$	1 m
Bunch spacing, T	15 ns
Bunch charge	1 nC
HOM resonant frequency, f_r	8.4 GHz
HOM transverse impedance, $R_\perp$	50 MΩ/m
HOM quality factor	11 000
Cell length	17.5 cm

From Eq. (53), we see that one possible way to reduce the BBU instability is to act on the dimensionless BBU strength given by Eq. (54). For example, we can reduce the bunch charge Q_b or the betatron function, i.e. increase the focusing strength. A better approach is to remove the source of the instability by damping the transverse dipole mode, for example with an improved electromagnetic design of the accelerating cells.

The other main approach to BBU instability suppression is to detune the cell frequencies to introduce a spread in the resonance frequency of the dangerous mode so that it will no longer be excited coherently by the beam. Indeed, by properly detuning each cell, a damping of the BBU instability is produced via a decoherence of the various cell wakefields. It has been demonstrated [22] that a Gaussian distribution of the cell frequencies, which provides a rapid drop in the wakefield for a given total frequency spread, would be optimal. The analytical approach to determine the effectiveness of this detuning technique for the BBU multibunch instability can be found in Ref. [14], where it is also shown that the damping increases with the amplitude of the frequency spread.

References

[1] CST AG, Darmstadt, Germany, www.cst.com.

[2] W. Bruns, GdfidL: a finite difference program with reduced memory and CPU usage, 1997 Particle Accelerator Conf., Vancouver, B.C., Canada (1997), pp. 2651–2653.

[3] C. Ng *et al*., State of the art in EM field computation, Proc. EPAC 2006, Edinburgh, Scotland (2006), pp. 2763–2767.

[4] Y.H. Chin, KEK Report 2005-06, KEK, Tsukuba, Japan (2005).

[5] K. Yokoya, DESY 86-084 (1986).

[6] K. Bane, SLAC-PUB-4169 (1986).

[7] K. Bane and M. Sands, SLAC-PUB-4441 (1987).

[8] A. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators*, (John Wiley and Sons Inc., New York, 1993).

[9] L. Palumbo, V.G. Vaccaro, and M. Zobov, in Proceedings of the CAS-CERN Accelerator School: 5th Advanced Accelerator Physics Course, Rhodes, Greece, 20 September - 1 October 1993, CERN-1995-006 (CERN, Geneva, 1995), pp. 331-390. http://dx.doi.org/10.5170/CERN-1995-006.331

[10] B.W. Zotter and S.A. Kheifets, *Impedances and Wakes in High-Energy Particle Accelerators*, (World Scientific, Singapore, 1998).

[11] G.V. Stupakov, Wake and impedance, SLAC-PUB-8683 (2000).

[12] K. Bane, Emittance control for very short bunches, Proc. EPAC 2004, Lucerne, Switzerland (1982).

[13] M. Ferrario, M. Migliorati, and L. Palumbo, Space charge effects, these proceedings.

[14] A. Mosnier, in Proceedings of the CAS-CERN Accelerator School: 5th Advanced Accelerator Physics Course, Rhodes, Greece, 20 September - 1 October 1993, CERN-1995-006 (CERN, Geneva, 1995), pp. pp. 459–514, http://dx.doi.org/10.5170/CERN-1995-006.459

[15] V.E. Balakin, A.V. Novokhatsky, and V. P. Smirnov, Proc. 12th Int. Conf. on High Energy Accelerators, Batavia (1983).

[16] K. Bane, SLAC-PUB-9663 (2003).

[17] P.B. Wilson, SLAC-PUB-2884 (1982).

[18] W.K.H. Panofsky and W.A. Wenzel, *Rev. Sci. Instrum*. 27 (1956).

[19] H. Altenmueller *et al*., Beam break-up experiments at SLAC, 1966 Linear Accelerator Conf., Los Alamos; also SLAC-PUB-224 (Oct. 1966).

[20] M. Ferrario *et al*., Wake fields effects in the Sparc Photoinjector, Proc. EPAC 2004, Lucerne, Switzerland (2004).

[21] E.I. Jury, *Theory and Application of the Z-transform Method* (John Wiley & Sons, New York, 1964).

[22] K.L.F. Bane and R. L. Gluckstern, SLAC-PUB-5783 (1992).

[23] N.A. Vinokurov, in *High Quality Beams*, Eds. S.I. Kurokawa *et al*. (American Institute of Physics, 2001).

Appendix A: Power radiated by a bunch passing through a taper

In the case of a uniform charge distribution, and $\gamma \to \infty$, the electric field lines of a beam passing inside a perfectly conducting circular pipe are perpendicular to the direction of motion and travel together with the charge [9], as shown in Fig. A.1. In other words, the field map does not change during the charge flight, as long as the trajectory is parallel to the pipe axis. Under this condition, the transverse field's intensity can be computed as in the static case, applying Gauss's and Ampere's laws:

$$\int_S \varepsilon_0 \mathbf{E}\cdot\mathbf{n}\ \mathrm{d}S = \int_V \rho\ \mathrm{d}V, \qquad \oint \mathbf{B}\cdot\mathrm{d}\mathbf{l} = \mu_0 \int_S \mathbf{J}\cdot\mathbf{n}\ \mathrm{d}S \tag{A.1}$$

Let us consider a cylindrical beam of radius a and current I, with uniform charge density $\rho = I/\pi a^2 v$ and current density $J = I/\pi a^2$, propagating with relativistic speed $v = \beta c$ along the axis of a cylindrical perfectly conducting pipe of radius b, as shown in Fig. A.1.

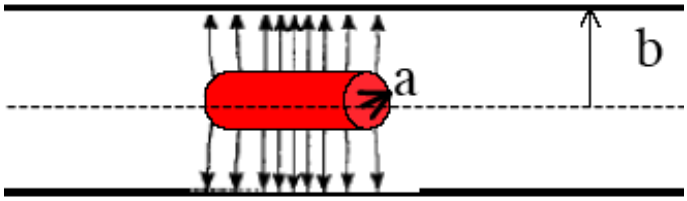


Fig. A.1: Cylindrical bunch of radius a propagating inside a cylindrical perfectly conducting pipe of radius b

By applying Eqs. (A.1), one obtains for the radial component of the electric field:

$$E_r = \frac{I}{2\pi\varepsilon_0 a^2 v} r \quad \text{for} \quad r \le a,$$

$$E_r = \frac{I}{2\pi\varepsilon_0 v}\frac{1}{r} \quad \text{for} \quad r > a,$$

and the relation $B_\vartheta = \dfrac{\beta}{c} E_r$ holds.

The electrostatic potential satisfying the boundary condition $\varphi(b) = 0$ is given by

$$\varphi(r,z) = \int_r^b E_r(r',z)\,\mathrm{d}r' = \begin{cases} \dfrac{I}{4\pi\varepsilon_0 v}\left(1 + 2\ln\dfrac{b}{a} - \dfrac{r^2}{a^2}\right) & \text{for} \quad r \le a \\ \dfrac{I}{2\pi\varepsilon_0 v}\ln\dfrac{b}{r} & \text{for} \quad a \le r \le b. \end{cases}$$

How can a perturbation of the boundary conditions affect the beam dynamics? Let us consider the following example: a smooth transition of length L (taper) from a beam pipe of radius b to a larger beam pipe of radius d is experienced by the beam [9]. To satisfy the boundary condition of a perfectly conducting pipe also in the tapered region, the field lines are bent, as shown in Fig. A.2. Therefore there must be a longitudinal electric field $E_z(r,z)$ in the transition region.

A test particle moving outside the beam charge distribution will experience along the transition of length L a voltage difference given by (e.g. see Ref. [21]):

$$V = -\int_z^{z+L} E_z(r,z')\,\mathrm{d}z' = -\left[\varphi(r,z+L) - \varphi(r,z)\right] = -\frac{I}{2\pi\varepsilon_0 v}\ln\frac{d}{b},$$

which is decelerating if $d > b$. The power lost by the beam to sustain the induced voltage is given by

$$P_{\text{lost}} = VI = \frac{I^2}{2\pi\varepsilon_0 v}\ln\frac{d}{b}. \tag{A.2}$$

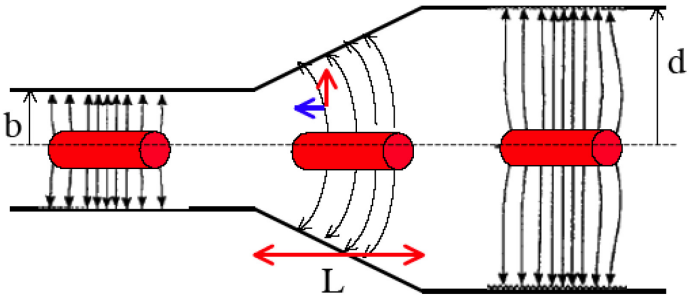


Fig. A.2: Smooth transition of length L (taper) from a beam pipe of radius b to a larger beam pipe of radius d

This means that for $d > b$ the power is deposited into the energy of the fields: moving from left to right of the transition, the beam induces the fields in the additional space around the bunch (i.e. in the region $b < r < d$, $0 < z < l_0$) at the expense of the only available energy source, which is the kinetic energy of the beam itself.

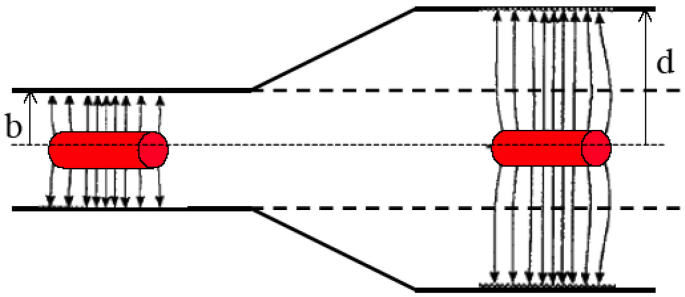


Fig. A.3: During the beam propagation in the taper, additional e.m. power flow is required to fill the new available space.

To verify this interpretation, let us now compute the e.m. power radiated by the beam to fill the additional space available around the bunch, as shown in Fig. A.3. On integrating the Poynting vector through the surface $\Delta S = \pi(d^2 - b^2)$ representing the additional power passing through the right part of the beam pipe, one obtains

$$P_{\text{em}} = \int_{\Delta S}\left(\frac{1}{\mu}\mathbf{E}\times\mathbf{B}\right)\cdot\mathbf{n}\,\mathrm{d}S = \int_b^d \frac{E_r B_\vartheta}{\mu} 2\pi r\,\mathrm{d}r = \frac{I^2}{2\pi\varepsilon_0 v}\ln\frac{d}{b},$$

which is exactly the same expression as in Eq. (A.2). Note that if $d < b$ the beam gains energy. If $d \to \infty$, the power goes to infinity. Such an unphysical result is nevertheless consistent with the original assumption of an infinite energy beam ($\gamma \to \infty$).

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam–Beam Effects

W. Herr and T. Pieloni
CERN, Geneva, Switzerland

Abstract
One of the most severe limitations in high-intensity particle colliders is the beam–beam interaction, i.e., the perturbation of the beams as they cross the opposing beams. This introduction to beam–beam effects concentrates on a description of the phenomena that are present in modern colliding beam facilities.

Keywords
Collider; beam-beam; non-linear dynamics; tune shift; Hamiltonian.

1 Introduction

The problem of the beam–beam interaction has been the subject of many studies since the introduction of the first particle colliders. It has been and will be one of the most important limits to performance and therefore attracts interest at the design stage of a new colliding beam facility. A particle beam is a collection of a large number of charges and represents an electromagnetic potential for other charges. It will therefore exert forces on itself and other beams. The forces are most important for high-density beams, i.e., high intensities and small beam sizes, which are the key to high luminosity.

The electromagnetic forces from particle beams are very non-linear and result in a wide spectrum of consequences for the beam dynamics. Furthermore, as a result of the interaction, the charge distribution creating the disturbing fields can change as well. This has to be taken into account in the evaluation of beam–beam effects and, in general, a self-consistent treatment is required.

Although we now have a good qualitative understanding of the various phenomena, a complete theory does not exist and exact predictions are still difficult. Numerical techniques, such as computer simulations, have been used with great success to improve the picture of some aspects of the beam–beam interaction, while for other problems the available models are not fully satisfactory in their predictive power.

2 Beam–beam force

In the rest frame of a beam, we have only electrostatic fields and, to find the forces on other moving charges, we have to transform the fields into the moving frame and to calculate the Lorentz forces (see Refs. [1–5] and references therein).

The fields are obtained by integrating over the charge distributions. The forces can be defocusing or focusing, since the test particle can have the same or opposite charge with respect to the beam producing the forces.

The distribution of particles producing the fields can follow various functions, leading to different fields and forces. It is not always possible to integrate the distribution to arrive at an analytical expression for the forces, in which case either an approximation or numerical methods have to be used. This is in particular true for hadron beams, which usually do not experience significant synchrotron radiation and damping. For e^-e^+ colliders, the distribution functions are most likely Gaussian with truncated tails.

In the two-dimensional case of a beam with bi-Gaussian beam density distributions in the transverse planes, i.e., $\rho(x, y) = \rho_x(x) \cdot \rho_y(y)$ with r.m.s. of σ_x and σ_y,

$$\rho_u(u) = \frac{1}{\sigma_u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma_u^2}\right), \text{ where } u = x, y\,, \tag{1}$$

https://doi.org/10.23730/CYRSP-2017-003.185

one can give the two-dimensional potential $U(x, y, \sigma_x, \sigma_y)$ as a closed expression:

$$U(x, y, \sigma_x, \sigma_y) = \frac{ne}{4\pi\epsilon_0} \int_0^\infty \frac{\exp(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q})}{\sqrt{(2\sigma_x^2+q)(2\sigma_y^2+q)}} \mathrm{d}q\,, \tag{2}$$

where n is the line density of particles in the beam, e the elementary charge, and ϵ_0 the permittivity of free space [6]. From the potential, one can derive the transverse fields $\vec{E}$ by taking the gradient $\vec{E} = -\nabla U(x, y, \sigma_x, \sigma_y)$.

2.1 Elliptical beams

For the case of bi-Gaussian distributions (i.e., elliptical beams with $\sigma_x \neq \sigma_y$), the fields can be derived and, for the case of $\sigma_x > \sigma_y$, we have [7]

$$E_x = \frac{ne}{2\epsilon_0\sqrt{2\pi(\sigma_x^2-\sigma_y^2)}} \mathrm{Im}\left[\mathrm{erf}\left(\frac{x+\mathrm{i}y}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right) - \mathrm{e}^{\left(-\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)} \mathrm{erf}\left(\frac{x\frac{\sigma_y}{\sigma_x}+\mathrm{i}y\frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right)\right], \tag{3}$$

$$E_y = \frac{ne}{2\epsilon_0\sqrt{2\pi(\sigma_x^2-\sigma_y^2)}} \mathrm{Re}\left[\mathrm{erf}\left(\frac{x+\mathrm{i}y}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right) - \mathrm{e}^{\left(-\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)} \mathrm{erf}\left(\frac{x\frac{\sigma_y}{\sigma_x}+\mathrm{i}y\frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right)\right]. \tag{4}$$

The function erf(t) is the complex error function:

$$\mathrm{erf}(t) = \mathrm{e}^{-t^2}\left[1 + \frac{2\mathrm{i}}{\sqrt{\pi}} \int_0^t \mathrm{e}^{z^2} \mathrm{d}z\right]. \tag{5}$$

The magnetic field components follow from

$$B_y = -\beta_r E_x/c \quad \text{and} \quad B_x = \beta_r E_y/c\,. \tag{6}$$

The Lorentz force acting on a particle with charge q is, finally,

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right). \tag{7}$$

2.2 Round beams

With the simplifying assumption of round beams ($\sigma_x = \sigma_y = \sigma$), one can re-write Eq. (7) in cylindrical co-ordinates:

$$\vec{F} = q\,(E_r + \beta c B_\Phi) \times \vec{r}\,. \tag{8}$$

From Eq. (2), and with $r^2 = x^2 + y^2$, one can immediately write the fields from Eq. (8) as

$$E_r = -\frac{ne}{4\pi\epsilon_0} \cdot \frac{\delta}{\delta r} \int_0^\infty \frac{\exp\left(-\frac{r^2}{(2\sigma^2+q)}\right)}{(2\sigma^2+q)} \mathrm{d}q\,, \tag{9}$$

and

$$B_\Phi = -\frac{ne\beta c\mu_0}{4\pi} \cdot \frac{\delta}{\delta r} \int_0^\infty \frac{\exp\left(-\frac{r^2}{(2\sigma^2+q)}\right)}{(2\sigma^2+q)} \mathrm{d}q\,. \tag{10}$$

We find from Eqs. (9) and (10) that the force of Eq. (8) has only a radial component. Equations (9) and (10) can easily be evaluated when the derivative is done first and $1/(2\sigma^2+q)$ is used as the integration variable. We can now express the radial force in a closed form (using $\epsilon_0\mu_0 = c^{-2}$):

$$F_r(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \cdot \frac{1}{r} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{11}$$

and, for the Cartesian components in the two transverse planes, we get

$$F_x(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \cdot \frac{x}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{12}$$

and

$$F_y(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \cdot \frac{y}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] . \tag{13}$$

The forces of Eqs. (12) and (13) are computed when the charges of the test particle and the opposing beam have opposite signs. For equally charged beams, the forces change sign. The shape of the force as a function of the amplitude is given in Fig. 1.

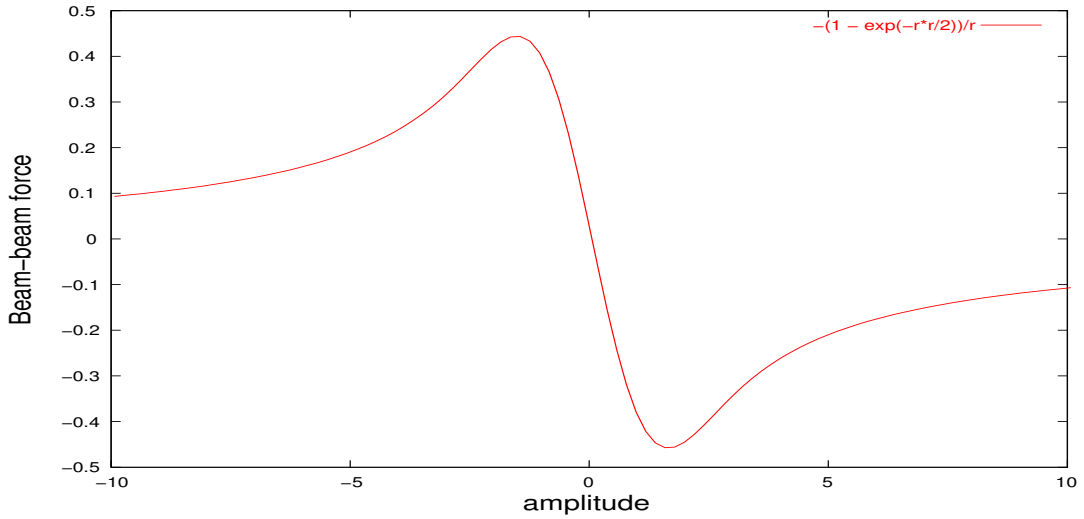


Fig. 1: Beam–beam force for round beams: force in arbitrary units, amplitude in units of r.m.s. beam size

For small amplitudes the force is approximately linear and a particle crossing a beam at small amplitudes will experience a linear field. This results in a change of the tune as in a quadrupole. At larger amplitudes (i.e. above $\approx 1\sigma$) the force deviates strongly from this linear behaviour. Particles at larger amplitudes will also experience a tune change; however, this tune change will depend on the amplitude. From the analytical form of Eq. (13), one can see that the beam–beam force includes higher multipoles.

3 Incoherent effects—single-particle effects

The force we have derived is the force of a beam on a single test particle. It can be used to study single-particle or incoherent effects. For that, we treat a particle crossing a beam as if it were moving through a static electromagnetic lens. A static field is an approximation that is valid as long as the beam–beam interaction is small compared with effects of other machine elements. We have to expect all effects that are known from resonance and non-linear theory, such as:

– unstable or irregular motion;
– beam blow-up or bad lifetime.

4 Beam–beam parameter

We can derive the linear tune shift of a small-amplitude particle crossing a round beam of a finite length. We use the force to calculate the kick it receives from the opposing beam, i.e., the change of the slope

of the particle trajectory. Starting from the two-dimensional force and multiplying by the longitudinal distribution, which depends on both position s and time t, and which we assume has a Gaussian shape with a width of σ_s, we obtain

$$F_r(r,s,t) = -\frac{Ne^2(1+\beta^2)}{\sqrt{(2\pi)^3}\epsilon_0\sigma_s}\cdot\frac{1}{r}\cdot\left[1-\exp\left(-\frac{r^2}{2\sigma^2}\right)\right]\cdot\left[\exp\left(-\frac{(s+vt)^2}{2\sigma_s^2}\right)\right]. \tag{14}$$

Here, N is the total number of particles. We make use of Newton's law and integrate over the collision to get the radial deflection:

$$\Delta r' = \frac{1}{mc\beta\gamma}\int_{-\infty}^{\infty} F_r(r,s,t)\mathrm{d}t\,. \tag{15}$$

The radial kick $\Delta r'$ that a particle with a radial distance r from the opposing beam centre receives is then

$$\Delta r' = -\frac{2Nr_0}{\gamma}\cdot\frac{1}{r}\cdot\left[1-\exp\left(-\frac{r^2}{2\sigma^2}\right)\right], \tag{16}$$

where we have re-written the constants and used the classical particle radius:

$$r_0 = e^2/4\pi\epsilon_0 mc^2\,, \tag{17}$$

where m is the mass of the particle. For small amplitudes r, one can derive the asymptotic limit:

$$\Delta r'|_{r\to 0} = -\frac{Nr_0 r}{\gamma\sigma^2} = -r\cdot f\,. \tag{18}$$

This limit is the slope of the force at $r = 0$ and the force becomes linear with a focal length as the proportionality factor.

It is well known how the focal length relates to a tune change and one can derive a quantity ξ, which is known as the *linear beam–beam parameter*:

$$\xi = \frac{Nr_0\beta^*}{4\pi\gamma\sigma^2}\,. \tag{19}$$

Here, r_0 is the classical particle radius (e.g., $r_\mathrm{e}, r_\mathrm{p}$) and β^* is the optical amplitude function (β-function) at the interaction point.

For small values of ξ and a tune far enough away from linear resonances, this parameter is equal to the linear tune shift ΔQ.

The beam–beam parameter can be generalized for the case of non-round beams and becomes

$$\xi_{x,y} = \frac{Nr_0\beta^*_{x,y}}{2\pi\gamma\sigma_{x,y}(\sigma_x+\sigma_y)}\,. \tag{20}$$

The beam–beam parameter is often used to quantify the strength of the beam–beam interaction; however, it does not reflect the non-linear nature.

5 Linear effects

For a small-amplitude linear force, like a quadrupole with focal length f

$$\frac{1}{f} = \frac{\Delta x'}{x} = \frac{Nr_0}{\gamma\sigma^2} = \left[\frac{\xi\cdot 4\pi}{\beta^*}\right]. \tag{21}$$

the transformation matrix over the interaction becomes (like a thin quadrupole):

$$\begin{pmatrix} 1 & 0 \\ \dfrac{1}{-f} & 1 \end{pmatrix}. \tag{22}$$

The full turn matrix including the tune shift ΔQ is computed from the unperturbed full turn matrix (see other lectures) plus interaction

$$\begin{pmatrix} \cos(2\pi(Q+\Delta Q)) & \beta^* \sin(2\pi(Q+\Delta Q)) \\ -\frac{1}{\beta^*}\sin(2\pi(Q+\Delta Q)) & \cos(2\pi(Q+\Delta Q)) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 \\ \frac{1}{-2f} & 1 \end{pmatrix} \circ \begin{pmatrix} \cos(2\pi Q) & \beta_0^* \sin(2\pi Q) \\ -\frac{1}{\beta_0^*}\sin(2\pi Q) & \cos(2\pi Q) \end{pmatrix} \circ \begin{pmatrix} 1 & 0 \\ \frac{1}{-2f} & 1 \end{pmatrix} . \quad (23)$$

Solving this equation gives

$$\cos(2\pi(Q+\Delta Q)) = \cos(2\pi Q) - \frac{\beta_0^*}{2f}\sin(2\pi Q)\,, \quad (24)$$

and

$$\frac{\beta^*}{\beta_0^*} = \sin(2\pi Q)/\sin(2\pi(Q+\Delta Q))\,. \quad (25)$$

The tune is changed by ΔQ and the β-function is also changed (β-beating) for large values of the tune change.

The tune change as a function of the unperturbed tune Q is shown in Fig. 2. For small ξ and Q, not too close to 0.0 and 0.5, we have:

$$\Delta Q \approx \xi\,, \quad (26)$$

and

$$\frac{\beta^*}{\beta_0^*} = \frac{\sin(2\pi Q)}{\sin(2\pi(Q+\Delta Q))} = \frac{\beta_0}{\sqrt{1+4\pi\xi\cot(2\pi Q) - 4\pi^2\xi^2}}\,. \quad (27)$$

Close to tune values of 0.0 and 0.5, the tune change can be significantly smaller or large than ξ, and β can become smaller or larger at the interaction point. This effect is called dynamic β.

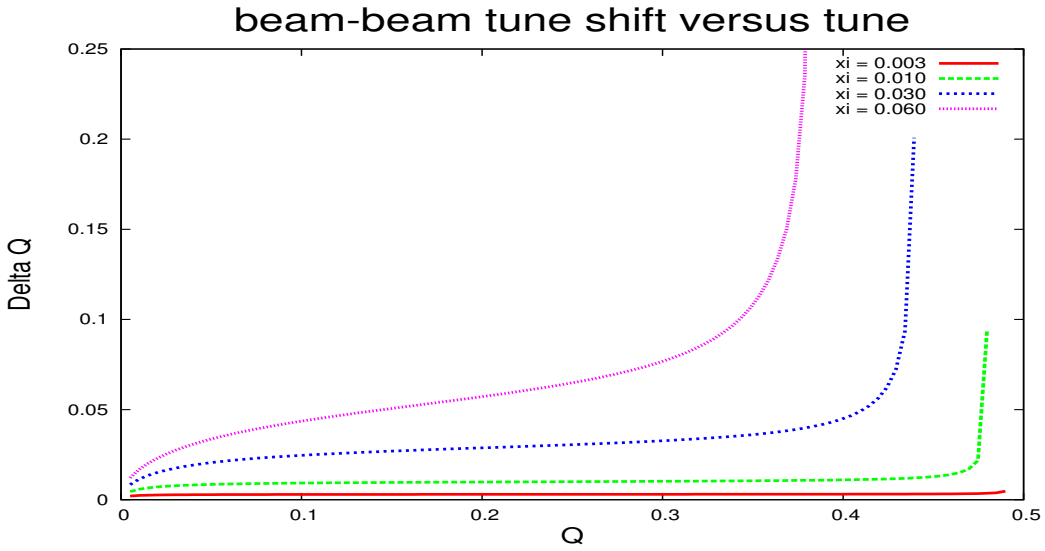


Fig. 2: Tune shift as function of tune and different beam–beam parameters

6 Non-linear effects

Since the beam–beam forces are strongly non-linear, the study of beam–beam effects encompasses the entire field of non-linear dynamics [8], as well as collective effects.

First, we briefly discuss the immediate effect of the non-linearity of the beam–beam force on a single particle. It manifests as an amplitude-dependent tune shift and, for a beam with many particles, as a tune spread. The instantaneous tune shift of a particle when it crosses the other beam is related to the derivative of the force with respect to the amplitude $\delta F/\delta x$. For a particle performing an oscillation with a given amplitude, the tune shift is calculated by averaging the slopes of the force over the range (i.e. the phases) of the particle's oscillation amplitudes.

The derivative of the beam–beam force from Fig. 1 is plotted for the one-dimensional case in Fig. 3.

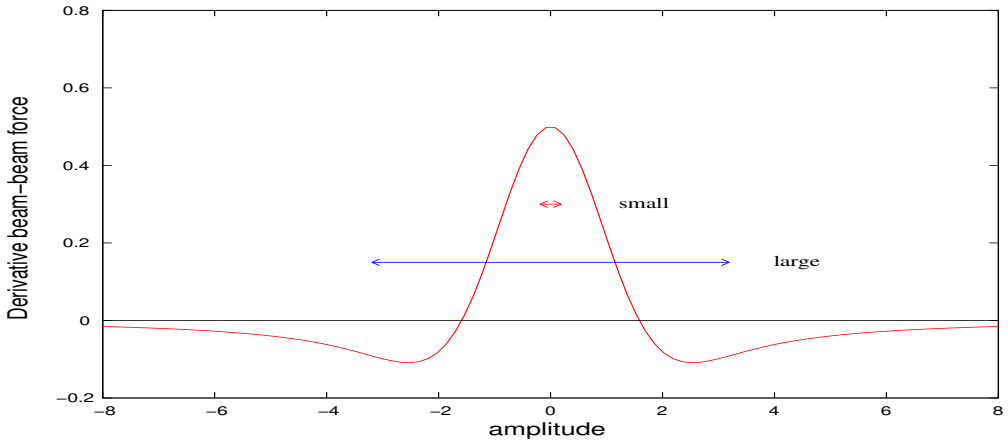


Fig. 3: Derivative of beam–beam force for round beams. Oscillation range of particles with large and small amplitudes.

The derivation of the detuning is given in Appendix B, using a classical method. An elegant calculation can be made using the Hamiltonian formalism [3] developed for non-linear dynamics, as discussed in Ref. [8] using the Lie formalism. It is demonstrated in Appendix C for one and two interaction points and includes the computation of the invariant Hamiltonian. We get the formula for the non-linear detuning with the amplitude J:

$$\Delta Q(J) = \xi \cdot \frac{2}{J} \cdot \left(1 - I_0(J/2) \cdot \mathrm{e}^{-J/2}\right) , \tag{28}$$

where $I_0(x)$ is the modified Bessel function and $J = \epsilon\beta/2\sigma^2$ in the usual units. Here,ϵ is the particle 'emittance' and not the beam emittance.

In the two-dimensional case, the tune shifts (ΔQ_x, ΔQ_y) of a particle with amplitudes x and y depend on both horizontal and vertical amplitudes. The detuning must be computed and presented in a two-dimensional form, i.e., the amplitude (x, y) is mapped into the tune space (Q_x, Q_y) or, alternatively, to the two-dimensional tune change (ΔQ_x, ΔQ_y). Such a presentation is usually called a 'tune footprint' and an example is shown in Fig. 4 (right-hand side); it maps the amplitudes into the tune space and each 'knot' of the mesh corresponds to a pair of amplitudes. Amplitudes between 0 and 6σ in both planes are used. The cross indicates the original, unperturbed tunes without the beam–beam interaction.

The maximum tune spread for a single head-on collision is equal to the tune shift of a particle with small amplitudes and, for small tune shifts, is equal to the beam–beam parameter ξ. In the simple case of a single head-on collision, the parameter ξ is therefore a measure for the tune spread in the beam.

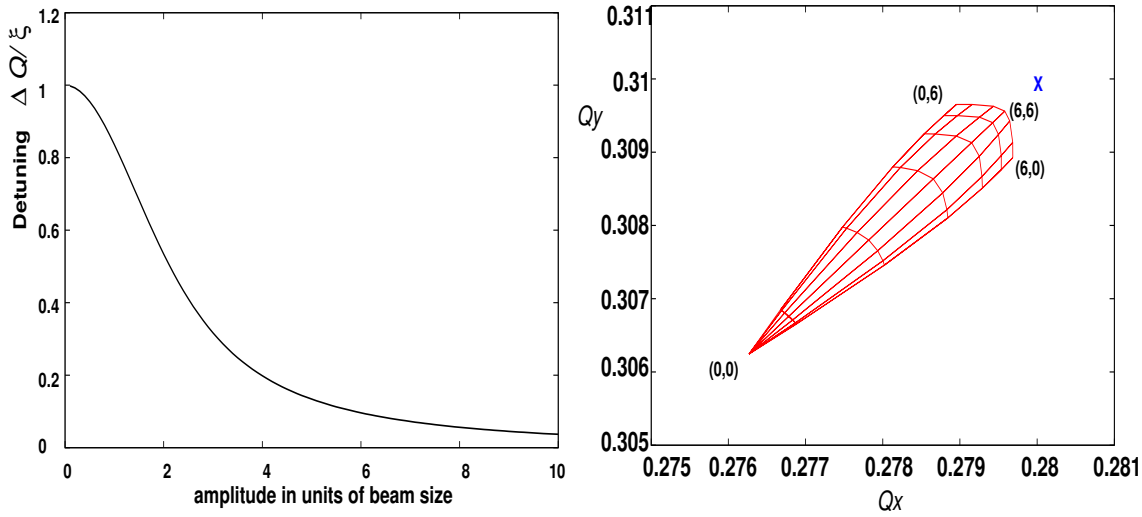


Fig. 4: Left-hand side: tune shift (non-linear detuning) as a function of the amplitude; right-hand side: two-dimensional tune footprint.

7 Beam stability

When the beam–beam interaction becomes too strong, the beam can become unstable or the beam dynamics is strongly distorted. One can distinguish different types of distortion; a few examples are:

- non-linear motion can become stochastic and can result in a reduction of the dynamic aperture and particle loss and bad lifetime;
- distortion of beam optics: dynamic beta (LEP) [2];
- vertical blow-up above the so-called beam–beam limit.

Since the beam–beam force is very non-linear, the motion can become 'chaotic'. This often leads to a reduction of the available dynamic aperture. The dynamic aperture is the maximum amplitude where the beam remains stable. Particles outside the dynamic aperture can eventually get lost. The dynamic aperture is usually evaluated by tracking particles with a computer program through the machine, where they experience the fields from the machine elements and other effects, such as wake fields or the beam–beam interaction.

As an example, we show such a study done for the LHC [9]. The dynamic aperture was evaluated for different configurations and different angles in the x–y plane. To find a good working point for the machine, the horizontal and vertical tunes have be varied and the dynamic aperture was computed for every step. The results of such a tune scan are shown in Fig. 5.

In the horizontal plane we find two regions where the dynamic aperture is reduced to significantly smaller values and we identify these with resonances near $q = 4/13$ and $q = 5/16$ (q is the fractional part of the tune).

Since the beam–beam interaction is basically a very non-linear lens in the machine, it distorts the optical properties and it may create a noticeable beating of the β-function around the whole machine and at the location of the beam–beam interaction itself. This can be approximated by inserting a quadrupole that produces the same tune shift at the position of the beam–beam interaction. The r.m.s. beam size at the collision point is now proportional to $\sqrt{\beta^*_{\rm p}}$, where $\beta^*_{\rm p}$ is the perturbed β-function, which can be significantly different from the unperturbed β-function β^*. This, in turn, changes the strength of the beam–beam interaction and the parameters have to be found in a self-consistent form. This is called the dynamic beta effect. This is a deviation from our assumption that the beams are static non-linear lenses.

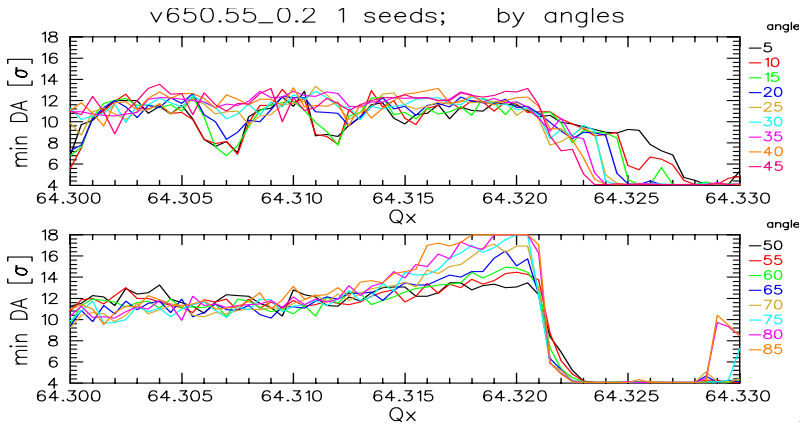


Fig. 5: Tune scan showing resonances. The dynamic aperture in units of the beam size is plotted as a function of the horizontal tune and for different angles in the x–y plane.

A strong dynamic beta effect was found in LEP [10], owing to its very large tune shift parameters. The β^* was decreased from 5 cm to $\beta^* = 2.5$ cm, leading to a substantial gain of luminosity [10].

Another effect that can be observed, in particular in e^+e^- colliders, is the blow-up of the emittance, which naturally limits the reachable beam–beam tune shifts.

8 Beam–beam limit

In e^+e^- colliders, the beam sizes are usually an equilibrium between the damping due to the synchrotron radiation and heating mechanisms, such as quantum excitation, intra-beam scattering and, very importantly, the beam–beam effect. This leads to a behaviour that is not observed in a hadron collider. When the luminosity is plotted as a function of the beam intensity, it should increase approximately as the current squared [11], in agreement with

$$\mathcal{L} = \frac{N^2 \cdot kf}{4\pi\sigma_x\sigma_y} \,. \tag{29}$$

Here, k is the number of bunches per beam and f the revolution frequency [11]. At the same time, the beam–beam parameter ξ should increase linearly with the beam intensity according to Eq. (20):

$$\xi_y = \frac{N \cdot r_e \beta_y}{2\pi\gamma\sigma_y(\sigma_x + \sigma_y)} \,. \tag{30}$$

In all e^+e^- colliders, the observation can be made that above a certain current, the luminosity increases approximately proportionally to the current, or at least much less than with the second power [12]. Another observation is that, at the same value of the intensity, the beam–beam parameter ξ saturates. This is shown for three e^+e^- colliders in Fig. 6 and is schematically illustrated in Fig. 7. This limiting value of ξ is commonly known as the *beam–beam limit*.

When we re-write the luminosity as

$$\mathcal{L} = \frac{N^2 \cdot kf}{4\pi\sigma_x\sigma_y} = \frac{N \cdot kf}{4\pi\sigma_x} \cdot \frac{N}{\sigma_y} \,, \tag{31}$$

we get an idea of what is happening. In e^+e^- colliders, the horizontal beam size σ_x is usually much larger than the vertical beam size σ_y and changes very little. For the luminosity to increase proportionally to the number of particles N, the factor N/σ_y must be constant.

This implies that with increasing current the vertical beam size increases in proportion above the beam–beam limit. This has been observed in all e^+e^- colliders and, since the vertical beam size is usually

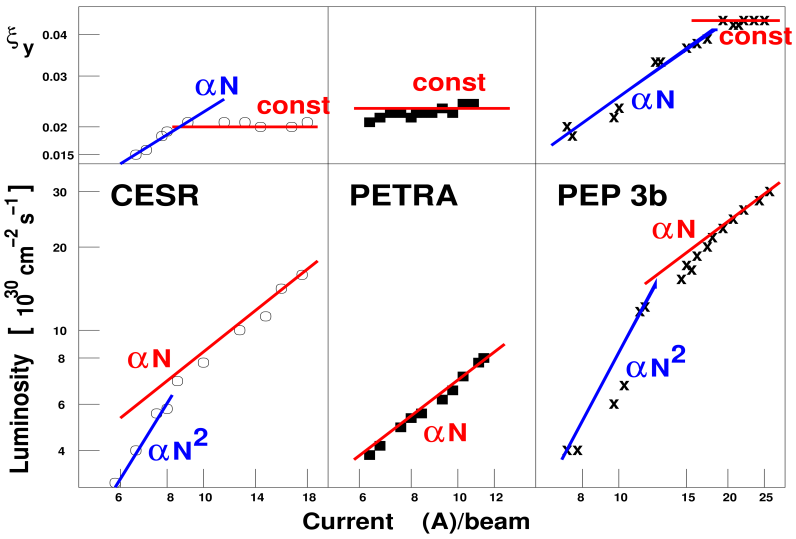


Fig. 6: Measurements of luminosity and beam–beam limit in e^+e^- colliders. Logarithmic scale of the axes to demonstrate change of exponent.

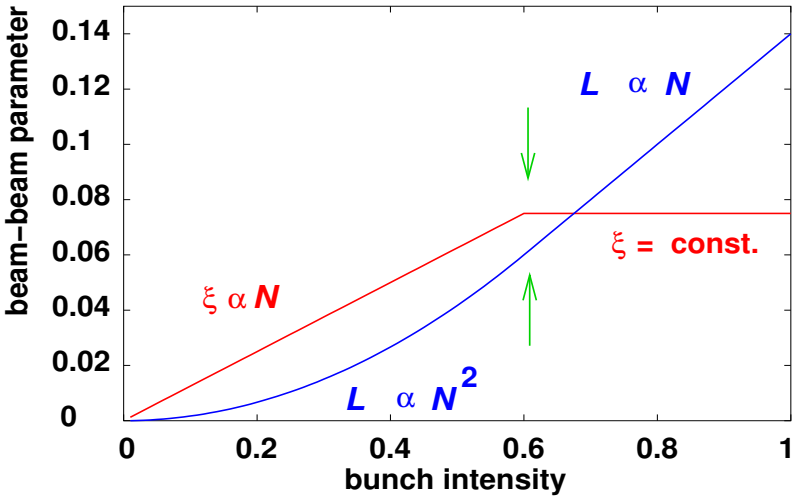


Fig. 7: Schematic illustration of beam–beam limit in e^+e^- colliders

small, this emittance growth can be very substantial before the lifetime of the beam is affected or beam losses are observed.

The dynamics of machines with high synchrotron radiation are dominated by the damping properties and the beam–beam limit is not a universal constant, nor can it be predicted. Simulation of beams with many particles can provide an idea of the order of magnitude [13, 14].

9 Crossing angle

To reach the highest luminosity, it is desirable to operate a collider with as many bunches as possible, since the luminosity is proportional to their number [11].

In a single-ring collider, such as the SPS, Tevatron, or LEP, the operation with k bunches leads to $2{\cdot}k$ collision points. When k is a large number, most of these are unwanted and must be avoided to reduce the perturbation due to the beam–beam effects.

Various schemes have been used to avoid these unwanted 'parasitic' interactions. Two prominent examples are shown in Fig. 8. In the SPS, Tevatron, and LEP, so-called Pretzel schemes were used. When the bunches are equidistant, this is the most promising method. When two beams of opposite charge travel in the same beam pipe, they can be moved onto separate orbits using electrostatic separators. In a well-defined configuration, the two beams cross when the beams are separated (Fig. 8, left-hand side). To avoid a separation around the whole machine, the bunches can be arranged in so-called trains of bunches, following each other closely. In that case a separation with electrostatic separators is only needed around the interaction regions. Such a scheme was used in the LEP in the second phase [15] and is schematically illustrated in Fig. 8 (right-hand side). Contrary to the majority of the colliders, the LHC collides particles

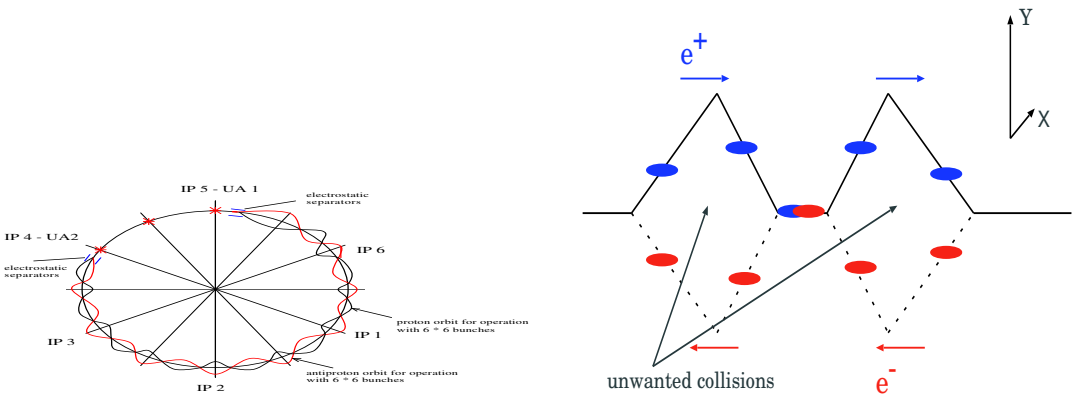


Fig. 8: Beam separation with: left-hand side, a Pretzel scheme (SPS, Tevatron, LEP); right-hand side, short bunch trains (LEP).

of the same type, which must therefore travel in separate beam pipes.

At the collision points of the LHC, the two beams are brought together and into collision (Fig. 9). An arrangement of separation and recombination magnets is used for the purpose of making the beams cross (Fig. 10).

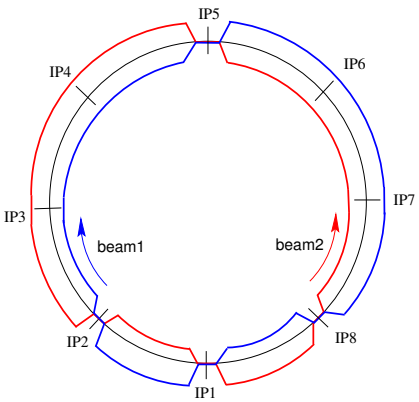


Fig. 9: Layout of the LHC collision points and beams

During this process it is unavoidable that the beams travel in a common vacuum chamber for more than 120 m. In the LHC, the time between the bunches is only 25 ns and therefore the bunches will meet in this region. To avoid the collisions, the bunches collide at a small crossing angle of 285 µrad.

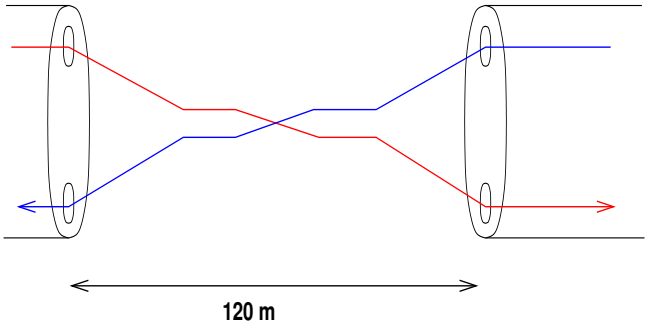


Fig. 10: Crossover between inner and outer vacuum chambers in the LHC

The basic principle is shown in Fig. 11: while two bunches collide at a small angle (quasi-head-on) at the centre, the other bunches are kept separated by the crossing angle. However, since they travel in a common beam pipe, the bunches still feel the electromagnetic forces from the bunches of the opposite beam. When the separation is large enough, these so-called long-range interactions should be weak. From the bunch spacing and the length of the interaction region, one can easily calculate that at each of the four LHC interaction points we must expect 30 of these long-range encounters, i.e. 120 interactions in total. The typical separation between the two beams is between 7 and 10 in units of the beam size of the opposing beam.

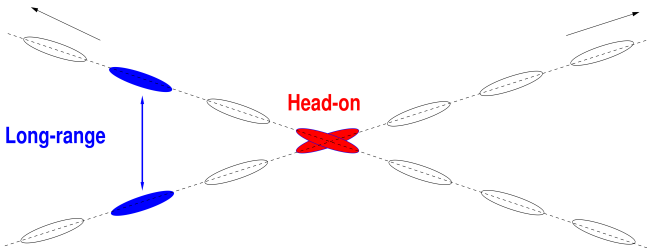


Fig. 11: Head-on and long-range interactions in a LHC interaction point

9.1 Long-range beam–beam effects

Although the long-range interactions distort the beams much less than a head-on interaction, their large number and some particular properties require careful studies.

- They break the symmetry between planes, i.e., odd resonances are also excited.
- While the effect of head-on collisions is strongest for small-amplitude particles, they mostly affect particles at large amplitudes.
- The tune shift caused by long-range interactions has the *opposite* sign in the plane of separation compared with the head-on tune shift.
- They cause changes of the closed orbit.
- They largely enhance the so-called PACMAN effects.

9.1.1 Opposite-sign tune shift

The opposite sign of the tune shift can easily be understood when we come back to the method for calculating the tune spread, explained with the help of Fig. 3.

We again average the oscillation of a small-amplitude particle as it samples the focusing force of the beam–beam interaction. Figure 12 shows the range of oscillation for central collisions and for the

interaction of separated beams, in both cases for particles with small oscillation amplitudes. When the separation is larger than $\approx 1.5\sigma$, the focusing (slope of the force as a function of the amplitude) changes sign and the resulting tune shift assumes the opposite sign.

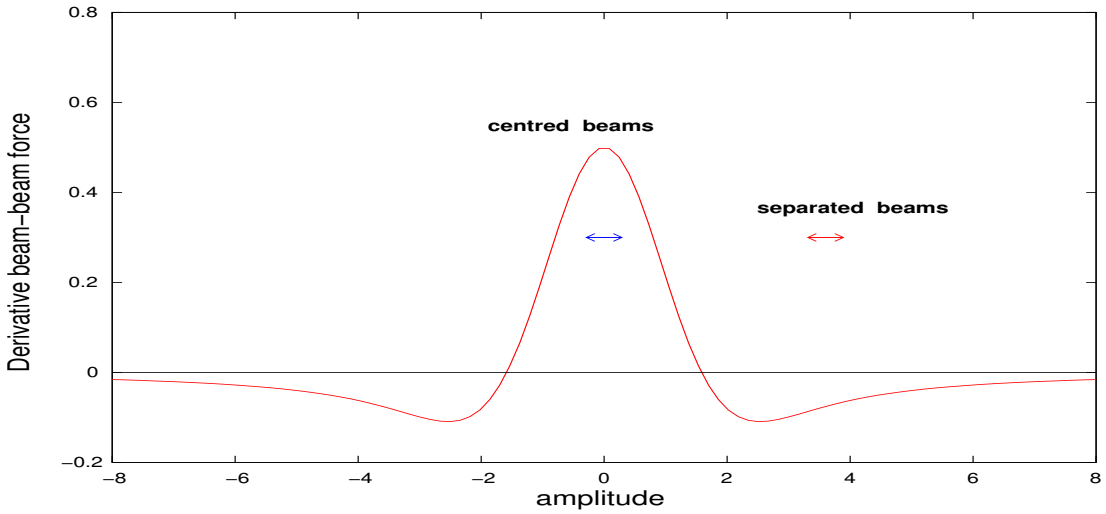


Fig. 12: Derivative of beam–beam force for round beams. Oscillation range of particles of centred and separated beams.

To some extent, this property could be used to partially compensate long-range interactions when a configuration is used where the beams are separated in the horizontal plane in one interaction region and in the vertical plane in another one. This is done in the LHC where the main experiments are exactly opposite in azimuth and therefore have the same interaction schedule. The crossing angles are vertical and horizontal in these two experiments.

9.1.2 Strength of long-range interactions

The geometry of a single encounter is shown in Fig. 13. The particles in the test bunch receive a kick (change of slope) $\Delta x'$.

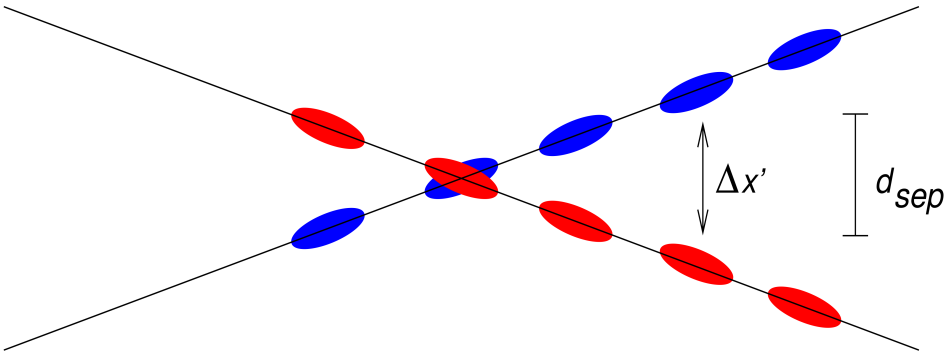


Fig. 13: Long-range interaction

Assuming a separation d in the horizontal plane, the kicks in the two planes can be written as

$$\Delta x' = -\frac{2Nr_0}{\gamma} \cdot \frac{(x+d)}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{32}$$

with $r^2 = (x+d)^2 + y^2$. The equivalent formula for the plane orthogonal to the separation is

$$\Delta y' = -\frac{2Nr_0}{\gamma} \cdot \frac{y}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] . \tag{33}$$

It is fairly obvious that the effect of long-range interactions must strongly depend on the separation. The calculation shows that the tune spread $\Delta Q_{\rm lr}$ from long-range interactions alone follows an approximate scaling (for large enough separation, i.e. above $\approx 6\sigma$):

$$\Delta Q_{\rm lr} \propto -\frac{N}{d^2} , \tag{34}$$

where N is the bunch intensity and d the separation. Small changes in the separation can therefore result in significant differences.

9.1.3 Footprint for long-range interactions

Contrary to the head-on interaction, where the small-amplitude particles are mostly affected, now the large-amplitude particles experience the strongest long-range beam–beam perturbations. This is rather intuitive, since the large-amplitude particles are the ones that can come closest to the opposing beam as they perform their oscillations. We must therefore expect a totally different tune footprint.

Such a footprint for only long-range interactions is shown in Fig. 14. Since the symmetry between the two planes is broken, the resulting footprint shows no symmetry. In fact, the tune shifts have different signs for x and y, as expected. For large amplitude, one observes a 'folding' of the footprint. This occurs when large-amplitude particles are considered, for which the oscillation amplitudes extend across the central maximum in Fig. 12, i.e., when the oscillation amplitude is larger than the separation between the beams.

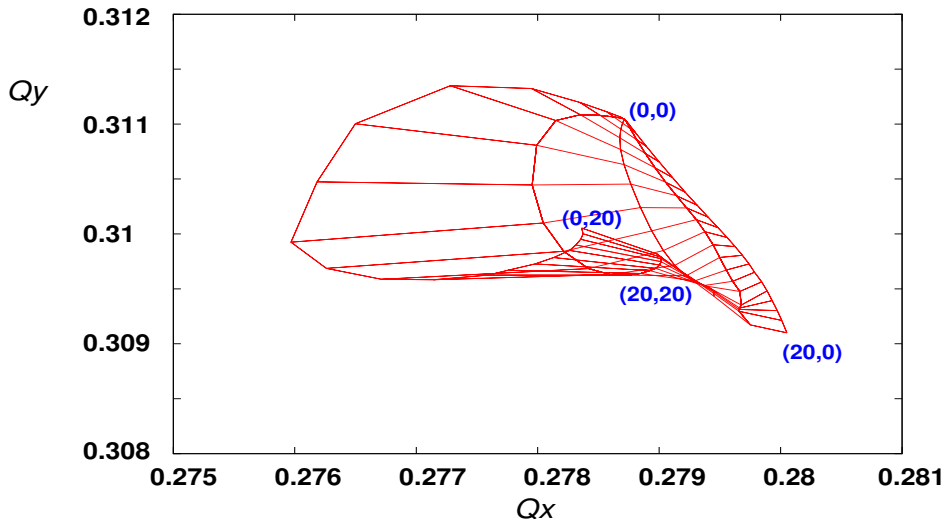


Fig. 14: Tune footprint for long-range interactions only. Vertical separation and amplitudes between 0 and 20 σ

Such large amplitudes are treated in Fig. 12 to demonstrate this feature. In practice, these amplitudes are usually not important since in real machines no particles reach these amplitudes.

In Fig. 15 (left), we show separately the footprints for two head-on proton–proton collisions, and long-range interactions with horizontal separation and vertical separation, respectively.

The differences, in particular the different sign of the long-range tune shift, are clearly visible. The long-range footprints are shifted away from the original tune (0.28, 0.31) in opposite directions.

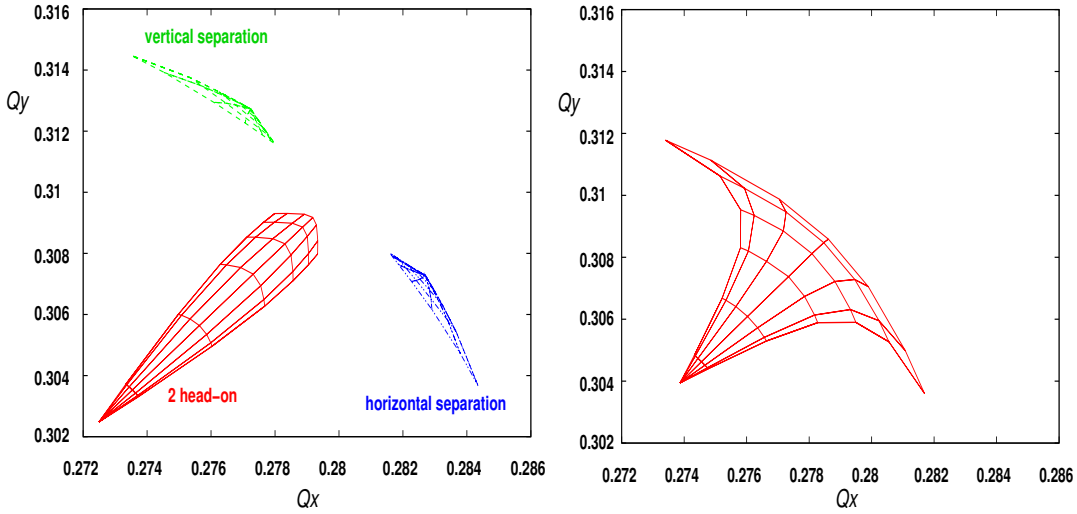


Fig. 15: Left-hand side: Footprint for two head-on interactions (proton–proton), long-range interactions in the horizontal plane, and long-range interactions in the vertical plane. Right-hand side: Combined head-on and long-range interactions, one horizontal and one vertical crossing.

Amplitudes between 0 and 6σ r are shown in the figure. In the same figure on the right, we show the combined footprint, i.e., for particles that experience two head-on collisions, long-range interactions in one interaction point with horizontal separation and a second with vertical separation. A partial compensation can be seen and the footprint is again symmetrical in x and y. In particular, the linear tune shifts of the central parts are very well compensated.

9.1.4 Dynamic aperture reduction due to long-range interactions

For too small separation, the tune spread induced by long-range interactions can become very large and resonances can no longer be avoided. The motion can become irregular and, as a result, particles at large amplitudes can get lost.

This is demonstrated in Fig. 16. The emittance increase of a large-amplitude particle (5σ) is computed for different values of the crossing angle and hence of the beam separation. For large enough angles, the emittance increase is very small, but it increases over several orders of magnitude when the crossing angle drops significantly below 300 µrad. These results are obtained by particle tracking [16,17].

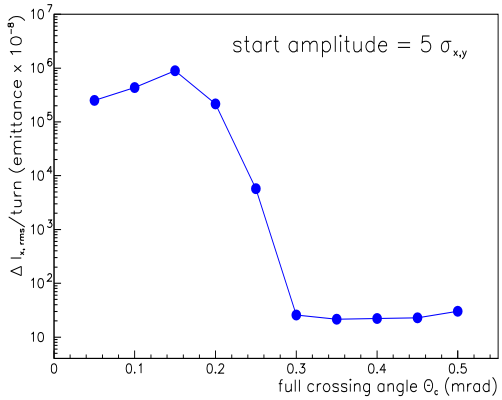


Fig. 16: Emittance increase as function of crossing angle

To evaluate the dynamic aperture in the presence of beam–beam interactions, a simulation of the complete machine is necessary and the interplay between the beam–beam perturbation and possible machine imperfections is important [18].

For the current LHC parameters, we consider the minimum crossing angle to be 285 µrad.

10 Beam–beam-induced orbit effects

When two beams do not collide exactly head-on, the force has a constant contribution, which can easily be seen when the kick $\Delta x'$ (from Eq. (32), for sufficiently large separation) is developed in a series:

$$\Delta x' = \frac{\text{const.}}{d} \cdot \left[1 - \frac{x}{d} + O\left(\frac{x^2}{d^2}\right) + \cdots\right]. \tag{35}$$

A constant contribution, i.e., more precisely an amplitude-independent contribution, changes the orbit of the bunch as a whole (Fig. 17).

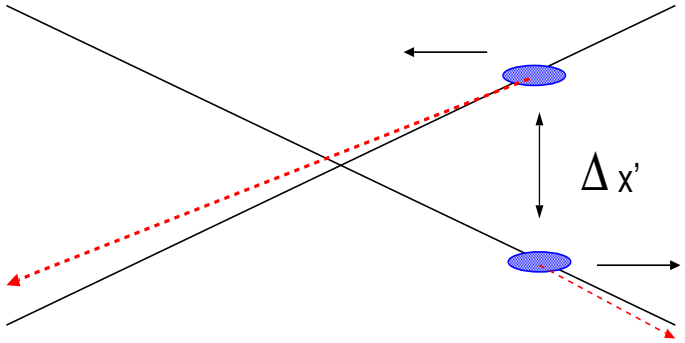


Fig. 17: Beam–beam deflection leading to orbit changes

When the beam–beam effect is strong enough, i.e., for high intensity or small separation, the orbit effects are large enough to be observed.

When the orbit of a beam changes, the separation between the beams will change as well, which will, in turn, lead to a slightly different beam–beam effect, and so on. The orbit effects must therefore be computed in a self-consistent way [19], in particular when the effects are sizable. The closed orbit of an accelerator can usually be corrected; however, an additional effect, which is present in some form in many colliders, sets a limit to the correction possibilities. A particularly important example is the LHC and we shall therefore use it to illustrate this feature.

11 PACMAN bunches

The bunches in the LHC do not form a continuous train of equidistant bunches spaced by 25 ns, but some empty space must be provided to allow for the rise time of kickers. These gaps and the number of bunches per batch are determined by requirements from the LHC injectors (PS, SPS, etc.) and the preparation of the LHC beam (bunch splitting). The whole LHC bunch pattern is composed of 39 smaller batches (trains of 72 bunches) separated by gaps of various length followed by a large abort gap for the dump kicker at the end. Figure 18 shows the actual LHC filling scheme with the various gaps in the train.

In the LHC, only 2808 out of 3564 possible bunches are present with this filling scheme. Owing to the symmetry, bunches normally meet other bunches at the head-on collision point. For the long-range interactions this is no longer the case.

This is illustrated in Fig. 19. Bunches at the beginning and end of a small batch will encounter a hole and, as a result, experience fewer long-range interactions than bunches from the middle of a

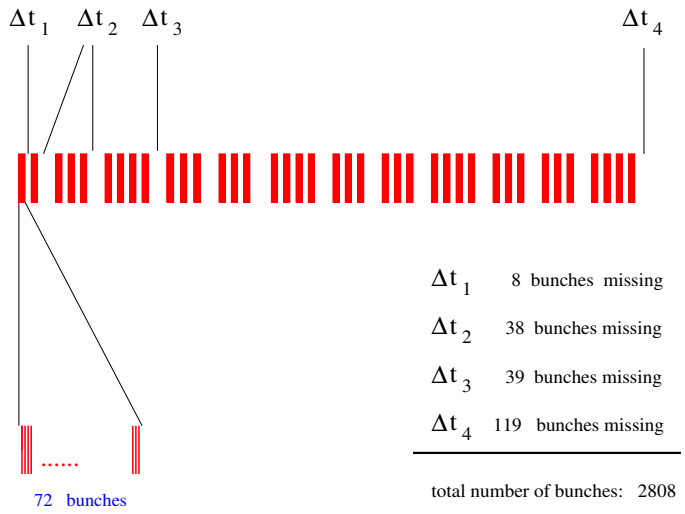


Fig. 18: Bunch filling scheme in the LHC

batch [20]. In the limit, the first bunch of a batch near a large gap encounters no opposing bunch before the central collision and the full number of bunches after.

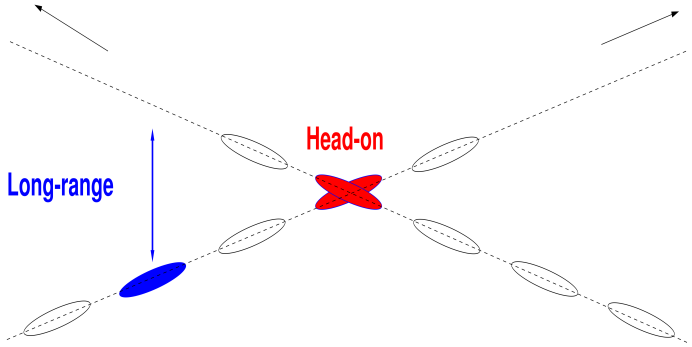


Fig. 19: Long-range interactions with missing bunches

Bunches with fewer long-range interactions have a very different integrated beam–beam effect and different dynamics must be expected. In particular, bunches will have a different tune and occupy a different area in the working diagram; therefore, they may be susceptible to resonances that can be avoided for nominal bunches. The overall space needed in the working diagram is therefore largely increased [20, 21].

Another consequence of reduced long-range interactions is the different effect on the closed orbit of the bunches. We have to expect a slightly different orbit from bunch to bunch.

This effect is demonstrated in Fig. 20, where we show the horizontal position at one head-on collision point for 432 bunches (out of 2808). The bunches in the middle of a batch have all interactions and therefore the same orbit while the bunches at the beginning and end of a batch show a structure that exhibits the decreasing number of long-range interactions. The orbit spread is approximately 10–15% of the beam size. Since the orbits of the two beams are not the same, it is impossible to make all bunches collide exactly head-on. A significant fraction will collide with an offset. Although the immediate effect on the luminosity is small [11], collisions at an offset can potentially affect the dynamics and are undesirable. The LHC design should try to minimize these offsets [20, 21].

A second effect, the different tunes of the bunches, is shown in Fig. 21. For three batches, it shows

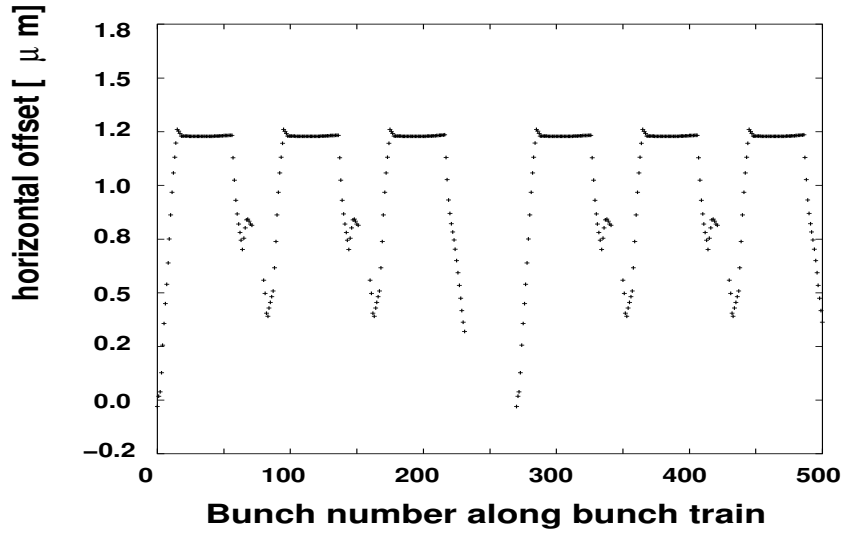


Fig. 20: Horizontal orbits of the first 432 bunches at IP1

a sizable spread from bunch to bunch; without compensation effects [21], it may be too large for safe operation. The tune spread between bunches is of the same order as the spread from the beam–beam effect and reduces the available tune space significantly.

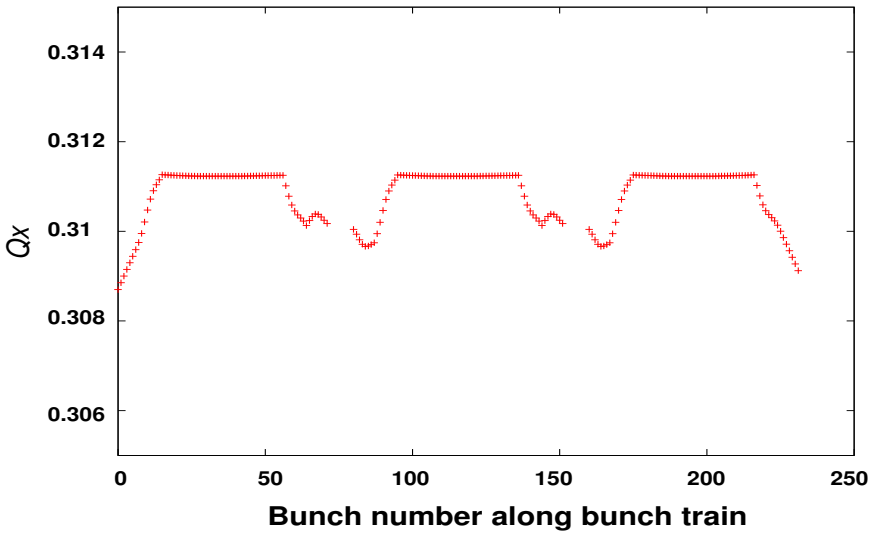


Fig. 21: Tune variation along the LHC bunch pattern

12 Compensation of beam–beam effects

For the case where the beam–beam effects limit the performance of a collider, several schemes have be proposed to compensate all or part of the detrimental effects. The basic principle is to design correction devices that act as non-linear 'lenses' to counteract the distortions from the non-linear beam–beam 'lens'.

For both head-on and long-range effects, schemes have been proposed:

- head-on effects:
  - electron lenses;
  - linear lens to shift tunes;

  - non-linear lens to decrease tune spread.
- long-range effects:
  - at large distance: beam–beam force changes like $1/r$;
  - same force as a wire.

12.1 Electron lenses

The basic principle of a compensation of proton–proton (or proton–antiproton) collisions with an 'electron lens' implies that the proton (antiproton) beam travels through a counter-rotating high-current electron beam ('electron lens') [22, 23]. The negative electron space charge can reduce the effect from the collision with the other proton beam.

An electron beam with a size much larger than the proton beam can be used to shift the tune of the proton beam ('linear lens'). When the current in the electron bunches can be varied quickly enough, the tune shift can be different for the different proton bunches, thus correcting PACMAN tune shifts.

When the electron charge distribution is chosen to be the same as the counter-rotating proton beam, the non-linear focusing of this proton beam can be compensated ('non-linear lens'). When it is correctly applied, the tune spread in the beam can be strongly reduced.

Such lenses have been constructed at the Tevatron at Fermilab [23] and experiments are in progress.

12.2 Electrostatic wire

To compensate the tune spread from long-range interactions, one needs a non-linear lens that resembles a separated beam. At large enough separation, the long-range force changes approximately with $1/r$ and this can be simulated by a wire parallel to the beam [24].

To compensate PACMAN effects, the wires have to be pulsed according to the bunch filling scheme. Tests are in progress at the SPS to study the feasibility of such a compensation for the LHC.

12.3 Möbius scheme

The beam profiles of e^+e^- colliders are usually flat, i.e., the vertical beam size is much smaller than the horizontal beam size. Some studies indicate that the collision of round beams, even for e^+e^- colliders, shows more promise for higher luminosity, since larger beam–beam parameters can be achieved. Round beams can always be produced by strong coupling between horizontal and vertical planes. A more elegant way is the so-called Möbius lattice [25,26]. In this lattice, the horizontal and vertical betatron oscillations are exchanged by an insertion. A horizontal oscillation in one turn becomes a vertical oscillation in the next turn and vice versa. Tests with such a scheme have been made at CESR at Cornell [26].

References

[1] W. Herr, Beam–beam effects, Proc. CERN Accelerator School, Zeuthen, 2003, CERN-2006-002 (2006), p. 379.

[2] A. Chao, The beam–beam instability, SLAC-PUB-3179 (1983).

[3] L. Evans and J. Gareyte, Beam–beam effects, CERN Accelerator School, Oxford, 1985, CERN 87-03 (1987), p. 159.

[4] A. Zholents, Beam–beam effects in electron–positron storage rings, Joint US–CERN School on Particle Accelerators (Lecture Notes in Physics No. 400, Springer, Heidelberg, 1992), p. 321.

[5] E. Keil, Beam–beam dynamics, CERN Accelerator School, Rhodes, 1993, CERN 95-06 (1995), p. 539.

[6] S. Kheifets, PETRA-Kurzmitteilung 119, DESY (1976).

[7] M. Basetti and G.A. Erskine, Closed expression for the electrical field of a two-dimensional Gaussian charge, CERN-ISR-TH/80-06 (1980).

[8] W. Herr, Mathematical and numerical methods for non-linear beam dynamics, Proc. CERN Accelerator School, Advanced Accelerator Physics Course, Trondheim, 2013, CERN-2014-009 (2014), p. 157.

[9] W. Herr and D. Kaltchev, Effect of phase advance between interaction points in the LHC on the beam–beam interaction, CERN LHC Project 1082 (2008).

[10] D. Brandt *et al.*, Is LEP beam–beam limited at its highest energy?, Proc. Particle Accelerator Conf., New York, 1999, p. 3005. http://dx.doi.org/10.1109/pac.1999.792127

[11] W. Herr, Concept of luminosity, Proc. CERN Accelerator School, Zeuthen, 2003, CERN-2006-002 (2006), p. 361.

[12] J.T. Seeman, Observations of the beam–beam interaction (Lecture Notes in Physics No. 247, Springer, Heidelberg, 1986), p. 121. http://dx.doi.org/10.1007/bfb0107349

[13] S. Myers, *IEEE Trans. Nucl. Sci.* **NS-28** (1981) 2503. http://dx.doi.org/10.1109/TNS.1981.4331737

[14] S. Myers, Review of beam–beam simulations (Lecture Notes in Physics No. 247, Springer, Heidelberg, 1986), p. 176. http://dx.doi.org/10.1007/bfb0107351

[15] B. Goddard *et al.*, *Part. Accel.* **57** (1998) 237.

[16] T. Sen *et al.*, Effect of the beam–beam interactions on the dynamic aperture and amplitude growth in the LHC, Proc. Workshop on Beam–Beam Effects in Large Hadron Colliders, LHC99, Geneva, 12–17 April 1999, CERN-SL-99-039 (AP) (1999), p. 85.

[17] Y. Papaphilippou and F. Zimmermann, Weak–strong beam–beam simulations for the LHC, Proc. Workshop on Beam–Beam Effects in Large Hadron Colliders, LHC99, Geneva, 12–17 April 1999, CERN-SL-99-039 (AP) (1999), p. 95.

[18] Y. Luo and F. Schmidt, Dynamic aperture studies for LHC optics version 6.2 at collision, CERN LHC Project Note 310 (2003).

[19] H. Grote and W. Herr, Self-consistent orbits with beam–beam effects in the LHC, Proc. 2001 Workshop on Beam–Beam Effects, Fermilab, Chicago, IL, 25–27 June 2001.

[20] W. Herr, Effects of PACMAN bunches in the LHC, CERN LHC Project Report 39 (1996).

[21] W. Herr, Features and implications of different LHC crossing schemes, LHC Project Report 628 (2003).

[22] V.D. Shiltsev *et al.*, Compensation of beam–beam effects in the Tevatron collider with electron beams, Proc. Particle Accelerator Conf., New York, 1999, p. 3728. http://dx.doi.org/10.1109/pac.1999.792426

[23] V.D. Shiltsev *et al.*, *Phys. Rev. ST Accel. Beams* **2** (1999) 071001. http://dx.doi.org/10.1103/PhysRevSTAB.2.071001

[24] J.P. Koutchouk *et al.*, Correction of the long-range beam–beam effect in LHC using electromagnetic lenses, Proc. Particle Accelerator Conf., New York, 1999, p. 1681.

[25] R. Talman, *Phys. Rev. Lett.* **74** (1995) 1590. http://dx.doi.org/10.1103/PhysRevLett.74.1590

[26] S. Henderson *et al.*, Investigation of the Möbius accelerator at CESR, CBN 99-5 (1999).

[27] A.G. Webster, *Partial Differential Equations of Mathematical Physics* (Hafner, New York, 1950).

[28] I. Bronstein *et al.*, *Taschenbuch der Mathematik* (Harri Deutsch, Frankfurt, 2000).

[29] O.D. Kellog, *Foundations of Potential Theory* (Dover, New York, 1953).

[30] L. Evans, The beam–beam interaction, CERN Accelerator Physics Course on Proton–Antiproton Colliders, CERN 84-15 (1984), p.319.

[31] O. Brüning, Non-linear imperfections, these proceedings.

[32] A. Dragt, *IEEE Trans. Nucl. Sci.* **NS-26** (1979) 3601.
http://dx.doi.org/10.1109/TNS.1979.4330114

[33] A. Dragt, Analysis of the beam–beam interaction using transfer maps, Proc. Beam–Beam Interaction Seminar, SLAC 1980, SLAC-PUB-2624 (1980).

[34] E. Forest, *Beam Dynamics, A New Attitude and Framework* (Harwood Academic, Berkeley, CA, 1998).

[35] A. Chao, *et al.*, Non-linear beam–beam resonances (Lecture Notes in Physics No. 247, Springer, Heidelberg, 1986), p. 77.

[36] A. Chao, Lecture notes on topics in accelerator physics, SLAC (2001).

[37] D. Kaltchev, On beam–beam resonances observed in LHC tracking, TRI-DN-07-9 (TRIUMF, Vancouver, Canada, 2007).

Appendices

A Appendix: Potential of the beam-beam force for a Gaussian beam

In practice, one usually derives the potential $U(x, y, z)$ from the Poisson equation, which relates the potential to the charge-density distribution $\rho(x, y, z)$:

$$\Delta U = -\frac{1}{\epsilon_0}\rho(x, y, z) \tag{A.1}$$

and computes the fields from

$$\vec{E} = -\nabla U(x, y, \sigma_x, \sigma_y)\,. \tag{A.2}$$

The Poisson equation can be solved using e.g., the Green's function method (e.g., Ref. [27]) since the Green's function for this boundary value problem is well known. The formal solution using a Green's function $G(x, y, z, x_0, y_0, z_0)$ is

$$U(x, y, z) = \frac{1}{\epsilon_0}\int G(x, y, z, x_0, y_0, z_0)\cdot\rho(x_0, y_0, z_0)\mathrm{d}x_0\mathrm{d}y_0\mathrm{d}z_0\,. \tag{A.3}$$

For the solution of the Poisson equation, we get [28]

$$U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi\epsilon_0}\int\int\int \frac{\rho(x_0, y_0, z_0)\mathrm{d}x_0\mathrm{d}y_0\mathrm{d}z_0}{\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}}\,. \tag{A.4}$$

In the case of a beam with Gaussian beam density distributions, we can factorize the density distribution $\rho(x_0, y_0, z_0) = \rho(x_0)\cdot\rho(y_0)\cdot\rho(z_0)$ with r.m.s. of σ_x, σ_y and σ_z:

$$\rho(x_0, y_0, z_0) = \frac{Ne}{\sigma_x\sigma_y\sigma_z(\sqrt{2\pi})^3}\mathrm{e}^{\left(-\frac{x_0^2}{2\sigma_x^2}-\frac{y_0^2}{2\sigma_y^2}-\frac{z_0^2}{2\sigma_z^2}\right)}\,. \tag{A.5}$$

Here, N is the number of particles in the bunch. We therefore have

$$U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi\epsilon_0}\frac{Ne}{\sigma_x\sigma_y\sigma_z(\sqrt{2\pi})^3}\int\int\int \frac{\mathrm{e}^{\left(-\frac{x_0^2}{2\sigma_x^2}-\frac{y_0^2}{2\sigma_y^2}-\frac{z_0^2}{2\sigma_z^2}\right)}\mathrm{d}x_0\mathrm{d}y_0\mathrm{d}z_0}{\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}}\,. \tag{A.6}$$

This is difficult to solve; we would rather follow the proposal by Kheifets [6] to solve the diffusion equation

$$\Delta V - A^2\cdot\frac{\delta V}{\delta t} = -\frac{1}{\epsilon_0}\rho(x, y, z) \qquad (\text{for } t \geq 0) \tag{A.7}$$

and obtain the potential U by going to the limit $A \rightarrow 0$, i.e.,

$$U = \lim_{A \to 0} V \,. \tag{A.8}$$

The reason for this manipulation is that the Green's function to solve the diffusion equation takes a more appropriate form [28]:

$$G(x, y, z, t, x_0, y_0, z_0) = \frac{A^3}{(2\sqrt{\pi t})^3} \cdot \mathrm{e}^{-A^2/4t \cdot ((x-x_0)^2+(y-y_0)^2+(z-z_0)^2)} \tag{A.9}$$

and we get for $V(x, y, z, \sigma_x, \sigma_y, \sigma_z)$:

$$\frac{Ne}{\sigma_x \sigma_y \sigma_z (\sqrt{2\pi})^3 \epsilon_0} \int_0^t \mathrm{d}\tau \int\int\int \mathrm{e}^{\left(-\frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2} - \frac{z_0^2}{2\sigma_z^2}\right)} \frac{A^3 \cdot \mathrm{e}^{-A^2/4\tau((x-x_0)^2+(y-y_0)^2+(z-z_0)^2)}}{(2\sqrt{\pi\tau})^3} \mathrm{d}x_0 \mathrm{d}y_0 \mathrm{d}z_0 \,. \tag{A.10}$$

This allows us to avoid the denominator in the integral and to collect the exponential expressions, which can then be integrated. Changing the independent variable τ to $q = 4\tau/A^2$ and using the formula [28] for the three integrations,

$$\int_{-\infty}^{\infty} \mathrm{e}^{-(au^2+2bu+c)} \mathrm{d}u \;=\; \sqrt{\frac{\pi}{a}} \mathrm{e}^{(\frac{b^2-ac}{a})} \qquad (\text{for} \quad u = x_0, y_0, z_0) \,, \tag{A.11}$$

we can integrate Eq. (A.3) and with Eq. (A.8) we get the potential $U(x, y, z, \sigma_x, \sigma_y, \sigma_z)$ [6, 29]:

$$U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi\epsilon_0} \frac{Ne}{\sqrt{\pi}} \int_0^\infty \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q} - \frac{z^2}{2\sigma_z^2+q}\right)}{\sqrt{(2\sigma_x^2+q)(2\sigma_y^2+q)(2\sigma_z^2+q)}} \mathrm{d}q \,. \tag{A.12}$$

Since we are interested in the transverse fields, we can work with the two-dimensional potential:

$$U(x, y, \sigma_x, \sigma_y) = \frac{ne}{4\pi\epsilon_0} \int_0^\infty \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q}\right)}{\sqrt{(2\sigma_x^2+q)(2\sigma_y^2+q)}} \mathrm{d}q \,, \tag{A.13}$$

where n is the line density of particles in the beam, e is the elementary charge, and ϵ_0 is the electrostatic constant. In this case, we do not yet make any assumptions on the longitudinal distribution.

B Appendix: tune shifts and spread due to beam-beam interactions

B.1 Linear tune shift

For the one-dimensional case, we write the betatron motion of a single particle as a simple harmonic oscillator and use the 'smooth approximation':

$$r = R\cos(\Phi) \tag{B.1}$$

and its derivative:

$$r' = -\frac{R}{\beta}\sin(\Phi) \,. \tag{B.2}$$

After a small kick from the beam–beam interaction, the phase Φ will be shifted and we can calculate $\mathrm{d}\Phi$, which is the instantaneous tune change ΔQ_i multiplied by 2π (Fig. B.1). We have [30, 31]

$$2\pi\Delta Q_\mathrm{i} = -\frac{\mathrm{d}r' \cdot \cos(\Phi) \cdot \beta}{R} \,. \tag{B.3}$$

The deflection $\mathrm{d}r'$ was calculated in Eq. (16):

$$\mathrm{d}r'(r) = -\frac{2Nr_0}{\gamma r}\left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right]; \tag{B.4}$$

linearized for small amplitudes r, it becomes (see Eq. (18))

$$\delta r'|_{r\to 0} = -\frac{Nr_0 \cdot r}{\gamma\sigma^2} = -\frac{Nr_0 \cdot R\cos(\Phi)}{\gamma\sigma^2}, \tag{B.5}$$

$$2\pi\Delta Q_\mathrm{i} = -\frac{Nr_0 \cdot R\cos(\Phi)}{\gamma\sigma^2} \cdot \frac{\cos(\Phi)\beta}{R}, \tag{B.6}$$

$$\Delta Q_\mathrm{i} = -\frac{Nr_0\beta}{2\pi\gamma\sigma^2} \cdot \cos^2(\Phi). \tag{B.7}$$

After averaging Φ from 0 to 2π:

$$\Delta Q = \frac{1}{2\pi}\int_0^{2\pi} \Delta Q_\mathrm{i}\mathrm{d}\Phi, \tag{B.8}$$

we get the linear beam–beam tune shift:

$$\Delta Q = \xi = -\frac{Nr_0\beta}{4\pi\gamma\sigma^2}. \tag{B.9}$$

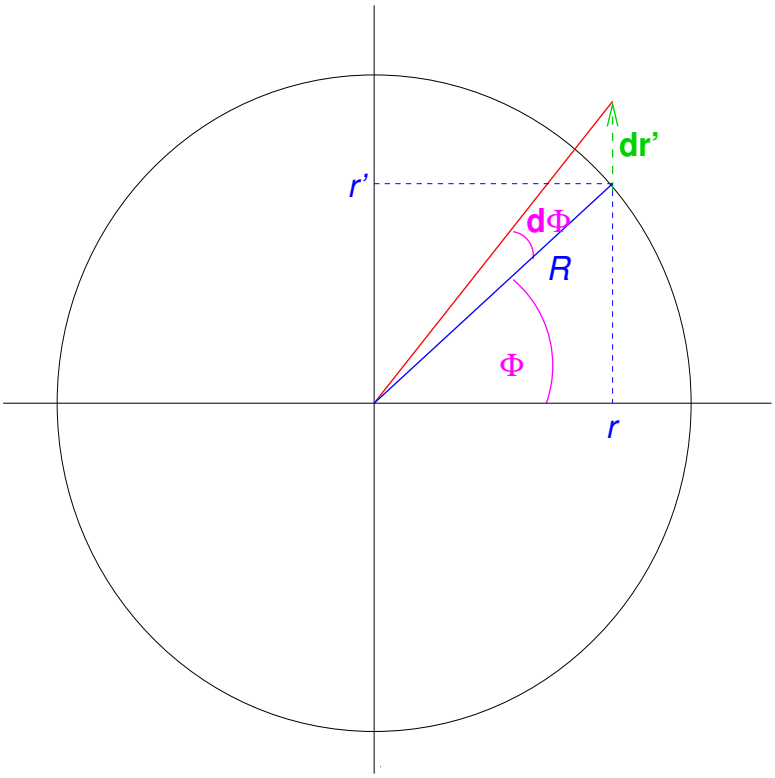


Fig. B.1: Phase space before and after the beam–beam kick; change of phase $\mathrm{d}\Phi$

B.2 Non-linear tune shift

For the non-linear tune shift, we must not linearize the beam–beam force; we get, for the instantaneous tune shift:

$$\Delta Q_\mathrm{i,nl} = -\frac{Nr_0\beta}{\pi\gamma} \cdot \frac{1 - \mathrm{e}^{-\frac{R^2}{2\sigma^2}\cos^2(\Phi)}}{R^2}. \tag{B.10}$$

To perform the integral, we first replace the $\cos^2(\Phi)$ term in the exponential with the expression

$$\cos^2(\Phi) = \frac{1}{2}\left(1 + \cos(2\Phi)\right) \tag{B.11}$$

and then perform the integral using

$$\frac{1}{\pi}\int_0^\pi \mathrm{e}^{x\cos(\Phi)}\mathrm{d}\Phi = I_0(x)\,, \tag{B.12}$$

where $I_0(x)$ is the modified Bessel function, and get the formula for the non-linear detuning with the amplitude J:

$$\Delta Q(J) = \xi \cdot \frac{2}{J} \cdot \left(1 - I_0(J/2) \cdot \mathrm{e}^{-J/2}\right)\,, \tag{B.13}$$

which is $J = \epsilon\beta/2\sigma^2$ in the usual units. Here, ϵ is the particle 'emittance' and not the beam emittance.

C Appendix: Non-linear perturbations due to beam-beam interactions

C.1 Classical approach

A standard treatment to assess non-linear perturbations is the s-dependent Hamiltonian and perturbation theory:

$$\mathcal{H} = \mathcal{H}_0 + \delta(s)\epsilon V\,, \tag{C.1}$$

where $\mathcal{H}_0$ is the unperturbed part of the Hamiltonian and ϵV describes the perturbation caused at the position s, specified by the δ function. The mathematical treatment is rather involved and, in most cases, cannot be carried beyond leading order in the perturbation. This can easily lead to wrong conclusions, which can still be found in the literature (e.g., fourth-order resonance cannot be driven by sextupoles) and one must ask whether this is the most appropriate tool to deal with this problem. In the case of isolated non-linearities caused by very local beam–beam interactions, we favour a map-based approach, as promoted by Dragt [32, 33] and described in detail by Forest [34].

C.2 Map-based approach using Lie transforms and invariant

In this approach, the ring is represented by a finite sequence of maps, which describe the individual elements. Possible representations of these maps are Taylor maps and Lie maps [35].

In this study, we shall use the Lie maps, which are always symplectic; other advantages will become obvious [9]. This technique allows us to derive invariants of the motion in a straightforward way; in particular, the extension of the results to multiple beam–beam interactions becomes an easy task.

To answer the initial question, this is particularly relevant, since we want to investigate the effect of the number of interaction points and the relative phase advance on the beam dynamics.

In the first part, we derive the formulae for a single interaction point and later extend the method to multiple beam–beam interactions.

C.2.1 Single interaction point

The derivation for a single interaction point can be found in the literature (see a particularly nice derivation by Chao [35]).

In this simplest case of one beam–beam interaction, we can factorize the machine into a linear transfer map $e^{:f_2:}$ and the beam–beam interaction $e^{:F:}$, i.e.,

$$e^{:f_2:} \cdot e^{:F:} = e^{:h:} \tag{C.2}$$

with

$$f_2 = -\frac{\mu}{2}\left(\frac{x^2}{\beta} + \beta p_x^2\right), \tag{C.3}$$

where μ is the overall phase, i.e., the tune Q multiplied by 2π, and β is the β-function at the interaction point. The function $F(x)$ corresponds to the beam–beam potential[1]

$$F(x) = \int_0^x \mathrm{d}x' f(x')\,. \tag{C.4}$$

For a Gaussian beam, we use for $f(x)$ the well-known expression for round beams:

$$f(x) = \frac{2Nr_0}{\gamma x}\left(1 - \mathrm{e}^{\frac{-x^2}{2\sigma^2}}\right). \tag{C.5}$$

Here, N is the number of particles per bunch, r_0 is the classical particle radius, γ is the relativistic parameter, and σ is the transverse beam size.

For the analysis, we examine the invariant h, which determines the one-turn map written as a Lie transformation $e^{:h:}$. The invariant h is the effective Hamiltonian for this problem.

As usual we transform to action and angle variables A and Φ, related to the variables x and p_x through the transformations

$$x = \sqrt{2A\beta}\sin(\Phi)\,, \qquad p_x = \sqrt{\frac{2A}{\beta}}\cos(\Phi)\,. \tag{C.6}$$

With this transformation, we get a simple representation for the linear transfer map f_2:

$$f_2 = -\mu A\,. \tag{C.7}$$

The function $F(x)$ we write as a Fourier series:

$$F(x) \Rightarrow \sum_{n=-\infty}^{\infty} c_n(A)\mathrm{e}^{\mathrm{i}n\Phi} \tag{C.8}$$

with the coefficients $c_n(A)$:

$$c_n(A) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}n\Phi}F(x)\mathrm{d}\Phi\,. \tag{C.9}$$

For an evaluation of Eq. (C.9), see [35]. We take some useful properties of Lie operators (see any textbook, e.g., Refs. [34, 35]):

$$:f_2:g(A) = 0\,, \qquad :f_2:\mathrm{e}^{\mathrm{i}n\Phi} = \mathrm{i}n\mu\mathrm{e}^{\mathrm{i}n\Phi}\,, \qquad g(:f_2:)\mathrm{e}^{\mathrm{i}n\Phi} = g(\mathrm{i}n\mu)\mathrm{e}^{\mathrm{i}n\Phi} \tag{C.10}$$

and the CBH formula for the concatenation of the maps (see any textbook, e.g., Ref.s [34, 35]):

$$\mathrm{e}^{:f_2:}\mathrm{e}^{:F:} = \mathrm{e}^{:h:} = \exp\left[:f_2 + \left(\frac{:f_2:}{1-\mathrm{e}^{-:f_2:}}\right)F + \mathcal{O}(F^2):\right], \tag{C.11}$$

which gives immediately for h

$$h = -\mu A + \sum_n c_n(A)\frac{\mathrm{i}n\mu}{1-\mathrm{e}^{-\mathrm{i}n\mu}}\mathrm{e}^{\mathrm{i}n\Phi}$$

[1] For a discussion of the Lie representation as a generalized kick map, see Ref. [35].

$$\text{and} \quad h = -\mu A + \sum_n c_n(A) \frac{n\mu}{2\sin(\frac{n\mu}{2})} \mathrm{e}^{(\mathrm{i}n\Phi + \mathrm{i}\frac{n\mu}{2})} \,. \tag{C.12}$$

Away from resonances, a normal-form transformation gives

$$h = -\mu A + c_0(A) = \text{const.} \tag{C.13}$$

On resonance, i.e., for the condition

$$Q = \frac{p}{n} = \frac{\mu}{2\pi} \tag{C.14}$$

and, with $c_n \neq 0$, we have

$$\sin\left(\frac{n\pi p}{n}\right) = \sin(p\pi) \equiv 0 \quad \text{for all integers} \quad p$$

and the invariant h diverges. This is a well-known result and not surprising.

C.2.2 Non-linear beam–beam tune shift

Having derived the effective Hamiltonian,

$$h = -\mu A + c_0(A) = \text{const.}\,, \tag{C.15}$$

we can now easily write an expression for the non-linear beam–beam tune shift derived earlier:

$$\Delta Q = \frac{\mathrm{d}c_0(A)}{\mathrm{d}A} \,. \tag{C.16}$$

Using the force for a round Gaussian beam and action-angle variables, we write the beam–beam potential $F(x)$ as

$$F(x) = \frac{Nr_0}{\gamma} \int_0^{A\beta/2\sigma^2} \left(1 - \mathrm{e}^{-2\alpha\sin^2(\Phi)}\right) \frac{\mathrm{d}\alpha}{\alpha} \,. \tag{C.17}$$

The coefficients $c_n(A)$ become

$$c_n(A) = \frac{Nr_0}{\gamma} \int_0^{A\beta/2\sigma^2} \frac{\mathrm{d}\alpha}{\alpha} \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\Phi \mathrm{e}^{-\mathrm{i}n\Phi} \left(1 - \mathrm{e}^{-2\alpha\sin^2(\Phi)}\right) \,. \tag{C.18}$$

With the coefficient $c_0(A)$, we get for the tune shift as a function of the amplitude:

$$\Delta Q = \frac{1}{2\pi} \frac{Nr_0}{\gamma} \frac{\mathrm{d}}{\mathrm{d}A} \int_0^{A\beta/2\sigma^2} \frac{\mathrm{d}\alpha}{\alpha} \left(1 - \mathrm{e}^{-\alpha} I_0(\alpha)\right) \tag{C.19}$$

$$= \frac{1}{2\pi} \frac{Nr_0}{\gamma A} \left(1 - I_0\left(A\beta/2\sigma^2\right) \cdot \mathrm{e}^{-A\beta/2\sigma^2}\right) \,, \tag{C.20}$$

which, for $J = A\beta/\sigma^2$, is the result we obtained earlier.

C.2.3 Two interaction points

To study two interaction points we use a configuration as shown in Fig. C.1 and extend the treatment of a single beam–beam interaction in Ref. [36] to any number of beam–beam interactions. Following the LHC conventions, we label the interaction points IP1 and IP5. The phase advance between IP1 and IP5 is μ_1; from IP5 to IP1 it is μ_2; and the overall phase advance for one turn is $\mu = \mu_1 + \mu_2$.

For the computation, we have now two transfers f_2^1, f_2^2 and two beam–beam kicks F^1, F^2, assuming the first interaction point (IP5) at μ_1 and the second (IP1) at μ [37]. The one-turn map with the four transforms is then

$$= \mathrm{e}^{:f_2^1:} \mathrm{e}^{:F^1:} \mathrm{e}^{:f_2^2:} \mathrm{e}^{:F^2:} = \mathrm{e}^{:h_2:} \,. \tag{C.21}$$

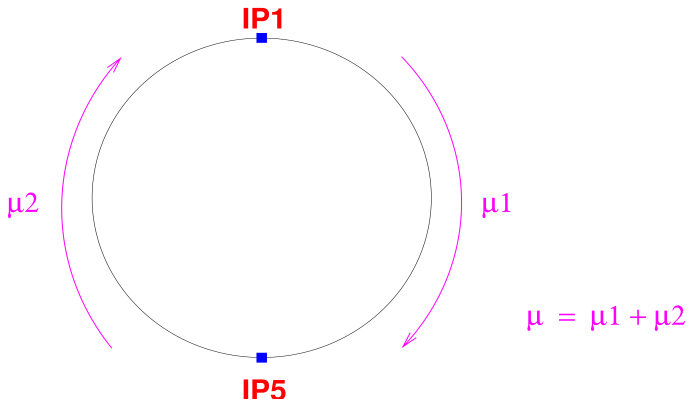


Fig. C.1: Two collision points with unequal phase advance

This we can easily re-write using the properties of Lie operators as

$$
\begin{aligned}
&= \mathrm{e}^{:f_2^1:}\mathrm{e}^{:F^1:}\mathrm{e}^{-:f_2^1:}\mathrm{e}^{:f_2^1:}\mathrm{e}^{:f_2^2:}\mathrm{e}^{:F^2:} = \mathrm{e}^{:h_2:} \\
&= \mathrm{e}^{:f_2^1:}\mathrm{e}^{:F^1:}\mathrm{e}^{-:f_2^1:}\mathrm{e}^{:f_2:}\mathrm{e}^{:F^2:}\mathrm{e}^{-:f_2:}\mathrm{e}^{:f_2:} = \mathrm{e}^{:h_2:} \\
&= \mathrm{e}^{:\mathrm{e}^{-:f_2^1:}F^1:}\mathrm{e}^{:\mathrm{e}^{-:f_2:}F^2:}\mathrm{e}^{:f_2:} = \mathrm{e}^{:h_2:}\,.
\end{aligned}
$$

Assuming now the simplification

$$
f_2 = -\mu A\,, \qquad f_2^1 = -\mu_1 A\,, \qquad \text{and} \qquad f_2^2 = -\mu_2 A \tag{C.22}
$$

and, remembering that $g(:\,f_2\,:)\mathrm{e}^{in\Phi} = g(\mathrm{i}n\mu)\mathrm{e}^{in\Phi}$, we have

$$
\mathrm{e}^{:f_2^1:}\mathrm{e}^{in\Phi} = \mathrm{e}^{\mathrm{i}n\mu_1}\mathrm{e}^{in\Phi} = \mathrm{e}^{in(\mu_1+\Phi)} \tag{C.23}
$$

and we find that the Lie transforms of the perturbations are phase-shifted (see e.g., Ref. [34]). Therefore,

$$
\mathrm{e}^{:\mathrm{e}^{-:f_2^1:}F^1:}\mathrm{e}^{:\mathrm{e}^{-:f_2:}F^2:}\mathrm{e}^{:f_2:} = \mathrm{e}^{:h_2:} \tag{C.24}
$$

becomes simpler with substitutions of $\Phi_1 \to \Phi + \mu_1$ and $\Phi \to \Phi + \mu$ in the functions G^1 and G:

$$
\mathrm{e}^{:G^1(\Phi_1):}\mathrm{e}^{:G(\Phi):}\mathrm{e}^{:f_2:} \qquad \Rightarrow \qquad \mathrm{e}^{:G^1(\Phi_1)+G(\Phi):}\mathrm{e}^{:f_2:} \tag{C.25}
$$

This reflects the phase-shifted distortions and we get for h_2

$$
h_2 = -\mu A + \sum_{n=-\infty}^{\infty} \frac{n\mu c_n(A)}{2\sin(n\frac{\mu}{2})}\left[\mathrm{e}^{-\mathrm{i}n(\Phi+\mu/2+\mu_1)} + \mathrm{e}^{-\mathrm{i}n(\Phi+\mu/2)}\right] \tag{C.26}
$$

or, re-written,[2]

$$
h_2 = -\mu A + 2c_0(A) + \underbrace{\sum_{n=2}^{\infty} \frac{2n\mu c_n(A)}{\sin\left(n\frac{\mu}{2}\right)}\cos\left[n(\Phi + \frac{\mu}{2} + \frac{\mu_1}{2})\right]\cos\left(n\frac{\mu_1}{2}\right)}_{\text{interesting part}}\,. \tag{C.27}
$$

Note well, because

$$
\mathrm{e}^{:F(\Phi):}\mathrm{e}^{:f_2:} \to \mathrm{e}^{:G^1(\Phi_1)+G(\Phi):}\mathrm{e}^{:f_2:}\,, \tag{C.28}
$$

[2]For head-on collisions only $c_n(A)$, for even orders in n are non-zero and the sum needs to be done only for even terms.

that the above treatment can be generalized to more interaction points, in particular including long-range interactions.

In practice, Eq. (C.27) is evaluated to a maximum order N, in our case up to order 40. We get

$$h_2 = -\mu A + 2c_0(A) + \underbrace{\sum_{n=2}^{N} \frac{2n\mu c_n(A)}{\sin(n\frac{\mu}{2})} \cos\left[n\left(\Phi + \frac{\mu}{2} + \frac{\mu_1}{2}\right)\right] \cos\left(n\frac{\mu_1}{2}\right)}_{\text{interesting part}} \tag{C.29}$$

with $N = 40$.

C.2.4 Comparison with numerical model

To test our result, we compare the invariant h with the results of a particle-tracking program [37].

The model we use in the program is rather simple:

- linear transfer between interactions;
- beam–beam kick for round beams;
- compute action $I = \frac{\beta^*}{2\sigma^2}\left(\frac{x^2}{\beta^*} + p_x^2\beta^*\right)$;
- compute phase $\Phi = \arctan\left(\frac{p_x}{x}\right)$;
- compare I with h as a function of the phase Φ.

The evaluation of the invariant of Eq. (C.12) is done numerically using Mathematica.

The comparison between the tracking results and the invariant h from the analytical calculation is shown in Fig. C.2 in the (I, Φ) space. Only one interaction point is used in this comparison and the particles are tracked for 1024 turns. The symbols are the results from the tracking and the solid lines are the invariants, computed as before. The two figures are computed for amplitudes of 5σ and 10σ. The agreement between the models is excellent. The analytical calculation was again made up to the order $N = 40$. Using a lower number, the analytical model can reproduce the envelope of the tracking results, but not the details.

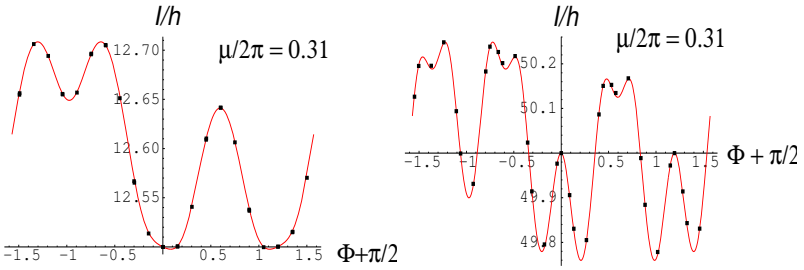


Fig. C.2: Comparison: numerical and analytical models for one interaction point. Shown for $5\sigma_x$ and $10\sigma_x$. Full symbols from numerical model and solid lines from invariant of Eq. (C.12).

Another comparison is made in Fig. C.3 for the case of two interaction points. The symbols are again from the simulation and the solid lines from the computation. For comparison, the invariant for a single interaction point is included to demonstrate the difference. Again, the agreement is excellent and shows the validity of the results.

C.3 Behaviour near a resonance

To show the behaviour of the system near a resonance, we show the invariant together with the tracking results near the third-order resonance in Fig. C.4.

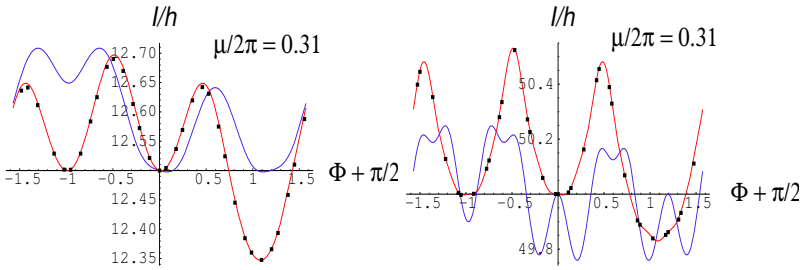


Fig. C.3: Comparison: numerical and analytical models for two interaction points. Shown for $5\sigma_x$ and $10\sigma_x$. Full symbols from numerical model and solid lines from invariant. Shown here are the invariant for one (Eq. (C.12), blue line, not passing through the full symbols) and two (Eq. (C.27), red line, passing through the full symbols) interactions to demonstrate the difference and the agreement with the tracking program.

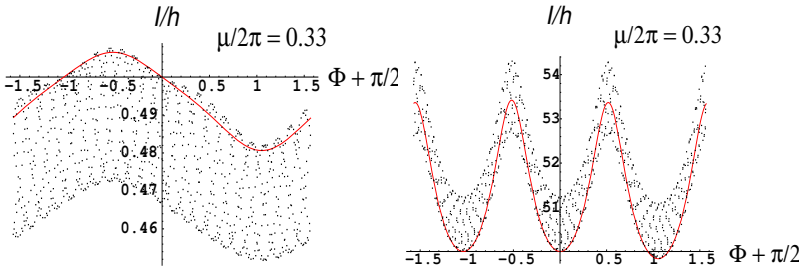


Fig. C.4: Comparison: numerical and analytical models for two interaction points on a resonance (third-order)

It is clearly demonstrated that the simulation differs quantitatively from the computed invariant at resonant tunes.

Some Effects Near Transition

E. Métral
CERN, Geneva, Switzerland

Abstract
This paper discusses some general longitudinal and transverse beam dynamics effects that occur near the transition energy and when it is crossed, with application to the particular cases of the CERN Proton Synchrotron and Super Proton Synchrotron machines.

Keywords
Transition energy; transition crossing; non-adiabatic motion; quadrupolar oscillation; transverse instability; gamma transition.

1 Transition energy

The momentum compaction factor α_p is a parameter which arises from the accelerator lattice (i.e., the transverse beam optics) and plays a major role in the longitudinal beam dynamics. It is given by [1, 2]

$$\alpha_p = \frac{\mathrm{d}C/C}{\mathrm{d}p/p} = \frac{1}{C}\int_0^C \frac{D_x(s)}{\rho(s)}\,\mathrm{d}s\,, \tag{1}$$

where C is the circumference of the machine (the length of the particle trajectory), p is the beam momentum, D_x is the horizontal dispersion, ρ is the bending radius, and s is the azimuthal coordinate around the machine. For most circular machines $\alpha_p > 0$, and the relativistic mass factor at transition (called 'gamma transition' and written γ_t) is given by [2]

$$\gamma_t = \frac{1}{\sqrt{\alpha_p}}. \tag{2}$$

However, $\alpha_p < 0$ is also possible (e.g., as in the LEAR machine at CERN in the past [3]), in which case the lattice is called a Negative Momentum Compaction (NMC) lattice or imaginary-γ_t lattice. Assuming a positive momentum compaction factor, it can be deduced from Eq. (1) that an increase in the beam momentum leads to an increase in the circumference. However, an increase in the beam momentum also leads to an increase in the beam velocity. Depending on the beam energy, the increase in the circumference may be either larger or smaller than the increase in the beam velocity, and the revolution frequency may either decrease or increase. At very high energy, the beam velocity is close to the velocity of light (c) and remains about constant, which leads to a decrease in the revolution frequency. At low energy, the beam velocity increases faster than the circumference and the revolution frequency increases. There is an energy at which the velocity variation is compensated by the circumference variation, i.e., $\mathrm{d}f_{\mathrm{rev}} = 0$ (where f_{rev} is the revolution frequency), and this energy is called the transition energy.

In fact, the important parameter for the longitudinal beam dynamics is the slip factor η, which is a measure of the distance from the transition energy and is given by [2]

https://doi.org/10.23730/CYRSP-2017-003.213

$$\eta = -\frac{\mathrm{d}f_{\mathrm{rev}} / f_{\mathrm{rev}}}{\mathrm{d}p / p} = \alpha_{\mathrm{p}} - \frac{1}{\gamma^2} = \frac{1}{\gamma_{\mathrm{t}}^2} - \frac{1}{\gamma^2}, \tag{3}$$

where γ is the relativistic mass factor of the beam. The slip factor is negative below transition, equal to zero at transition (this is called the isochronous condition), and positive above transition.

2 Longitudinal beam dynamics 'far' below or above transition

This is the usual (classical or adiabatic) case. The bucket separatrices are recalled in Fig. 1 for several values of the synchronous phase ϕ_{s} below and above transition [2], using the RF phase ϕ and its time derivative (a dot denotes the derivative with respect to time) as variables, where Ω_{s} is the angular synchrotron frequency, given by

$$\Omega_{\mathrm{s}} = \Omega_{\mathrm{rev}} \sqrt{-\frac{eV_{\mathrm{RF,peak}}\, h\eta \cos\phi_{\mathrm{s}}}{2\pi\beta^2 E_{\mathrm{total}}}}\ . \tag{4}$$

Here, Ω_{rev} is the angular frequency of revolution, e is the elementary charge, $V_{\mathrm{RF,peak}}$ is the peak RF voltage, h is the RF harmonic number, β is the relativistic velocity factor, and E_{total} is the total beam energy. As the term under the square root sign in Eq. (4) has to be positive (for stable particle motion), the synchronous phase ϕ_{s} has to be shifted to $\pi - \phi_{\mathrm{s}}$ above transition. Some examples of particle trajectories are shown in Fig. 2.

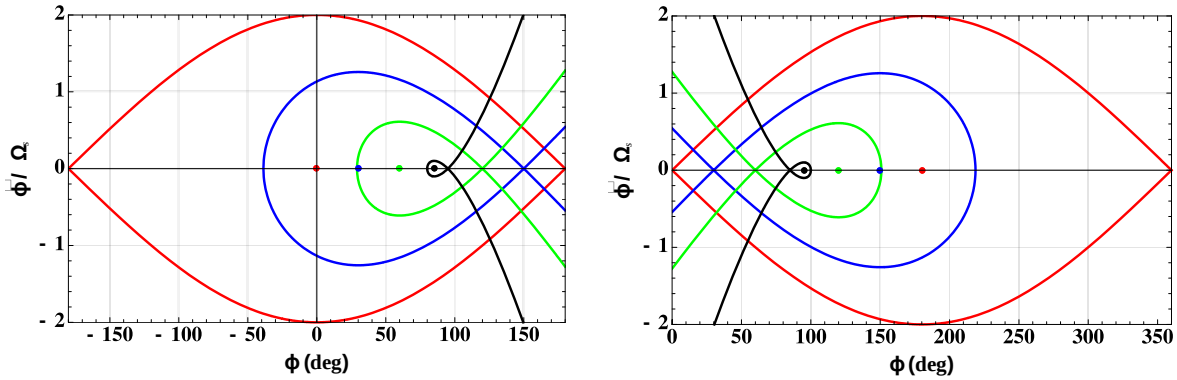


Fig. 1: RF bucket separatrices. Left, below transition for several synchronous phases: 0° (red), 30° (blue), 60° (green), 85° (black). Right, above transition for several synchronous phases: 180° (red), 150° (blue), 120° (green), 95° (black).

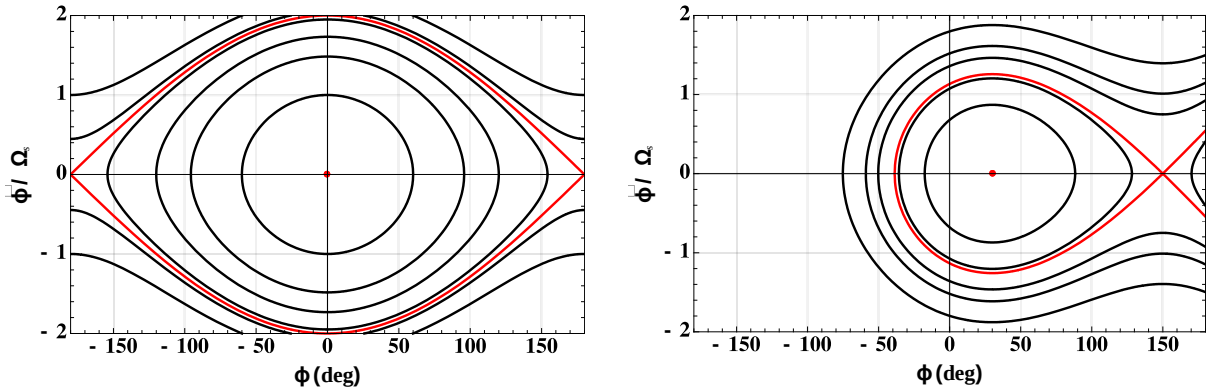


Fig. 2: Examples of particle trajectories (black) below transition, inside and outside the bucket separatrices (red), for two synchronous phases: 0° (left) and 30° (right).

3 Transition crossing

The case 'far' below or above transition discussed in Section 2 corresponds to the case where the adiabaticity condition

$$\frac{1}{\Omega_s^2}\left|\frac{d\Omega_s}{dt}\right| \ll 1 \tag{5}$$

is satisfied. 'Close' to transition, the adiabaticity condition is not satisfied any more and a non-adiabatic description of the synchrotron motion is necessary, as the particles will not be able to catch up with the rapid modification of the bucket shape [1, 4]. To describe the synchrotron motion close to transition, a non-adiabatic (characteristic) time T_c is defined by

$$T_c = \left(\frac{\beta^2 E_{rest} \gamma_t^4}{4\pi f_{rev}^2 \dot{\gamma} h V_{RF,peak} \left|\cos\phi_s\right|} \right)^{1/3}, \tag{6}$$

where E_{rest} is the beam rest energy. In the particular case of the bunch used for the nTOF facility [5] in the CERN Proton Synchrotron (PS), whose main parameters are reported in Table 1, the non-adiabatic time is ~2 ms.

Table 1: Relevant parameters for the CERN PS and the nTOF bunch [5]

Parameter	**Symbol (units)**	**Value**
Average machine radius	R (m)	100
Bending dipole radius	ρ (m)	70
Variation of the magnetic field	dB/dt (T/s)	2.2
Peak RF voltage	$V_{RF,peak}$ (kV)	200
RF harmonic number	h	8
Momentum compaction factor (gamma transition)	α_p (γ_t)	0.027 (6.1)
Longitudinal (total) emittance	ε_L (eV s)	2
Number of protons/bunch	N_b (10^{10} p/b)	800
Normalized r.m.s. transverse emittance	$\varepsilon^*_{x,y}$ (μm)	5
Transverse average betatron function	$\beta_{x,y}$ (m)	16
Beam pipe half-gaps	$b_{x,y}$ (cm)	3.5, 7.0
Transverse tunes	$Q_{x,y}$	6.25

The general variation of the bunch length, normalized by the value at transition, and of the momentum spread, obtained using a non-adiabatic theory (neglecting intensity effects for the moment), is depicted in Fig. 3, where several interesting features can be observed. First, contrary to the adiabatic theory, the bunch length does not decrease to zero at transition but reaches a minimum value. Second, the values of the bunch length and momentum spread below and above transition are symmetric with respect to the transition. Third, the product of the bunch length and the momentum spread is not constant, which means that the particle trajectory (which is an ellipse for small synchrotron amplitudes) is not always upright. The variation of the angle of the tilted ellipse around transition is shown on the left in Fig. 4, and the shape of the ellipse (using time and energy as variables) at transition is shown on the right.

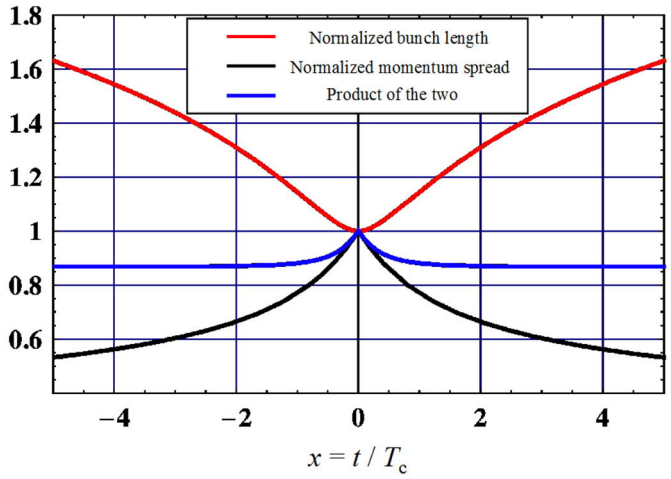


Fig. 3: General variation of the bunch length (normalized by the value at transition) and momentum spread close to transition obtained using the non-adiabatic theory [1, 4].

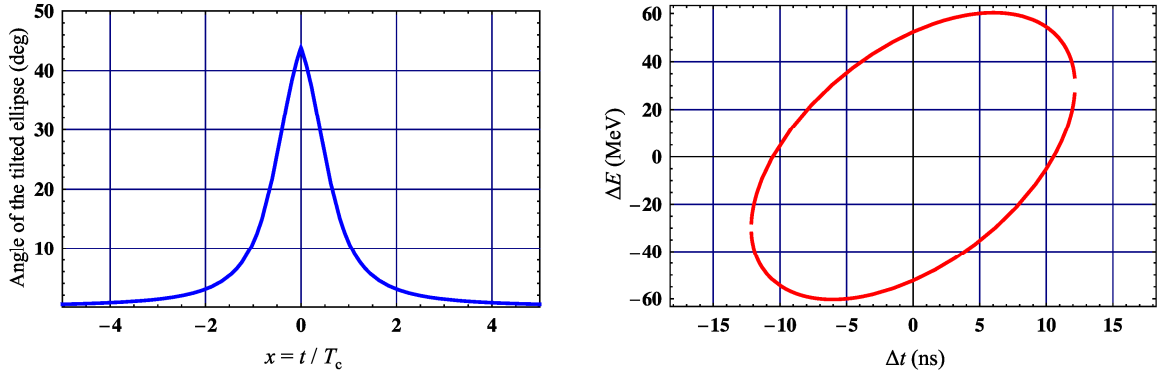


Fig. 4: Left, variation of the angle of the tilted ellipse (particle trajectory for small synchrotron amplitudes) around transition. Right, shape of the ellipse at transition.

As the bunch intensity increases, the effects of intensity (such as the longitudinal space charge effect and the interaction with the longitudinal impedance) cannot be neglected any more [4, 6, 7]. Considering first only the longitudinal space charge effect, as it is defocusing below transition, the final bunch length below transition will be longer than predicted above. Furthermore, as the longitudinal space charge effect is focusing above transition, the final bunch length above transition will be shorter than predicted above. This means that, because of the longitudinal space charge effect, an intensity-dependent step (i.e., a longitudinal mismatch) will thus appear at transition, as can be seen in Fig. 5 for the static case (i.e., when the transition is not crossed). In the dynamic case (i.e., when the transition is crossed), because of the longitudinal mismatch introduced by the intensity-dependent step at transition, the bunch length will oscillate, as can be seen in Fig. 6. Considering now, instead of the longitudinal space charge effect, the interaction with the inductive part of the longitudinal broadband impedance, the situation is similar but the bunch length is shorter below (and larger above) transition owing to the focusing (or defocusing) effect of the impedance (see Fig. 7). Detailed measurements (which can be used to estimate the imaginary part of the longitudinal impedance) on the CERN PS can be found in Ref. [8].

There are two main remedies for overcoming these intensity effects (and some other remedies, which will also be discussed below):

– Avoid crossing the transition in the design phase, i.e., use $\alpha_p < 0$.

– If transition crossing cannot be avoided, use a 'γ_t jump', which consists in an artificial increase in the transition-crossing speed by means of fast pulsed quadrupoles (at locations of non-zero dispersion) [1].

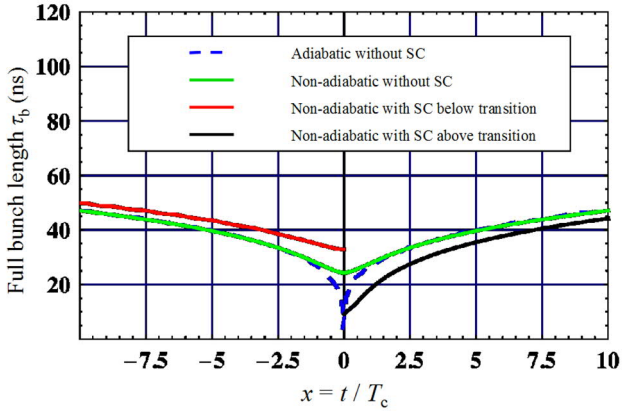


Fig. 5: Evolution of the full (4σ) bunch length versus time near transition (occurring at time 0) in the case of the adiabatic theory without Space Charge (SC), and in the case of the non-adiabatic theory with and without SC in the static case (i.e., without crossing the transition), applied to the CERN PS nTOF bunch (see Table 1).

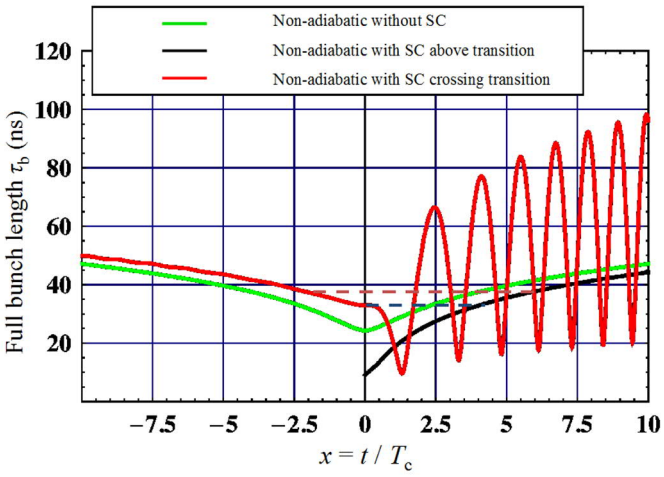


Fig. 6: Evolution of the full (4σ) bunch length versus time near transition (occurring at time 0) in the case of the non-adiabatic theory with and without space charge in the dynamic case (i.e., crossing the transition), applied to the CERN PS nTOF bunch (see Table 1).

The γ_t jump in the PS is shown in Fig. 8, where the effective crossing speed is increased by a factor ~50. As can be seen from Fig. 6, the idea of the γ_t jump is to switch, as fast as possible, from a certain bunch length below transition to the same bunch length above transition to avoid longitudinal quadrupolar oscillations (which ultimately lead to longitudinal-emittance blow-up through filamentation). One possibility could be to start the γ_t jump exactly at transition (see the dashed blue line in Fig. 6), which would lead to the smallest jump, but in this case we would move from a tilted to a non-tilted ellipse (see Fig. 4), which would still lead to some mismatch. It is cleaner to start the γ_t jump earlier, say at $\sim -2T_c$ (see Fig. 6), to avoid the effect of the tilted ellipse; this leads to a larger jump (see the dashed brown line in Fig. 6). In both cases, owing to the asymmetry introduced by intensity effects below and above transition, the optimum γ_t jump is asymmetric.

As discussed above, the important parameter for the longitudinal beam dynamics is the slip factor η, whose variation around transition is shown in Fig. 9. As can be seen, in the presence of the γ_t jump, the absolute value of the slip factor is kept above a certain value (which has to be defined

depending on the machine parameters and the issues observed) for most of the time, except for a very short time (~500 μs) close to transition.

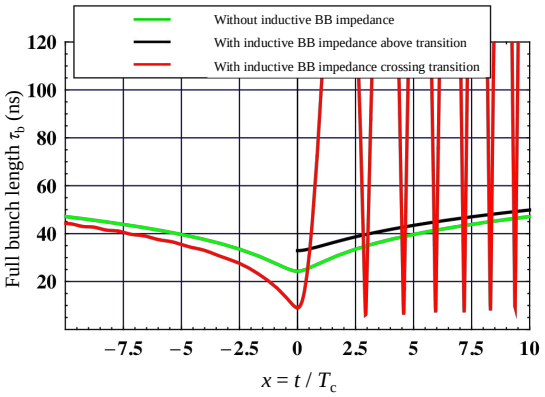


Fig. 7: Evolution of the full (4σ) bunch length versus time near transition (occurring at time 0) in the case of the non-adiabatic theory with (only) and without an inductive broadband (BB) impedance in the dynamic case (i.e., crossing the transition), applied to the CERN PS nTOF bunch (see Table 1). Here, a broadband impedance with an inductive part of 20 Ω has been used.

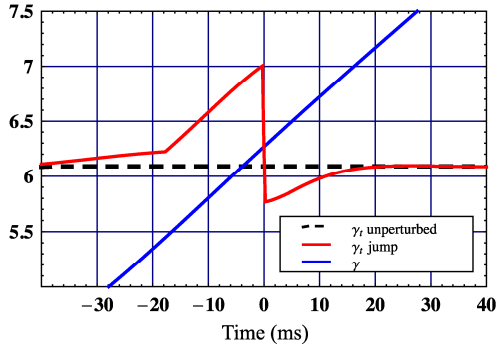


Fig. 8: Evolution of the relativistic mass factor γ (with $\mathrm{d}\gamma/\mathrm{d}t = 49.9\ \mathrm{s}^{-1}$) for the CERN PS nTOF bunch (see Table 1) around transition (occurring at time 0) and of γ_t, without (i.e., unperturbed) and with a jump.

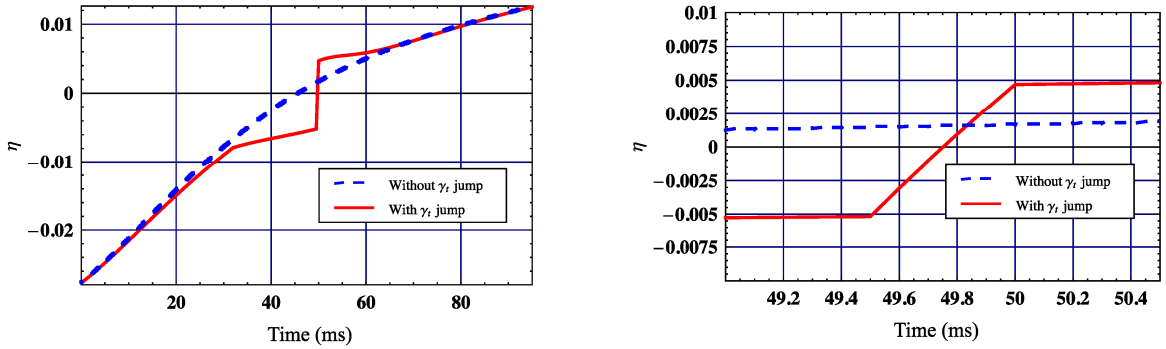


Fig. 9: Left, evolution of the slip factor η around transition (corresponding to Fig. 8) without and with the γ_t jump. Right, zoom close to transition, which reveals that the absolute value of the slip factor is 'small' for only a very short time (~500 μs).

4 Transverse (slow) head–tail instability

If the sign of the chromaticities (which are of order −1 for an uncorrected machine such as the CERN PS) is not changed (in both transverse planes) above transition, a (single-bunch) head–tail instability

may develop. This occurs only above a certain intensity in practice, even though this kind of instability has no threshold, owing to its rise time and to intrinsic non-linearities, which provide some Landau damping. This (slow) instability can be damped through Landau damping [6] using octupoles, which introduce some amplitude detunings. This method was used in the past to stabilize the PS beam. However, the better method of changing the sign of the chromaticities (and keeping them at small positive values) by acting on the optics with sextupoles was then adopted, and this has been a routine operation at the PS for many years. The transverse chromatic frequencies should be (slightly) positive to prevent the head–tail mode 0 (the most critical) from developing [1]:

$$f_{\xi_{x,y}} = Q_{x,y} f_{\mathrm{rev}} \frac{\xi_{x,y}}{\eta} > 0 \,, \tag{7}$$

where $\xi_{x,y}$ are the transverse chromaticities. An example of a transverse head–tail instability observed in the PS at injection (below transition) is shown in Fig. 10 (top), where several consecutive traces from a pick-up are superimposed and a clear 'standing-wave' pattern (with four nodes, called 'mode 4') can be observed. A comparison with an analytical prediction (for a bunch passing through the centre of the pick-up) is shown in Fig. 10 (bottom). The difference is mainly due to the fact that the measured bunch did not pass through the centre of the pick-up.

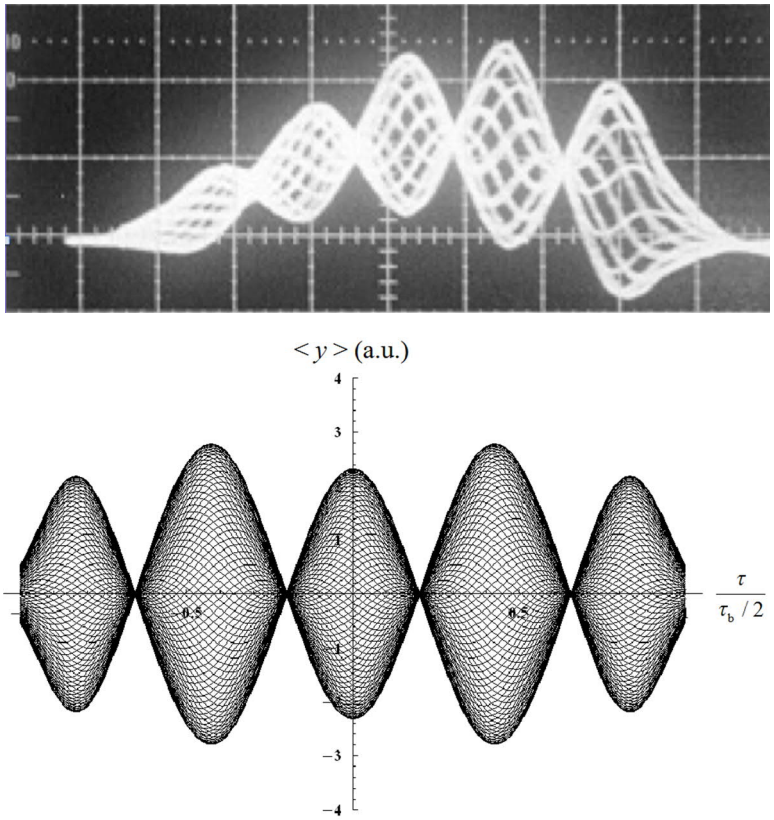


Fig. 10: Top, example of a transverse (slow) head–tail instability observed in the CERN PS at injection (below transition), with several consecutive traces from a pick-up superimposed; a clear 'standing-wave' pattern (with four nodes, called 'mode 4') can be observed. Bottom, comparison with an analytical prediction (for a bunch passing through the centre of the pick-up).

5 Fast (vertical) single-bunch instability

As the bunch intensity increases, the different head–tail modes can no longer be treated separately. In this regime, the wake fields couple the modes together and a wave pattern travelling along the bunch is created: this is the Transverse Mode-Coupling Instability (TMCI). A simple formula can be obtained from five seemingly diverse formalisms (in the absence of space charge and transverse feedback), assuming a broadband impedance and the long-bunch regime (two assumptions which are discussed below): (i) the coasting-beam approach with peak values, (ii) fast blow-up, (iii) beam break-up, (iv) post-head–tail, and (v) TMCI [1]. An example of a (simulated) TMCI for the case of the CERN Super Proton Synchrotron (SPS), assuming a broadband impedance and without taking into account the effects of both space charge and transverse damper, is shown in Fig. 11 (versus the bunch current I_b) [9], where the HEADTAIL tracking code [10] and MOSES calculations [11] are compared. Good agreement was observed between the two codes, which both showed a coupling between the (azimuthal) modes −2 and −3. Figure 11 (bottom) shows the imaginary part of the normalized tune shift, which is equal to

$$\frac{T_s}{2\pi\tau_{\text{instab}}}, \tag{8}$$

where T_s is the synchrotron period and τ_{instab} the rise time of the instability. Figure 11 (top) shows why the TMCI is more critical close to the transition energy, as the various azimuthal modes are spaced by the synchrotron tune Q_s, which tends to zero as the transition is approached.

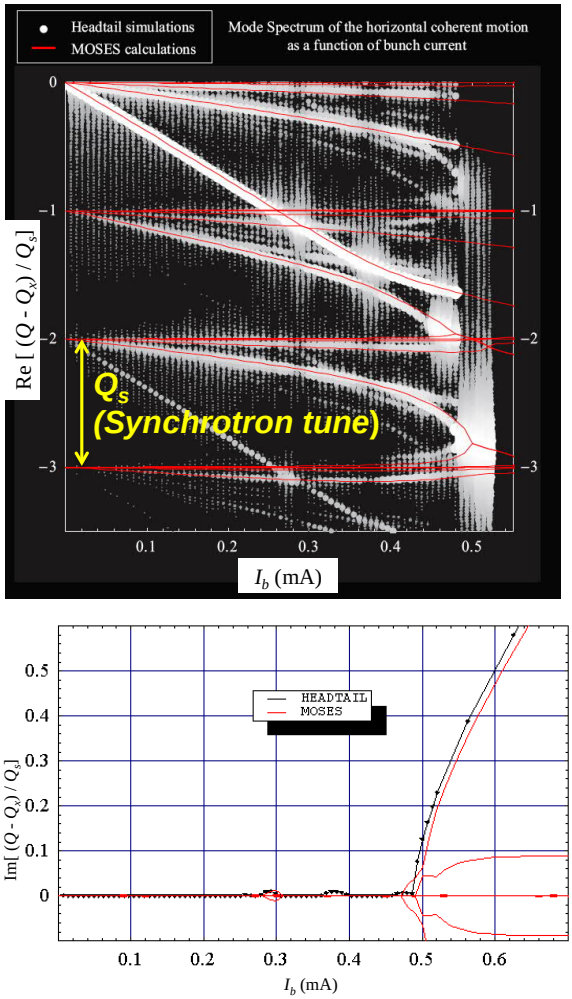


Fig. 11: Example of a simulated TMCI in the CERN SPS [9]: (top) real part and (bottom) imaginary part of the normalized tune shift versus the bunch current.

An example of a broadband impedance (i.e., a resonator impedance with a quality factor $Q = 1$), the first assumption discussed above, is shown in Fig. 12 for a case similar to that of the CERN SPS. The threshold for the TMCI long-bunch regime, the second assumption discussed above, is shown in Fig. 13: the long-bunch regime corresponds to the case where the product of the bunch length and the broadband resonance frequency f_r is much greater than ½.

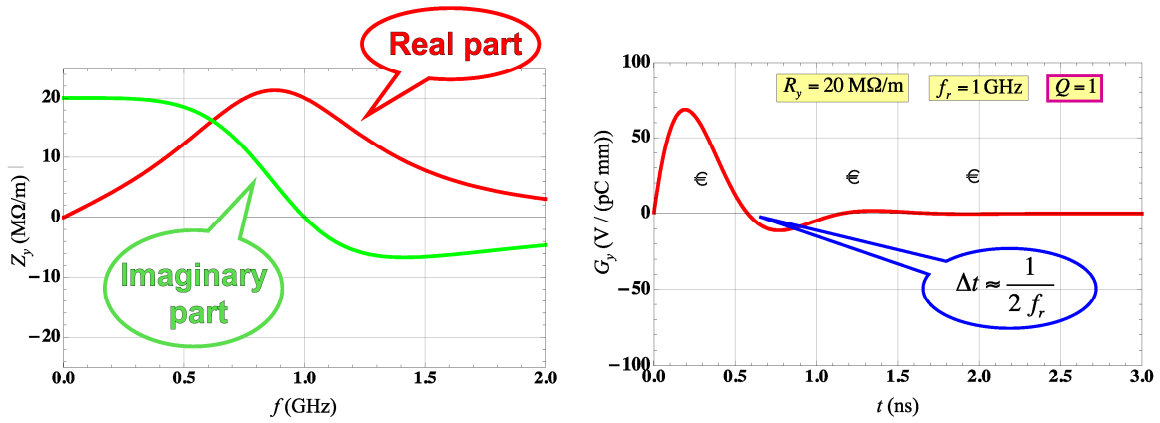


Fig. 12: Example of a (vertical) broadband impedance for a case similar to that of the CERN SPS: (left) real and imaginary parts versus frequency, (right) the corresponding wake function (or Green function) versus time.

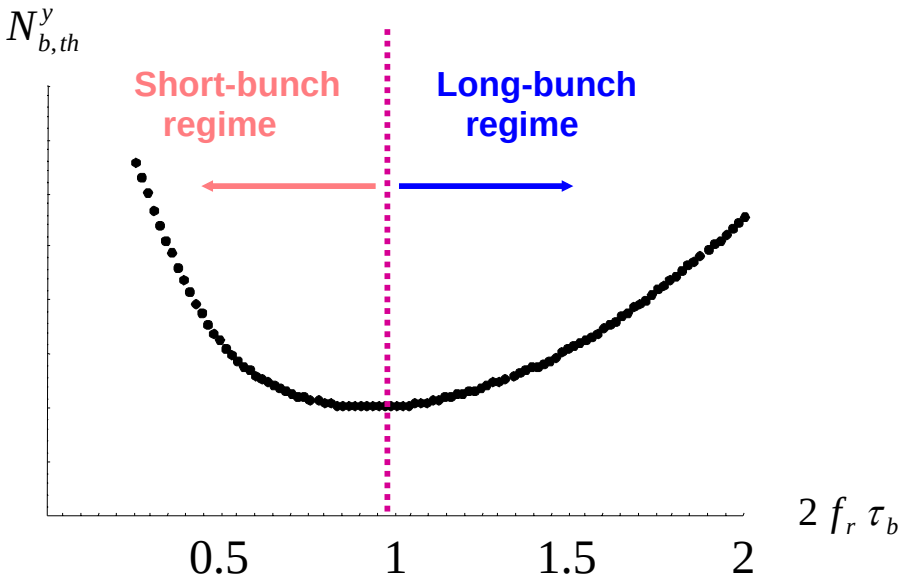


Fig. 13: TMCI threshold (number of particles per bunch) versus the product of the total (4σ) bunch length and the broadband resonance frequency.

Assuming that the two assumptions above are valid, a simple formula giving the threshold number of particles per bunch can be written as follows:

$$N_{b,th}^{y} \propto \frac{f_r}{|Z_y|}\,|\eta|\,Q_y\,\varepsilon_L\left(1+\frac{f_{\xi_y}}{f_r}\right) \tag{9}$$

- Increase the chromatic frequency - Chromaticity jump if transition has to be crossed

Increase the beam longitudinal emittance (when possible)

Try to decrease the impedance and/or increase the resonance frequency => Impedance reduction campaign

Change the optics to increase the betatron tune (decrease the beta function at critical impedances) and/or go further away from transition => New optics needed

where various means to increase the intensity threshold are also mentioned. Note that there is no dependence on Q_s in Eq. (9): what is important is the distance from transition, expressed through the slip factor. Equation (9) also shows that the TMCI does not disappear at high energy (even though the beam becomes increasingly stiff) but the intensity threshold saturates with the slip factor, as has been checked with the HEADTAIL code [12].

In the CERN PS, even in the presence of a γ_t jump together with a change in the sign of both chromaticities when the transition is crossed, a fast vertical single-bunch instability is observed (with the nTOF bunch) when no longitudinal-emittance blow-up is applied before the transition. An example of such an instability is shown in Fig. 14. The instability was suppressed by increasing the longitudinal emittance [13].

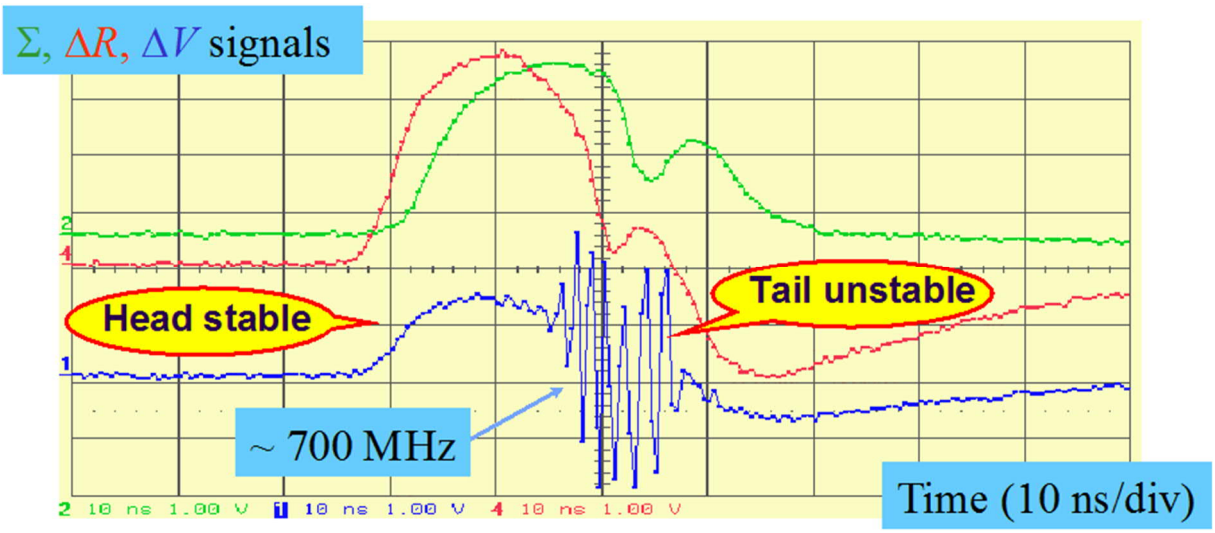


Fig. 14: Single-turn signals from a wideband pick-up in the CERN PS observed in 2000 with an nTOF bunch. From top to bottom: Σ (sum signal), Δ*R* (radial/horizontal signal), and Δ*V* (vertical signal). Time scale: 10 ns/div.

In the CERN SPS, a fast vertical single-bunch instability was first observed when the longitudinal emittance of the bunch was too small, an example of which is shown in Fig. 15 [14]. The instability was first stabilized by increasing the chromaticities, but high values of chromaticities can also have detrimental effects on the beam lifetime (in particular for the nominal multibunch beam in the LHC). In

contrast to the standing-wave pattern observed in the case of the (slow) head–tail instability shown in Fig. 10, this instability exhibited a 'travelling-wave' pattern at low chromaticity (see Fig. 16 (top); this can be seen better in the movie available with the slides for this course). The instability was suppressed at high chromaticity (see Fig. 16 (bottom); the movie shows clearly how a high chromaticity prevents the instability from developing).

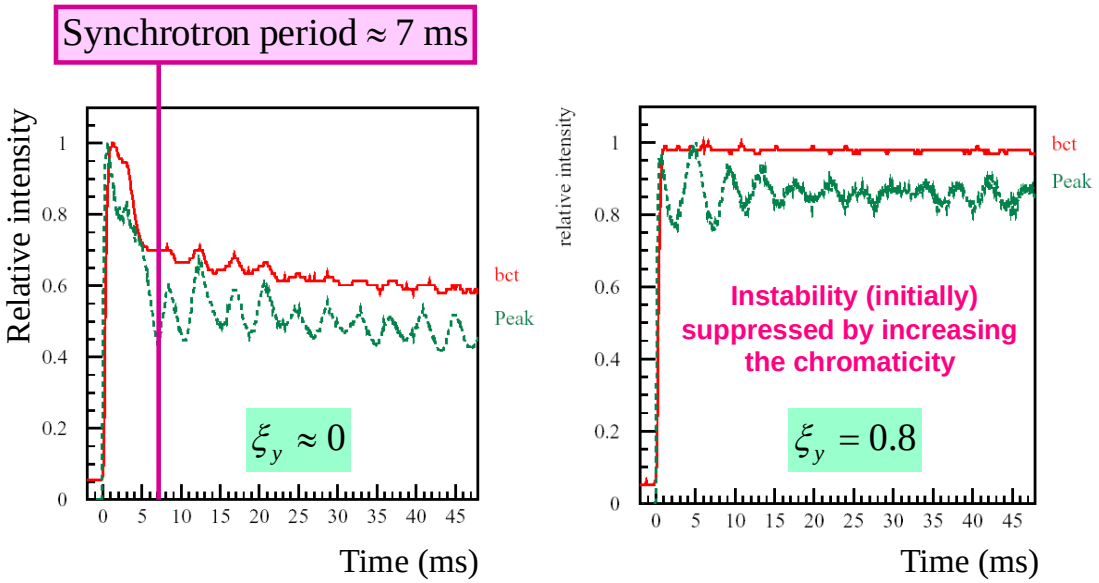


Fig. 15: Normalized bunch intensity (red curve) versus time after injection in the CERN SPS with (left) chromaticity ~0 and (right) chromaticity ~0.8.

Following Eq. (9), the value of γ_t for the SPS has recently been modified to increase the TMCI intensity threshold above the intensities foreseen in future upgrades. For a machine made of simple FODO cells, the following simple estimates of the horizontal dispersion and γ_t can be derived (approximating the machine radius by the bending radius):

$$D_x \approx \frac{\rho}{Q_x^2}, \qquad \gamma_t \approx Q_x . \tag{10}$$

Therefore, if one wants to modify γ_t (i.e., increase or decrease its value), one needs to modify the horizontal tune. This is what has been done in the SPS, where the integer tune was decreased from 26 (in the nominal optics, known as the Q26 optics) to 20 (in the new optics, known as the Q20 optics) [15]. In the case of the Q26 optics, $\gamma_t = 22.8$ and $|\eta| Q_y = 0.62 \times 10^{-3} \times 26.13 \approx 0.0162$ (at SPS injection at 26 GeV/c), while in the case of the Q20 optics, $\gamma_t = 18$ and $|\eta| Q_y = 1.80 \times 10^{-3} \times 20.13 \approx 0.0362$. Therefore, according to Eq. (9), the intensity threshold should be increased by a factor of $0.0362/0.0162 \approx 2.2$. As the TMCI intensity threshold with the Q26 optics was measured to be $\sim 1.7 \times 10^{11}$ p/b [9], the expected intensity threshold with the Q20 optics was $\sim 3.7 \times 10^{11}$ p/b. Measurements performed with the Q20 optics showed good agreement with Eq. (9), as the measured intensity threshold was $\sim 4.5 \times 10^{11}$ p/b, i.e., a gain by a factor of ~2.6 was achieved instead of the ~2.2 foreseen, as can be seen from Fig. 17 [15]. A detailed comparison of measurements and HEADTAIL simulations showed even better agreement, as can be seen from Fig. 18. It can be concluded that with the new Q20 optics, the predicted and measured intensity thresholds are $\sim 4.5 \times 10^{11}$ p/b, i.e., well above the maximum value discussed for future upgrades.

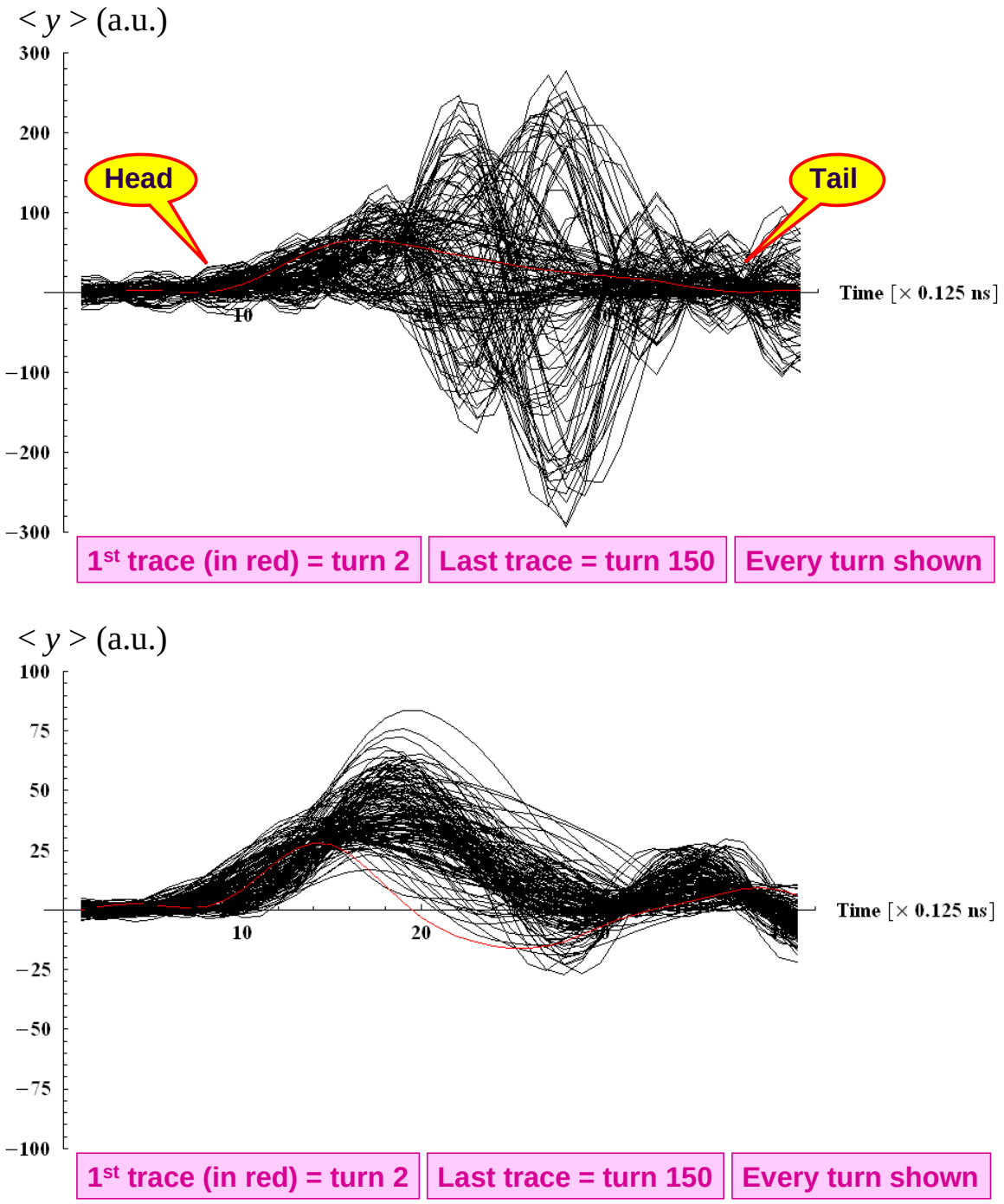


Fig. 16: Top, travelling-wave pattern along the bunch (which can be seen better in the movie available with the slides for this course) during an instability observed at low (~0.14) chromaticity. Bottom, a high chromaticity (~2.04) prevents the instability from developing.

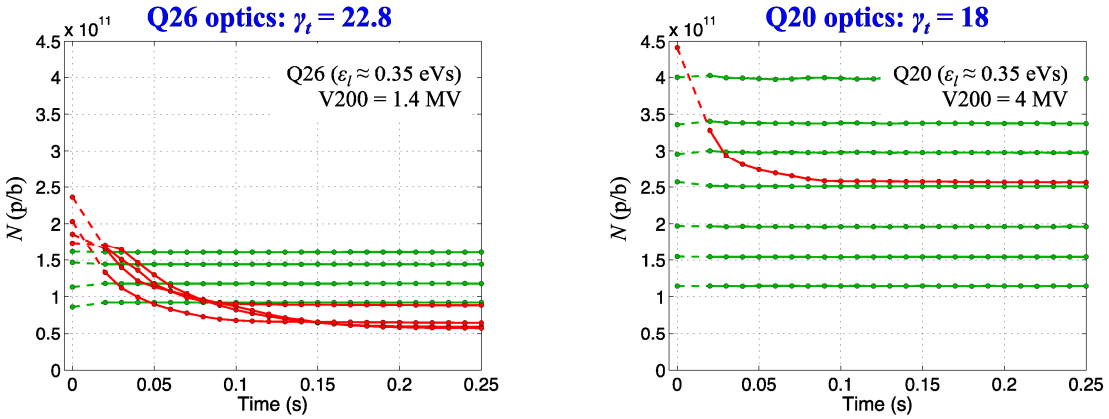


Fig. 17: Number of protons per bunch versus time after injection into the SPS at 26 GeV/*c*: (left) with the old Q26 optics, showing a threshold at ~1.7 × 10^{11} p/b [9], and (right) with the new Q20 optics, showing a threshold at ~4.5 × 10^{11} p/b [15].

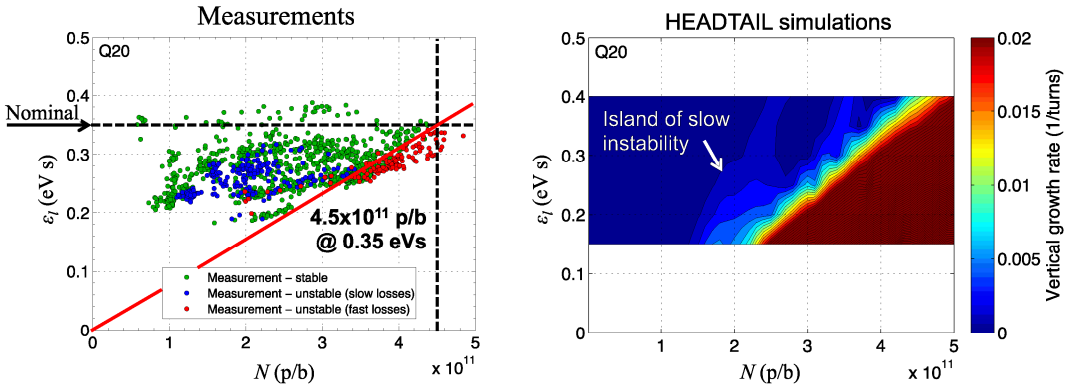


Fig. 18: Detailed comparison between (left) measurements and (right) HEADTAIL simulations of the longitudinal emittance required to reach stability, versus bunch intensity [15].

6 Conclusion

Many intricate beam dynamics effects are observed near the transition energy and when it is crossed. To ensure single-particle longitudinal phase stability, the first thing to be done when the transition is crossed is to shift the synchronous phase from ϕ_s below transition to $\pi - \phi_s$ above transition. Then, to avoid the most critical instability, namely the (slow) head–tail instability of (azimuthal) mode 0, the chromatic frequency should be kept positive, i.e., the signs of the two transverse chromaticities should be changed from negative below transition to positive above transition. Furthermore, above a certain intensity, a TMCI might develop. A simple approximate formula, valid in the presence of a broadband impedance and in the long-bunch regime (the effects of space charge and transverse feedback are still under discussion), shows that the various possible remedies for this are: (i) an impedance reduction campaign; (ii) an increase in the longitudinal emittance (as in the CERN PS); (iii) an increase in the absolute value of the slip factor (as in the SPS) and/or the tune; (iv) an increase in the chromatic frequency (below or above transition); and (v) a 'chromaticity jump' when the transition is crossed, i.e., not only do the signs of the transverse chromaticities need to be changed (because of the head–tail instability) but also the shape could be optimized (because of the possibility of a TMCI).

Furthermore, increasing the absolute value of the slip factor also helps in dealing with the longitudinal mode-coupling instability [16–18] and the fast single-bunch electron cloud instability [12, 15, 19]. Therefore, in most cases a very efficient way to push the performance of a machine near

transition to higher levels is either to go further away from the transition or to cross it as fast as possible (fast compared with the relevant instability rise times).

Finally, as was seen at the beginning of this paper, the operation of a synchrotron under isochronous or quasi-isochronous conditions can be attractive as it (naturally) achieves very short bunches, and for this reason it has been considered in several projects. This requires accurate control of the first high-order component α_1 of the momentum compaction factor (to provide the necessary momentum acceptance, with

$$C(\delta) = C_0\left[1+\alpha_0\,\delta\left(1+\alpha_1\,\delta+\alpha_2\,\delta^2+\ldots\right)\right], \tag{11}$$

where C_0 is the on-momentum circumference and $\delta = \Delta p/p$), as well as effective ways to damp all the collective instabilities.

References

[1] E. Métral and D. Möhl, Transition crossing, in Fifty years of the CERN proton synchrotron, Eds. S. Gilardoni and D. Manglunki, Vol. 1, CERN-2011-004 (2011), p. 59. http://dx.doi.org/10.5170/CERN-2011-004

[2] L. Rinolfi, Longitudinal beam dynamics (application to synchrotron), CERN/PS 2000-008 (LP) (2000).

[3] G. Plass (Ed.), Design study of a facility for experiments with low energy antiprotons (LEAR), CERN/PS/DL 80-7 (1980).

[4] K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, Singapore, 2006), p. 691.

[5] S. Andriamonje *et al.*, Neutron TOF facility (PS213): technical design report, CERN-INTC-2000-004 (2000).

[6] A. Chao, Beam instabilities in circular machines, these proceedings.

[7] A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).

[8] S. Aumon, High intensity beam issues in the CERN PS, CERN-THESIS-2012-261 (2012).

[9] B. Salvant, Impedance model of the CERN SPS and aspects of LHC single-bunch stability, CERN-THESIS-2010-087 (2010).

[10] G. Rumolo and F. Zimmermann, Practical user guide for HEADTAIL, CERN-SL-Note-2002-036-AP (2002).

[11] Y.H. Chin, User's guide for new MOSES version 2.0: mode-coupling single bunch instability in an electron storage ring, CERN/LEP-TH/88-05 (1988).

[12] G. Rumolo *et al.*, Simulation study on the energy dependence of the TMCI threshold in the CERN-SPS, CERN-AB-2006-075, Proc. EPAC'06, Edinburgh, UK (2006).

[13] S. Hancock and E. Métral, Ghost bunches and blow-up losses with high-intensity beams, PS/RF Note 2002-198 (2002).

[14] G. Arduini, H. Burkhardt, and E. Métral, Observation of a fast single bunch instability on protons in the SPS, CERN AB-Note-2003-093 (MD) (2003).

[15] H. Bartosik, Beam dynamics and optics studies for the LHC injectors upgrade, CERN-THESIS-2013-25 (2013).

[16] F.J. Sacherer, *IEEE Trans. Nucl. Sci.* **NS-24** (1977) 1393. http://dx.doi.org/10.1109/TNS.1977.4328955

[17] K. Ng, Potential-well distortion and mode-mixing instability in proton machines, FERMILAB-FN-630 (1995).

[18] E. Métral, Stability criterion for the longitudinal mode-coupling instability in the presence of both space-charge and resonator impedances, CERN/PS 2001-063 (AE) (2001).

[19] E. Métral, Effect of bunch length, chromaticity and linear coupling on the transverse mode-coupling instability due to electron cloud, CERN/PS 2002-009 (AE) (2002), Proc. ECLOUD02 Workshop, CERN Yellow Report CERN-2002-001, 211 (2002). http://dx.doi.org/10.5170/CERN-2002-001

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Observations and Diagnostics in High Brightness Beams

A. Cianchi
University of Rome Tor Vergata, Rome, Italy

Abstract
Brightness is a figure of merit largely used in light sources, such as free-electron lasers, but it is also fundamental in several other applications, for instance, Compton back-scattering sources, beam-driven plasma accelerators and terahertz sources. Advanced diagnostics is mandatory for the development of high brightness beams. 6D electron-beam diagnostics will be reviewed, with emphasis on emittance measurement.

Keywords
Electron beam diagnostics; emittance.

1 Introduction

Nowadays, a constant improvement of the characteristics of standard accelerators together with the development of new accelerators, techniques are opening completely new scenarios, paving the way to a complete revolution in the field of the accelerators.

The concept of electron-beam brightness was adopted from conventional optics, where beam brightness characterizes the quality of light sources. Electron-beam brightness is defined as current density per unit solid angle in the axial direction. While it is used for any kind of electron beam, from electron microscopes to free-electron lasers, it is very difficult to find a unique definition in the literature, especially because it is often confused with brilliance (see, for instance, Refs. [1], [2, p. 255], [3, p. 20], [4, p. 61], [5, p. 410], and [6, p. 73]). We follow the definition reported in Ref. [3]

$$B_n = \frac{2I}{\pi^2 \varepsilon_{nx} \varepsilon_{ny}} \tag{1}$$

where I is the beam current, and ε_{nx}, ε_{ny} are the transverse normalized emittance in the x- and y-directions, respectively. Brightness is measured in amps per square metre. Typical values of high brightness beams range between 10^{14} and 10^{16} A/m^2. This definition is sometimes called 5D brightness; when this quantity is divided by the energy spread, it is called 6D brightness, see, for instance, Ref. [7]. Although brightness is a figure of merit largely used in light sources, like free-electron lasers, it is, however, fundamental in several other applications, for instance Compton back-scattering sources, beam-driven plasma accelerators, and terahertz sources.

It is difficult to define high brightness from the point of view of diagnostics. It is important to focus on how the brightness is determined. Since brightness is the ratio between current and emittance, high values can be obtained by increasing the current or reducing the emittance or by combining these effects in different proportions. In addition, the meaning of 'brightness' has changed during the years. The LCLS, an X-ray free-electron laser, was planned to use bunches with a charge of about 1 nC, but today operates with bunch charges of 250 pC and sometimes as little as 20 pC, while the beam still preserves high brightness. The charge reduction, keeping charge density constant, decreases the emittance, and permits shorter bunches with higher currents. Using a 1 nC beam with a 1–2 mm-mrad normalized emittance means that conventional intercepting diagnostics can be difficult, especially if the beam is tightly focused. The beam can deposit enough energy on the device to destroy or severely damage it.

However, reducing the charge and squeezing the bunch down to 100 fs or even shorter opens new problems in the resolution of longitudinal diagnostics. So high brightness diagnostics covers a wide

https://doi.org/10.23730/CYRSP-2017-003.229

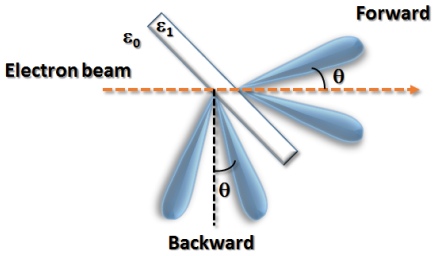


Fig. 1: Optical transition radiation emission. Emission is in both forward and backward directions

range of very different scenarios. In recent years, the growing interest for acceleration schemes based on plasma acceleration has even increased this range. In fact, in some of the proposed schemes there are two beams to be characterized at the same time: the beam injected into the plasma (in the case of external injection, the input beam) and the one leaving the plasma (the output beam, present in all of the schemes for diagnostics of self- [8] and external injection [9]).

2 Measuring beam size

Measurement of the beam dimensions is fundamental, to retrieve the value of the transverse emittance and to check proper beam matching through the machine. We are focusing here on high brightness machines, mainly linacs. Being single-passage devices they allow the use of intercepting diagnostics. Later, we will discuss the limits of using intercepting devices, but for most applications they are adequate.

A beam does not have a sharp edge, so it is quite difficult to define a clear boundary. To overcome this problem, the r.m.s. (root mean square) dimension is usually adopted. Let us assume a 1D intensity distribution, which can represent the projection of the beam profile on one axis, with i points each of intensity I_i. The definition of the second momentum of the distribution is

$$\left\langle x^2 \right\rangle = \frac{\sum\limits_i I_i \left(x_i - \bar{x}\right)^2}{\sum\limits_i I_i} \,. \tag{2}$$

Be aware that a point with $x = \bar{x}$ gives no contribution, even if its intensity is very large, while a point very far away from $\bar{x}$ gives a huge contribution, even for a small intensity. In Ref. [10], the reader can find several approaches to data processing in order to discriminate the beam from the surrounding halo.

We consider here only three types of monitor. A more exhaustive review of monitors for transverse profiles can be found in Ref. [11]. A metallic foil, or often a silicon aluminized wafer, placed at 45° with respect to the beam direction, is used as source of OTR (optical transition radiation), see Fig. 1.

The transition radiation is produced when a charge passes through the interface between two media with different refractive indices and is emitted in a narrow cone, with an angle of about $1/\gamma$, making this diagnostics unsuitable for low energy beams (less than tens of megaelectronvolts for electrons, for instance) where the radiation emission angle is too large to be efficiently collected by imaging optics. Also the intensity of the radiation is weak, with an efficiency of conversion between electrons and photons between 10^{-3} and 10^{-4} in the visible wavelengths. However, the prompt and linear emission, mainly arising because it is a surface effect, makes this device the best choice for longitudinal and transverse measurements, with high energy and bunch charges higher than several tens of picocoulombs.

It is worth mentioning that OTR diagnostics might fail even for high-energy electron beams, owing to coherence effects in the emission process (coherent transition radiation (COTR); see, for instance. Ref. [12]), i.e., when the bunch length (even locally, owing to some microbunching) is of the order of the observed wavelength, or shorter than it.

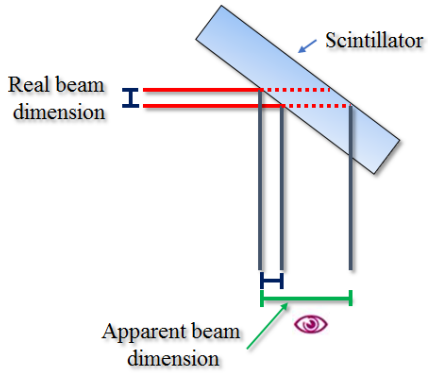


Fig. 2: Blurring due to transparent scintillator. The effect is highly exaggerated. The radiation is emitted all along the material, resulting in an apparent increase of the beam size.

Inorganic scintillators are widely used, owing to their large photon yield, even for low charges, and their high threshold radiation damage. Reference [13] provides an extensive review of their principal characteristics. While the resolution of the OTR monitor is mainly dominated by the radiation collection optics [14], in this case, the material itself presents a structure with grains that limits its value to the order of a few micrometres. Moreover this kind of radiator is transparent to its own radiation, resulting in an unwanted blurring effect that leads to an overestimation of the beam size (Fig. 2).

For this reason, the scintillators must be as thin as possible, of the order of 100 µm or even less. They can be used for beams of all energies, even with low charge. However, they experience saturation problems at high charges and their resolution is limited to the crystal grain size. Even scintillators can have trouble with COTR emission. It has only recently become possible to overcome this problem, using a new geometry [15] and playing with the difference between the isotropic emission of the crystal and the directional emission of the COTR.

Both OTR and scintillator screens are easy to implement and offer a 2D map of the charge distribution in a single shot. They are often called view screens. However, being intercepting devices, they scatter the beam all around, causing some problems, especially when high radiation levels are dangerous for the surrounding equipment. Also high-power density beams can seriously damage them, depositing too much energy.

To overcome these problems, a wire scanner is the popular solution. It is used in both circular and linear machines. It is basically a floating wire, moving transversally with respect to the beam orbit. When the beam hits the wire, a surrounding system of photomultiplier tubes detects a signal coming from the bremsstrahlung inside the wire material. By moving the wire scanner step by step, and recording signals in the photomultiplier tube, it is possible to correlate the beam intensity for every different spatial position of the wire. The wire enables the beam transverse size to be measured in the direction orthogonal to the wire itself, so the measurement is 1D and in multi-shots.

The main concerns in the design of wire scanners are mechanical stability of the mover and possible damage to the wire, see for instance, Ref. [16].

3 Emittance

The main parameter to be measured is the emittance.

Figure 3 shows two trace space diagrams. Remember that phase space diagrams have momentum on the y-axis, while the trace space diagram has the transverse angle. The area in the phase space is the Liouville invariant emittance, while the area in the trace space is just the geometrical emittance. This is the quantity that we measure. Because the beams do not have sharp edges and are very diffuse, the r.m.s.

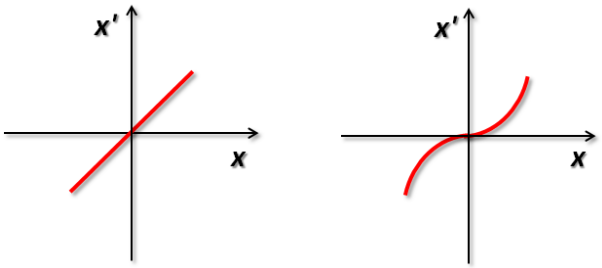


Fig. 3: Linear and quadratic correlation. Both figures have null area

emittance is used [17]. In both plots of Fig. 3 the area is zero, while the r.m.s. emittance is not always zero. The square of geometrical emittance is defined as

$$\varepsilon_{\mathrm{rms}}^2 = \langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2 \ , \tag{3}$$

and assuming a general correlation between x and x' such as $x = Cx'^n$, where C is a constant, the geometrical emittance can be rewritten as

$$\varepsilon_{\mathrm{rms}}^2 = C \langle x^2 \rangle \langle x^{2n} \rangle - \langle x^{n+1} \rangle^2 \ . \tag{4}$$

It is clear that if $n = 1$, even the r.m.s. emittance is zero, while this is not true for $n > 1$. So the r.m.s. emittance is more a figure of merit of the beam quality than the area of the trace space. For pure monochromatic beams, the normalized emittance is just $\beta\gamma$ times the geometrical emittance, where γ is the relativistic factor and β is the ratio between the particle speed and the speed of the light. Otherwise (see, for instance, Refs. [18] and [19]), there is also a contribution from the energy spread and the beam divergence. In fact, the Liouville invariant is the normalized emittance, defined as

$$\varepsilon_{\mathrm{n}}^2 = \langle x^2 \rangle \langle \beta^2 \gamma^2 x'^2 \rangle - \langle x \beta \gamma x' \rangle \ . \tag{5}$$

It has been demonstrated in Ref. [19] that this formula can be rewritten as

$$\varepsilon_{\mathrm{n}}^2 = \langle \gamma \rangle^2 \sigma_\varepsilon^2 \langle x^2 \rangle \langle x'^2 \rangle + \langle \beta\gamma \rangle^2 \left(\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2 \right) \ , \tag{6}$$

where σ_ε is the relative energy spread and the second term is the geometrical emittance multiplied by $\beta\gamma$. The first term is usually negligible in conventional accelerators, but it could be the leading one in some scenarios, such as in plasma-based accelerators. In such a case, both energy spread and angular divergence are needed, to measure the normalized emittance.

3.1 Emittance measurement with space charge

The envelope equation [4] in a drift space is

$$\sigma_x'' = \frac{\varepsilon_{\mathrm{n}}^2}{\gamma^2 \sigma_x^3} + \frac{I}{\gamma^3 I_0 \left(\sigma_x + \sigma_y \right)} \ , \tag{7}$$

where ε_{n} is the normalized emittance, σ_x and σ_y are the beam dimensions, I is the beam current and I_0 is the Alfvén current. When the term representing the space charge, the second term in this equation, is greater than the emittance term, the dynamic regime is said to be space charge dominated. Because the space charge contribution decreases as γ^2, this term is relevant only at low energies. However, as it arises from an internal pressure inside the beam, this term must be kept negligible to measure the emittance.

The most common technique is called the pepperpot technique. The principle is shown in Fig. 4. The beam is stopped or heavily scattered by a mask of a material of high atomic number, usually tungsten,

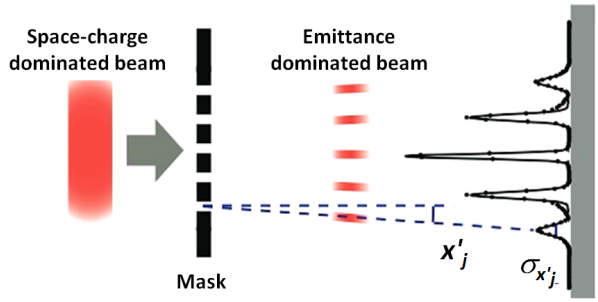


Fig. 4: Pepperpot technique: the beamlets emerging from the mask are emittance dominated

while some parts of the beam, called beamlets, pass through small holes or apertures. Since the charge is significantly suppressed, the beamlets are emittance dominated. After a proper drift, a scintillator screen produces an image of the beamlets. The relative intensity of each beamlet image is a direct measurement of the transverse beam distribution, $\langle x^2 \rangle$. In fact, it is a measurement of how much charge is entering any single hole; if the mask geometry is well known, it is a sampling of the beam transverse charge distribution. The width of each single spot gives a measurement of the beam divergence $\langle x'^2 \rangle$. A perfectly parallel beam produces images with dimensions equal to the hole size. The increase of this dimension is only due to the beam angular divergence, if the space charge contribution is negligible. Finally the mapping of the angular spread in different transverse positions allows the reconstruction of the correlation term between position and divergence,$\langle xx' \rangle$. In this way, we can measure the second term in Eq. (6), i.e., the geometrical emittance.

In implementing this diagnostic technique, there are some constrains to keep in mind. First of all, the hole dimensions must be large enough to allow some beam to pass through, to improve the signal-to-noise ratio, but at the same time they must ensure that the beamlets are emittance and not space charge dominated. The ratio between the two terms in the envelope equation is

$$R_0 = \frac{I\sigma_x^2}{2\gamma I_0 \varepsilon_{\mathrm{n}}^2} \,. \tag{8}$$

Assuming a uniform beam charge distribution in the holes, the r.m.s. is $\sigma_x = d/\sqrt{12}$, where d is the diameter of the hole. To ensure that the beamlets are emittance dominated, $R_0 \ll 1$. This constrains the value of d. In photoinjectors, the value of d is usually about $50 - 100$ µm.

The length L of the drift between the mask and the screen is another important parameter, in order to give to the beamlets enough space to develop a dimension greater than the hole size. The beam size at a distance L from the mask is

$$\sigma_x = \sqrt{(L \cdot \sigma_x')^2 + \frac{d^2}{12}} \,. \tag{9}$$

Since σ_x' is the parameter to measure, the first term in the square root must be much greater than the second one, setting the limit of the shorter acceptable drift L. In addition, the thickness l of the mask material is very important. It must be large enough to stop or heavily scatter the beam at a large angle (a critical issue at high energies), but a large thickness can limit the angular acceptance of the hole, which cannot be smaller than the expected angular divergence of the beam, i.e., $l \ll d/2\sigma_x'$.

It is worth mentioning that this technique not only enables the emittance and the Twiss parameters to be measured but can also be used to reconstruct the entire trace space. The holes are the sampling of the properties of such a space in different transverse positions. Figure 5 shows a comparison between simulated and measured trace spaces [20], demonstrating the effectiveness of this system.

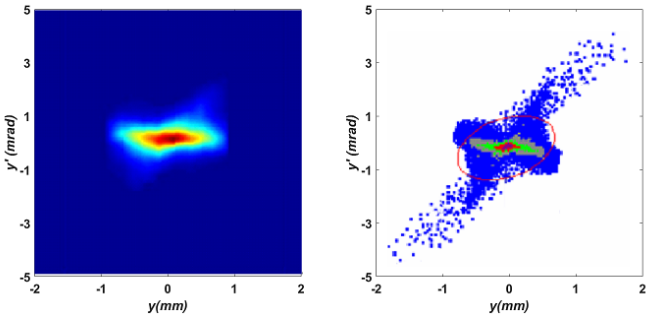


Fig. 5: Comparison between measured (left) and simulated (right) trace space [20]

3.1.1 Limits of this technique

Since the pepperpot technique is a sampling of the trace space, it can be affected by undersampling. Consider Fig. 6, in which two different trace spaces are compared [21].

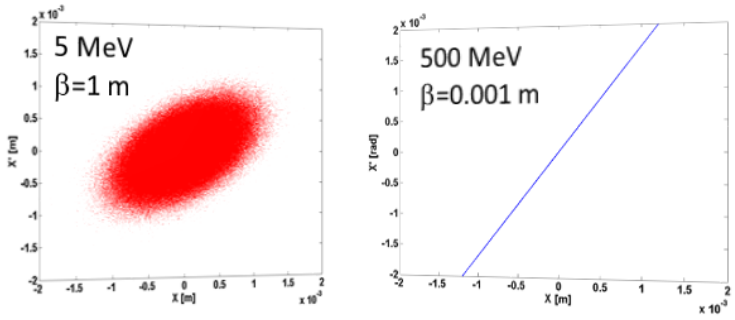


Fig. 6: Comparison between a typical trace space emerging from an RF photoinjector (left) and a possible trace space of a plasma-accelerated beam (right) [21].

The pepperpot measurement is a single-shot measurement and so is quite appealing from plasma-accelerated beams, where shot-to-shot instabilities makes a multi-shot technique unfeasible. However, the highly correlated trace space, proper to such beam types, severely affects the measurement resolution. The sampling error in the geometrical emittance has been established [22] is

$$\varepsilon_{\mathrm{err}} = \frac{2}{\pi}\left(x_{\max}\Delta x' + x'_{\max}\Delta x\right) \ , \tag{10}$$

where $x_{\max}$ is the beam size, $x'_{\max}$ is the maximum angular spread, Δx and $\Delta x'$ are the sampling intervals in position and angle. While in a conventional accelerated beam, as the energy increases, adiabatic damping decreases the values of the x and x', in a plasma-accelerated beam a significant angular spread could be present, even at high energy. In such a case, this error is larger than the measured emittance.

3.2 Emittance measurement without space charge

It has been demonstrated [23], both theoretically and experimentally, that when the space charge regime dominates, the pepperpot technique is the only method that can reliably measure the emittance. However, if the emittance term is the leading one in Eq. (7), another approach is followed.

The quadrupole scan [24] is the most used technique for measuring the emittance in such a regime. It relies on the measurement of the beam spot varying the strength of one or more quadrupoles. It is intrinsically a multi-shot measurement and, because it involves the use of magnetic elements, a large energy spread can spoil the emittance value [25]. There is an extensive treatment of this point in Ref. [26].

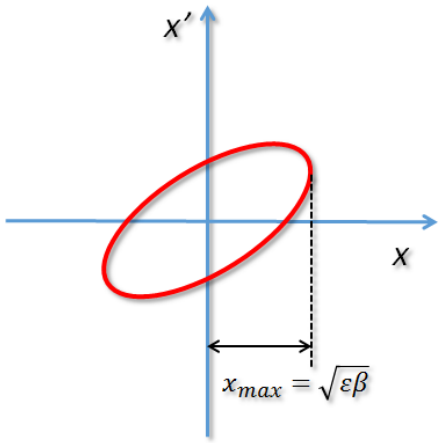


Fig. 7: Transverse trace space. The projection on the horizontal axis is the beam transverse size

Here, we want to focus on the main concepts. Consider two points in the transverse space, in different longitudinal positions along the machine, x and x_0, with their relative transverse momenta, x' and x'_0. The Courant–Snyder invariant must be the same in the two positions.

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = \varepsilon = \gamma_0 x_0^2 + 2\alpha_0 x_0 x_0' + \beta_0 x_0'^2 \,, \tag{11}$$

where α, β and γ are the Twiss parameters in x and α_0, β_0 and γ_0 in x_0 and M is a transport matrix between the two positions x and x_0 with the following elements:

$$M\left(x, x_0\right) = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} . \tag{12}$$

It has been demonstrated [27] that, using Eq. (12) in Eq. (11), the transport matrix for the Twiss parameters is

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & S'C + SC' & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix} . \tag{13}$$

Let us define a so-called σ matrix:

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \varepsilon \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix} . \tag{14}$$

Knowledge of the elements of this matrix allows a complete definition of the beam trace space. We focus on the σ_{11} element. It is equal to $\varepsilon\beta$. From Fig. 7 it is clear that it is just the square of the projection of the trace ellipse on the transverse axis. So this is a measurable quantity, being the beam size dimension, which can be retrieved using a view screen.

It is possible to demonstrate that the Courant–Snyder invariant in Eq. (11) can be rewritten as

$$\sigma_{11} x^2 + 2\sigma_{12} x x' + \sigma_{22} x'^2 = 1 \,. \tag{15}$$

Using Eq. (13), the transformation for the sigma matrix between x and x_0 can be written as

$$\sigma_1 = M \sigma_0 M^{\mathrm{T}} \,. \tag{16}$$

Following this transformation, the first element of the sigma matrix in the x position, i.e., σ_{11}^1, can be expanded as

$$\sigma_{11}^1 = C^2 \sigma_{11}^0 + 2SC\sigma_{12}^0 + S^2 \sigma_{22}^0 \,. \tag{17}$$

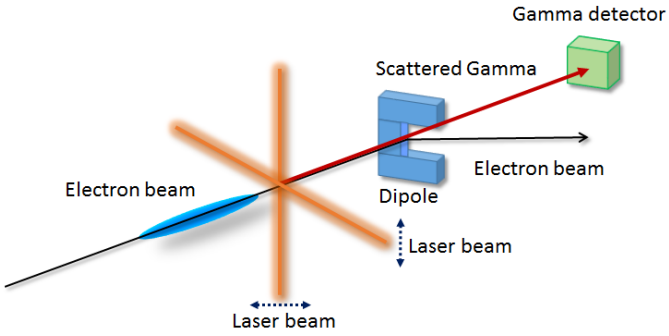


Fig. 8: Laser wire set-up. There is only a production of Compton photons when the electron beam intercepts the laser beams.

The square of the beam size is a linear combination of several elements, some of them well known, like the elements of the transport matrix C and S, and others unknown, as σ_{11}^0, σ_{12}^0 and σ_{22}^0, i.e., the elements of the beam matrix at the position where we want to know the beam trace space.

Three measurements in three different positions are sufficient to determine the unknown quantities, or alternatively at least three measurements changing the values of the beam transport matrix are needed, for instance, varying the field of a quadrupole to change the values of the elements of the transport matrix.

While there are examples of measurements in different positions (sometimes called multiscreen methods) and the reader can find interesting details about this technique in Ref. [28], the quadrupole scan method is widely used.

Between theory and practice, there is always a gap. In this scheme, sometimes the use of more than one quadrupole, to avoid strong defocalization of the beam in one plane, is necessary. However, it must be remembered that the more chromatic elements are used, the larger the effect on the emittance dilution will be [25]. The main error comes from the determination of the beam size [10]. The reader can find an excellent paper about the experimental uncertainties in such a measurement in Ref. [29].

While the pepperpot technique gives not only an emittance measurement but also a picture of the trace space, it is not the same for a quadrupole scan. To reconstruct the full trace space, more sophisticated strategies must be applied. The most common is tomography reconstruction, based on the reconstruction of an n-dimensional object starting from its $(n-1)$th dimensional projections. Several quadrupoles are needed to have a complete rotation of the trace space, so a dedicated beamline must be designed for such a task. This is why this system is rarely used. More details can be found in Ref. [30].

3.3 Non-intercepting diagnostics

The intercepting nature of this measurement could be a problem in the case of high repetition rate machines, or high charge beams, or if a shot-to-shot correlation between input and output beams in plasma wakefield accelerators is needed. Non-intercepting beam size measurements are not yet state of the art. Here, we report briefly three examples of promising techniques: laser wire, beam position monitor and diffraction radiation.

The laser wire [31] is a non-intercepting version of the wire scanner. Basically a tiny but intense laser wire is moved with respect to a beam, as shown in Fig. 8.

The Compton-scattered photons produced in the overlap between the laser wire and the electron beam, are collected by scintillators surrounding the vacuum chamber. Some tests are still ongoing but it a resolution in the submicrometre scale has already been demonstrated [32]. The main drawback of this technique is the intrinsic multi-shot nature and the stability of the laser wire alignment. Often, the whole chamber hosting the laser cavity is moved in front of the beam.

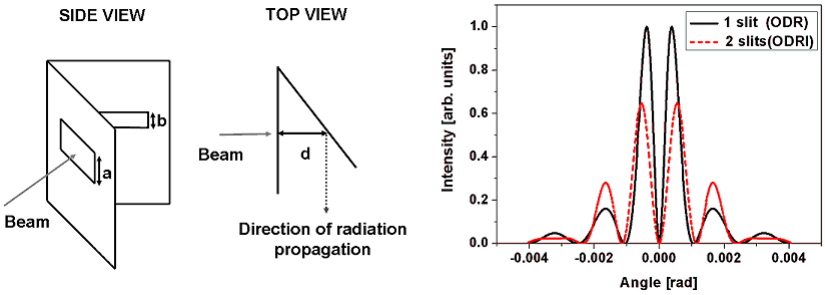


Fig. 9: Left: optical diffraction radiation interference (ODRI) set-up; right: angular distribution comparison between optical diffraction radiation (ODR) and ODRI.

It is possible to extract much more than the position of the beam from the signal emerging from a beam position monitor. It has been demonstrated that, by using a combination of signals coming from the different electrodes, it is also possible to retrieve the beam size [33]. A nice follow-up [34] demonstrated that this system is really interesting in circular machines, where there is a high repetition rate. Furthermore, Ref. [35] presents a refinement of this method, in which a beam position monitor is moved with a stepper motor at the micrometre scale. A recent paper [36] claimed that beam emittance could be measured using such a method. However, because the signal-to-noise ratio is very small, both the electromagnetic design of the electrodes and the mathematical treatment of the signal are fundamental.

When a charged particle passes through an aperture on a boundary between two media with different refraction indices, diffraction radiation is emitted in both the forward and backward directions. The diffraction radiation is emitted only when the dimension of the transverse electromagnetic field, at a given wavelength and beam energy, is larger than the aperture size. Since the beam passes through a hole, diffraction radiation provides a non-intercepting diagnostics tool, and is therefore well suited for measuring parameters of high charge density beams in a parasitic way.

The use of the angular distribution of the diffraction radiation as beam size monitor from a rectangular aperture was introduced in Ref. [37], while other authors have proposed a similar technique by using a circular aperture [38]. The choice of a rectangular slit shape has many advantages, e.g., mechanical machining and mathematical treatment, that were at the basis of the success of the first observation of diffraction radiation [39] as a diagnostics tool.

However, this result has also indicated some difficulties related to the experimental set-up; in particular, the low signal-to-noise ratio, mainly affected by the unavoidable synchrotron radiation background produced by the same beam in the upstream magnetic elements of the transport line and the requirement of an accurate and non-trivial control of the beam trajectory, owing to the ambiguity produced by a beam passing off-centre of the aperture.

The use of optical diffraction radiation interference (ODRI) [40], i.e. a two-slit system placed well inside the formation length (Fig. 9), can solve both of these problems; this technique was successfully used to measure the beam emittance [41].

4 Longitudinal parameters

Longitudinal measurements for high brightness beams are more established than transverse ones, except for the single-shot spectroscopic technique, as explained in the following. The main challenge of longitudinal diagnostics of high brightness beam is time resolution. Nowadays, we work with bunch lengths shorter than 100 fs, down to a few femtoseconds.

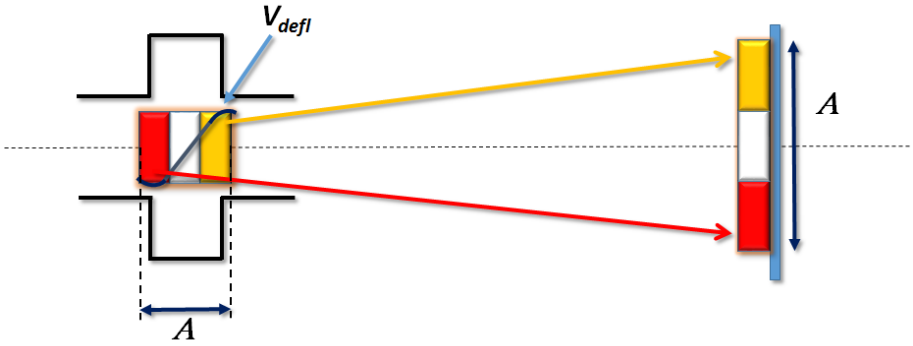


Fig. 10: Principle of operation of RF deflector. The longitudinal structure is mapped on the transverse profile because different longitudinal positions experience different transverse kicks.

4.1 Transverse deflecting structure

Transverse deflecting structures [42,43], sometimes also called RF deflectors, are powerful devices, able to attain resolutions of even a few femtoseconds in the X-ray band [44]. They are single-shot intercepting devices (for the measurement, but they need to be calibrated in multi-shot scenarios, as we'll see later), with the advantages of an inherent self-calibrating nature, simple implementation, and usage.

The working principle is shown in Fig. 10. A time-dependent transverse deflecting voltage is present in a standing or traveling wave structure. Different parts of the beam, in different longitudinal positions, explore a correlated force that imprints a transverse momentum on the bunch. After a drift, the imaging on a screen returns the longitudinal charge distribution. The change in the transverse momentum is $\Delta y' = qV/pc$, where V is the deflecting voltage, q is the charge, p is the momentum, and c is the speed of light. Assuming that the deflecting voltage is of the form $V = V_0 \sin(kz + \varphi)$, where $k = 2\pi/\lambda$ and λ is the RF wavelength, and considering that the bunch length is usually much smaller than such a wavelength (i.e., $kz \ll 1$), it is possible to expand the expression

$$\sin(kz + \varphi) = \sin(kz)\cos(\varphi) + \cos(kz)\sin(\varphi) \simeq kz\cos(\varphi) + \sin(\varphi) \ .$$

The expression of the change in the transverse momentum given by the structure is

$$\Delta y' = \frac{qV_0}{pc}\left[kz\cos(\varphi) + \sin(\varphi)\right] \ . \tag{18}$$

Using a phase value like $\varphi = 0$ or $\varphi = \pi$ nulls the second term, and the kick has a simple proportionality to the longitudinal position of the charge. Using a general expression of the transport matrix in terms of the Twiss parameter (see Ref. [45]) the value of the vertical displacement as a function of the longitudinal position is

$$y(z) = y_0 + \left(\sqrt{\beta\beta_0}\sin\Delta\right) y_0' \pm \frac{qV_0}{pc} kz \left(\sqrt{\beta\beta_0}\sin\Delta\right) \tag{19}$$

where β_0 and β are the Twiss parameters on the RF deflector and on the screen, respectively, y_0 is the displacement of the charge with respect the axis of the cavity, and Δ is the betatron phase advance between the RF deflector and the screen. The $\pm$ sign depends on which zero of the RF phase is chosen, $\sin(\varphi) = 0$ for both $\varphi = 0$ or $\varphi = \pi$. The second momentum of the distribution on the screen is given

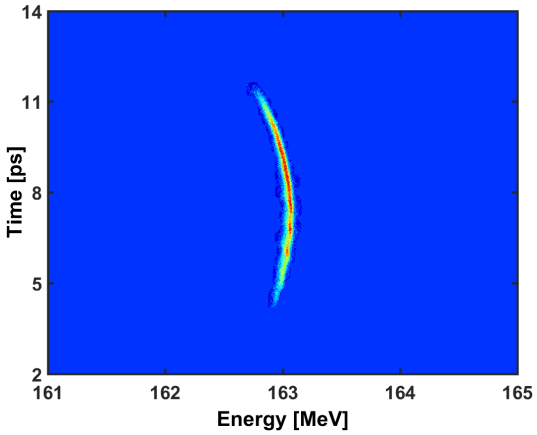


Fig. 11: Longitudinal trace space

by

$$\begin{aligned}\left\langle (y-\langle y\rangle)^2\right\rangle^{\pm} &= \left\langle y_0^2\right\rangle + \beta\beta_0 \sin^2\Delta \left\langle y_0{}'^2\right\rangle - \langle y_0\rangle^2 + \beta\beta_0 \sin^2\Delta \left(\tfrac{qV_0}{pc}k\right)^2 \left\langle z^2\right\rangle + \\ &-\beta\beta_0 \sin^2\Delta \langle y_0'\rangle^2 - \beta\beta_0 \left(\tfrac{qV_0}{pc}k\right)^2 \sin^2\Delta \langle z\rangle^2 + \\ &+2\sqrt{\beta\beta_0}\sin\Delta \langle y_0 y_0'\rangle \pm 2\sqrt{\beta\beta_0}\sin\Delta \left(\tfrac{qV_0}{pc}k\right)\langle y_0 z\rangle - 2\sqrt{\beta\beta_0}\sin\Delta \langle y_0\rangle \langle y_0'\rangle + \\ &\mp 2\sqrt{\beta\beta_0}\sin\Delta \left(\tfrac{qV_0}{pc}k\right)\langle y_0\rangle \langle z\rangle \pm 2\beta\beta_0 \sin^2\Delta \left(\tfrac{qV_0}{pc}k\right)\langle y_0' z\rangle + \\ &\mp 2\beta\beta_0 \sin^2\Delta \left(\tfrac{qV_0}{pc}k\right)\langle y_0'\rangle \langle z\rangle\end{aligned} \tag{20}$$

Usually $\langle y_0\rangle = \langle z\rangle = \langle y_0'\rangle = 0$.

When there is no power in the RFD, and using the former condition, we get

$$\sigma_0^2 = \left\langle y_0^2\right\rangle + \beta\beta_0 \sin^2\Delta \left\langle y_0'^2\right\rangle + 2\sqrt{\beta\beta_0}\sin\Delta \left\langle y_0 y_0'\right\rangle \tag{21}$$

It represents the dimension of the beam on the screen without any power in the RFD.

Several terms appear in 20 with different sign for a different choice of the phase φ. An average of the measurements of the second momentum with the two opposite phases φ=0 and φ=π is required in order to cancel them. Finally the dimension measured is

$$\sigma = \sqrt{\sigma_0^2 + \sigma_z^2} \tag{22}$$

where $\sigma_z^2 = \beta\beta_0 \sin^2\Delta \left(\tfrac{qV_0}{pc}k\right)^2 \left\langle z^2\right\rangle$

To increase the resolution of the device the term σ_0^2 must be much smaller than σ_z^2, so the spot on the screen with RFD off must be the smallest achievable. Looking at formula Eq. (19), we can see that other ways to improve the resolution are to increase the deflecting voltage (it is the value of the integrated voltage along the structure so, to some extent, increasing the device length can also work) or to decrease the RF wavelength.

One of the most interesting features of the RF deflector is its inherent self-calibrating nature. Basically, simply changing the phase by a precise amount and recording the variation in the centre-of-mass position on a view screen gives the conversion scale between pixels and phase, and thus also time.

The use of a dipole (i.e., a energy dispersive element) together with the RF deflector (i.e., a time-dispersive element) allows measurement of the longitudinal trace space in a single shot (Fig. 11).

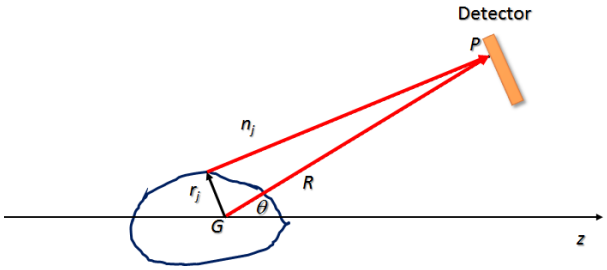


Fig. 12: Coordinate system

However, owing to the Panofsky–Wenzel theorem [46] the RF deflector introduces an energy spread that must be evaluated, to understand whether it is comparable or not with the slice energy spread [47].

4.2 Coherent radiation based measurement

A bunch charge can emit coherent radiation at a wavelength longer than its length. The characteristics of this radiation depend on the physical phenomena involved, such as synchrotron radiation, transition radiation, Smith–Purcell, and so on. Consider an electron bunch whose centre of mass is the origin of a coordinate system G, see Fig. 12.

Let us write R as the distance from the source to the detector and $\hat{n}_j$ the unit vector that selects the observation direction for the jth particle, $\vec{r}$ being the position vector of the jth electron relative to its bunch centre. We can express the electric field generated by the jth electron as

$$E_j = \mathrm{e}^{\mathrm{i}\frac{2\pi}{\lambda}(ct-\hat{n}_i\cdot\vec{r}_j)} \, .$$

If the distance from the source to the detector is much larger than the extent of the bunch σ_z, i.e., $R \gg \sigma_z$, the total field is given by the sum of the fields of each of the N particles and the total intensity can be written, in the limit of the Fraunhofer scalar theory, as

$$I_{\mathrm{tot}}\left(\omega\right) = I_{\mathrm{sp}}\left(\omega\right) \sum_{j,k} E_j E_k^* \, , \tag{23}$$

where $I_{\mathrm{sp}}(\omega)$ is the spectrum produced by a single particle. The analysis of the spectrum of such radiation can reveal the bunch longitudinal structure. Expanding the former equation, the spectrum intensity is given by

$$I_{\mathrm{tot}}\left(\omega\right) = I_{\mathrm{sp}}\left(\omega\right) \left[N + N\left(N-1\right) F\left(\omega\right)\right] \, , \tag{24}$$

where $I_{\mathrm{sp}}(\omega)$ is the spectrum produced by a single particle, N is the number of particles and $F(\omega)$ is the so-called form factor. The physics of the emitting process is contained inside $I_{\mathrm{sp}}(\omega)$, while the information about the bunch is in $F(\omega)$. The first term in the square bracket corresponds to incoherent emission and scales as the number of the particles, while the second term scales as N^2 and comes from the coherent emission. The form factor can be expressed as a function of the charge distribution $\rho(z)$, as

$$F\left(\omega\right) = \left| \int_{-\infty}^{\infty} \rho\left(z\right) \mathrm{e}^{\mathrm{i}\frac{\omega}{c}z} \mathrm{d}z \right|^2 \, . \tag{25}$$

The inverse transformation gives the searched value of $\rho\left(z\right)$

$$\rho\left(z\right) = \frac{1}{\pi c} \int_0^{\infty} \sqrt{F\left(\omega\right)} \cos\left(\frac{\omega}{c} z\right) \, . \tag{26}$$

Being an even transformation, the odd terms in the longitudinal distribution are not considered and a phase reconstruction technique must be used to retrieve phase information [48]. So the information about the longitudinal parameter is inside the form factor, which can be obtained from the power spectrum. The principles of Fourier transform spectroscopy [49, 50] were discovered by Michelson and Rayleigh, who identified that the interference pattern from a two-beam interferometer, obtained by altering the path difference between the two beams, is the Fourier transform of the radiation passing through the interferometer,

$$I(\omega) \propto \int_{-\infty}^{\infty} I(\delta) \cos\left(\frac{\omega\delta}{c}\right) \mathrm{d}\delta \,, \tag{27}$$

where δ is the path length difference between the two interferometer arms: $I(\delta)$ is called the interferogram. So to estimate the bunch length we need to measure the spectrum of the coherent radiation, and to do this we have to record the intensity of the radiation on a detector in an interferometer, varying the path difference between two arms of this device. Since the bunch lengths for high brightness beams are of the order of picoseconds or even shorter, the coherent radiation is usually in the terahertz frequency range. For such frequencies, the most used detectors are thermal detectors, like Golay cells and pyroelectric detectors.

A Golay cell [51] is a thermo-acoustic detector consisting of a small cell filled with a gas, typically xenon because of its low thermal conductivity, and a sensitive heat-absorbing film with low thermal capacity, which ensures a flat response to different frequencies. The frequency dependence is given, in principle, only by the properties of the window material used.

Radiation goes through the window and is absorbed by the film, warming it and the contained gas up. The gas therefore expands and the resulting pressure change modifies the shape of a flexible aluminized membrane at the back of the cell, which acts as a mirror. To convert the membrane movement into an electrical signal, a light-emitting diode illuminates the back of the flexible mirror. The reflected light is then focused onto a photocell.

Pyroelectric detectors produce a signal in response to a change in their temperature. Below a temperature T_{c}, known as the Curie point, ferroelectric materials (triglycine sulfate, lithium, tantalate) exhibit a large spontaneous electrical polarization. If the temperature is altered by an incident radiation, the polarization changes; if electrodes are placed on opposite faces of a thin dielectric, forming a capacitor, the change in polarization can be observed as an electrical signal. The process is independent of the wavelength of the incident radiation, resulting in a flat response over a wide spectral range.

The main advantages of Golay cells are high sensitivity and wide bandwidth in the millimetre range, while the main drawback is the slow temporal response (decay time of the order of tens of milliseconds). However, pyroelectric detectors, even exhibiting less sensitivity, have a faster response time; they are more robust and reliable than Golay cells and they are more widely used.

One of the main issues in this kind of measurement is the precise reconstruction of the entire spectrum, as there is a low frequency cut-off, owing to the vacuum pipe, the interferometer components and the detector acceptance. The overall transfer function of the whole system must be known, including the transfer function of the vacuum window and the transport line.

For all these reasons, a practical approach to these measurement is not to make the inverse Fourier transform the experimental data, but to fit the spectrum directly to guessed and well known distributions [52].

Longitudinal diagnostics with this technique began about 20 years ago and are is now widespread, but they have the big disadvantage of being multi-shot methods. However, some groups already performed single-shot measurements. In Ref. [53], Wesch *et al.* used a multi-stage spectrometer with a series of blazed reflection gratings. In any stage, wavelengths shorter than a threshold are dispersed while the longer wavelengths are reflected. With every grating focused on an array of detectors, it is

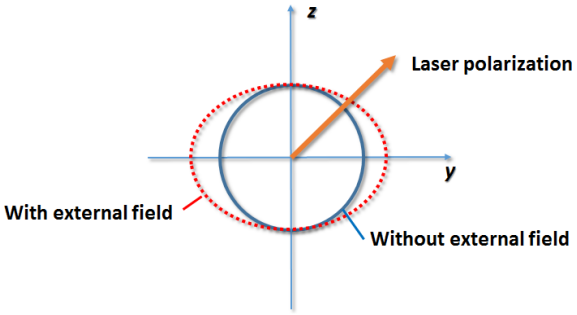


Fig. 13: Index ellipse for electro optic crystal. Without external field the index of refraction is the same in every direction. It is not true anymore with an external field.

possible to acquire in a single shot the whole spectrum for terahertz and infrared spectroscopy. The main challenge of such a device is the alignment of the various stages. To overcome this problem, in [54] Maxwell *et al.* used a single-stage single-shot device, based on a KRS-5 (thallium bromoiodide) prism. This material is quite interesting, even if it is hygroscopic, because it has a quite flat response between 0.8 and 40 µm. The radiation is dispersed and then sent to 128 lead zirconate titanate pyroelectric elements with 100 µm spacing line array.

The major advantage of these techniques based on the coherent radiation is the time resolution. The emitted wavelength scales as the bunch length. So the only requirement is a device sensitive to such wavelengths. For instance, 1 ps means 300 µm, while 1 fs needs a wavelength sensitivity of 300 nm, where a lot of detectors are available.

4.3 Electro-optic sampling

The electro-optic diagnostics technique is a non-destructive and single-shot method.

Large electric fields (of the order of megavolts per metre) applied to optically active crystals lead to the linear electro-optic effect (the Pockels effect), which is the basis of electro-optic detection techniques. In an isotropic medium, the polarization induced by an electric field is always parallel and linear with the electric field vector and related to the field by a scalar factor, the susceptibility. If the medium is anisotropic, the induced polarization is still linear but not necessarily parallel to the electric field, so the susceptibility is a tensor. In such a case, the crystal becomes birefringent, with different refraction indices along its principal axes.

The Coulomb electric field co-propagating with the relativistic electron bunch induces a time-dependent birefringence in an optically active crystal. A near infrared laser propagates along the crystal. Figure 13 shows the so-called ellipse index.

Without any external applied field, the crystal is isotropic and the refraction indices are equal along both axes. However, when the field is applied, the ellipse is distorted and the indices are different. The polarization of the external laser is in between these axes. Owing to the different refraction indices, the propagation speeds along these different axes are different, and this results in a rotation in the laser polarization. If we place an orthogonal polarizer before the crystal and another after it, we can observe the laser transmission only when there is an applied electric field. So the rotation angle encodes information about the field strength, which reflects the amount of the charge in the beam bunch that produces such a field.

There are mainly three schemes to obtain this information. The first is called spectral decoding [55].

As shown in Fig. 14, a femtosecond laser pulse is stretched in a picosecond long stretcher and linearly chirped, i.e., different wavelengths are placed in different longitudinal positions. A first polar-

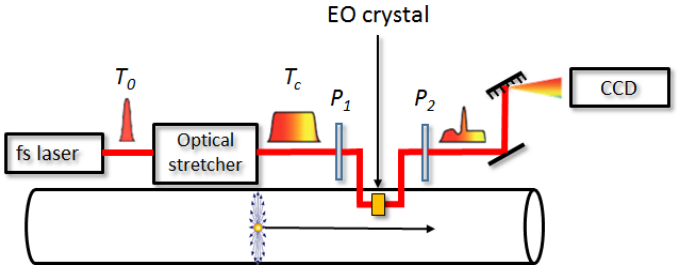


Fig. 14: Electro Optic Sampling spectral decoding scheme. A short pulse is stretched and sent in an electro optic crystal. The two polarizers P_1 and P_2 are orthogonal. Without any beam passing near the crystal there is no output through the second polarizer. Otherwise there is a polarization rotation and the longitudinal profile is encoded in the spectrum.

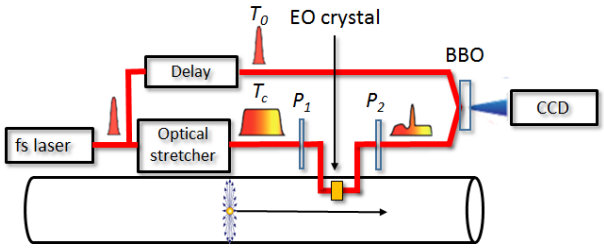


Fig. 15: Electro-optic sampling temporal decoding scheme [56]. BBO, barium borate; EO, electro-optic crystal; P_1,P_2, polarizers.

izer (P) selects a polarization. The laser crosses a non-linear crystal, in this case gallium phosphide, which is one of the most used for its large bandwidth. A second polarizer (A), orthogonal to the first one, allows the signal to cross it only when the polarization is rotated, which happens when the beam passes close (of the order of a few millimetres) to the crystal. A quarter-wave plate removes residual birefringence of the electro-optic crystal by minimizing transmitted light through the analyzer in the absence of an electron bunch. Owing to the linear chirp, different longitudinal positions in the probe laser pulse are rotated in different ways in the crystal, since refractive index is wavelength dependent. The information of the longitudinal beam profile is encoded into the spectrum and can be retrieved by a grating spectrometer. While this was the first implemented scheme, the time resolution is limited, owing to frequency mixing of the Fourier components of the electric field and it is about $T_{\rm lim} \approx 2.6\sqrt{T_0 T_{\rm c}}$ [57], a value in the range of some hundreds of femtoseconds.

The problem of frequency mixing is solved in the electro-optic temporal decoding scheme shown in Fig. 15. It is similar to spectral decoding but the reconstruction of the beam longitudinal profile is different. The intensity-modulated long laser pulse is measured in a single shot using a second short laser probe, realizing a cross-correlation with the principal pulse. The technique is based on the generation of second-harmonic light by crossing the two laser pulses in a non-linear crystal (e.g., barium borate) at an angle. The short gate pulse overlaps different temporal slices of the electro-optic pulse at different spatial positions of the barium borate crystal. Thus the temporal modulation of the electro-optic pulse is transferred to a spatial distribution of the second-harmonic light. The resolution of such a device can be about 40–50 fs, being limited mainly by the crystal absorption bands and the length of the laser probe. The main drawback is the low-efficiency second-harmonic process, requiring $\approx$1 mJ laser pulse energy.

Another single-shot technique is based on non-collinear propagation of the electric pulse and the laser pulse onto the electro-optic crystal, bending the crystal with respect to the beam axis propagation; this is called spatial decoding [58]. At any moment in time, the field in the electro-optic crystal overlaps with only a spatial fraction of the laser pulse. There is a direct correlation with the transverse spatial

position within the laser beam and the relative time delay between both pulses because different points across the transverse profile of the probe beam experiences an electric field at different instances in time, and the temporal image of the electric pulse can be impressed on the transverse spatial variation of the probe beam.

While the temporal resolution is, in principle, the same as temporal decoding, the advantages of this system lie in the simpler set-up and the possibility of avoiding a second-order harmonic generation. However, a very high surface uniformity of the crystal is required because in such a scheme the laser cannot be focused on the crystal, as in the other schemes, since the spatial size is proportional to the temporal window. This fact could be crucial in the future development of crystals with larger bandwidth, up to 20 THz, as well as better time resolution, like diethylaminosulfur trifluoride crystals [59], which so far appear to have too rough surfaces to be used for such a task.

Acknowledgements

I would like to thank Michele Castellano, Enrica Chiadroni, Andrea Mostacci, Riccardo Pompili and Luca Sabato for the useful discussions and the help in revising the document.

References

[1] C. Lejeune and J. Aubert, *Emittance and Brightness: Definitions and Measurements* (Academic Press, New York, 1980).

[2] A. Chao *et al.*, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1999).

[3] C.A. Brau, What brightness means, Proc. ICFA Workshop on the Physics and Applications of High Brightness Beams, Sardinia (World Scientific, Singapore, 2002).

[4] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, Weinham, 2008).

[5] S.-Y. Lee, *Accelerator Physics* (World Scientific, Singapore, 1999).

[6] J.A. Clarke, *The Science and Technology of Undulators and Wigglers* (Oxford University Press, Oxford, 2004).

[7] S. Dimitri and M. Cornacchia, *Phys. Rep.* **539** (2014) 1.

[8] W.P. Leemans *et al.*, *Nat. Phys.* **2** (2006) 696. http://dx.doi.org/10.1038/nphys418

[9] C.E. Clayton and L. Serafini, *IEEE Trans. Plasma Sci.* **24** (1996) 400.

[10] Mini Workshop on Characterization of High Brightness Beams, Zeuthen, 2008, https://indico.desy.de/conferenceTimeTable.py?confId=806#20080526.

[11] E. Bravin, in Proc. of the CAS-CERN Accelerator School: Course on Beam Diagnostics, 28 May–6 Jun 2008, Dourdan, France, edited by D. Brandt, CERN-2009-005 (CERN, Geneva, 2009), pp. 377, http://dx.doi.org/10.5170/CERN-2009-005.377

[12] S. Wesch and B. Schmidt, Summary of COTR effect, Proc. DIPAC11, Hamburg, 2011.

[13] G. Kube *et al.*, Resolution studies of inorganic scintillator screens for high energy and high brilliance electron beams, Proc. IPAC10, Kyoto, 2010.

[14] M. Castellano and V. Verzilov, *Phys. Rev. Spec. Top. Accel. Beams* **1** (1998) 062801.

[15] R. Ischebeck *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **18** (2015) 082802. http://dx.doi.org/10.1103/PhysRevSTAB.18.082802

[16] M. Sapinski, Model of carbon wire heating in accelerator beam, CERN-AB-2008-030-BI, CERN, Geneva (2008).

[17] J. Buon *et al.*, in Proc. of the CAS-CERN Accelerator School: 5th General Accelrator Physics Course, 7–18 September 1992, Jyváskylá, Finland, edited by S. Turner, CERN-1994-001 (CERN, Geneva, 1994), pp. 89-116, http://dx.doi.org/CERN-1994-001.116

[18] K. Floettmann, *Phys. Rev. Spec. Top. Accel. Beams*, **6** (2006) 034202.

[19] M. Migliorati *et al.*, *Phys. Rev. Special. Top. Accel. Beams* **16** (2013) 011302. http://dx.doi.org/10.1103/PhysRevSTAB.16.011302

[20] A. Cianchi *et al.*, *Phys. Rev.Spec. Top. Accel. Beams* **11** (2008) 032801. http://dx.doi.org/10.1103/PhysRevSTAB.11.032801

[21] A. Cianchi *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **720** (2013) 153. http://dx.doi.org/10.1016/j.nima.2012.12.012

[22] T. Ludwig *et al.*, *Rev. Sci. Instrum.* **65** (1994) 1462. http://dx.doi.org/10.1063/1.1144946

[23] S.G. Anderson *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **5** (2002) 014201. http://dx.doi.org/10.1103/PhysRevSTAB.5.014201

[24] M.G. Minty and F. Zimmermann, *Measurement and Control of Charged Particle Beams* (Springer, Berlin, 2003). http://dx.doi.org/10.1007/978-3-662-08581-3

[25] A. Mostacci *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **15** (2012) 082802.

[26] J. Rees and L. Rivkin, On measuring emittance and sigma matrices, SLAC-PUB-3305 (1984).

[27] H. Wiedemann, *Particle Accelerator Physics* (Springer, Berlin, 2013), Vol. 1.

[28] P. Castro, Monte Carlo simulations of emittance measurements at TTF2, Dt. Elektronen-Synchrotron DESY (2003).

[29] F. Löhl *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **9** (2006) 092802.

[30] D. Stratakis *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **9** (2006) 112801. http://dx.doi.org/10.1103/PhysRevSTAB.9.112801

[31] I. Agapov *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **10** (2007) 112801.

[32] L.J. Nevay *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **17** (2014) 072802.

[33] S.J. Russell, *Rev. Sci. Instrum.* **70** (1999) 1362.

[34] A. Jansson, *Phys. Rev. Spec. Top. Accel. Beams* **5** (2002) 072803.

[35] R. Assmann *et al.*, Use of movable beam position monitors for beam size measurements, Proc. EPAC, Wien, 20000.

[36] Z.-J. Wang *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **816** (2016) 171. http://dx.doi.org/10.1016/j.nima.2016.01.074

[37] M. Castellano *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **394** (1997) 275.

[38] R.B. Fiorito *et al.*, *AIP Conf. Proc.* **472** (1999) 725.

[39] P. Karataev *et al.*, *Phys. Rev. Lett.* **93** (2004) 244802.

[40] A. Cianchi *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **14** (2011) 102803.

[41] A. Cianchi *et al.*, *New J. Phys.* **16** (2014) 113029.

[42] P. Emma *et al.*, A transverse RF deflecting structure for bunch length and phase space diagnostics, LCLS Technical Note, 12 (2000).

[43] C. Behrens *et al.*, Measurement and control of the longitudinal phase space at high-gain free-electron lasers, Proc. 33rd International Free Electron Laser Conf., Shanghai, 2011.

[44] C. Behrens *et al.*, *Nat. Commun.* **5** (2014) 3762.

[45] J. Rossbach, in Proc. of the CAS-CERN Accelerator School: 5th General Accelerator Physics Course, 7-18 September 1992, Jyváskylá, Finland, edited by S. Turner, CERN-1994-001 (CERN, Geneva, 1994) pp. 17-88, http://dx.doi.org/10.5170/CERN-1994-001.17

[46] W. Panofsky and W.A. Wenzel, *Rev. Sci. Instrum.* **27** (1956) 967. http://dx.doi.org/10.1063/1.1715427

[47] D. Alesini, *Int. J. Mod. Phys A* **22** (2007) 3693. http://dx.doi.org/10.1142/S0217751X07037378

[48] Y. Shibata *et al.*, *Phys. Rev. E* **50** (1994) 1479.

[49] M.F. Kimmitt, *Far-Infrared Techniques* (Pion, London, 1970).

[50] M. Cuisenier *et al.*, J. Optics (Paris) **23** (1992) 179.

[51] H.A. Zahl and M.J.E. Golay, *Rev. Sci. Instrum.* **17** (1946) 511. http://dx.doi.org/10.1063/1.1770416

[52] A. Murokh *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **410** (1998) 452. http://dx.doi.org/10.1016/S0168-9002(98)00177-6

[53] S. Wesch *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **665** (2011) 40. http://dx.doi.org/10.1016/j.nima.2011.11.037

[54] T.J. Maxwell *et al.*, *Phys. Rev. Lett.* **111** (2013) 184801. http://dx.doi.org/10.1103/PhysRevLett.111.184801

[55] I. Wilke *et al.*, *Phys. Rev. Lett.* **88** (2002) 12. http://dx.doi.org/10.1103/PhysRevLett.88.124801

[56] B. Steffen *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **12** (2009) 032802. http://dx.doi.org/10.1103/physrevstab.12.032802

[57] J.R. Fletcher, *Opt. Express* **10** (2002) 1425. http://dx.doi.org/10.1364/OE.10.001425

[58] A.L. Cavalieri *et al.*, *Phys. Rev. Lett.* **94** (2005) 114801. http://dx.doi.org/10.1103/PhysRevLett.94.114801

[59] Y. Okayasu *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **16** (2013) 052801.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Numerical Methods I & II

K. Li
CERN, Geneva, Switzerland

Abstract
Numerical methods are a fundamental field of research and application in accelerator physics and beam dynamics. They are widely used to study intensity effects and limitations in accelerators and become more important still in regimes where simple analytical models break down and experiments cannot be easily conducted. In this course we will introduce some basic concepts of numerical methods used to study collective effects in circular accelerators. The course covers macroparticle models, the implementation of the beam dynamics and applications relevant to intensity effects and limitations. This includes basic tracking in the transverse and longitudinal planes, modelling of impedance effects and finally two-stream effects such as electron clouds.

Keywords
Collective effects; instabilities; macroparticle models; numerical methods; simulation.

1 Introduction and concepts

Numerical methods are a fundamental field of research and application in accelerator physics and beam dynamics. They are a very powerful tool to overcome limitations from simplifications inherent to most analytical approaches or the complications connected to many experimental studies.

The advantage of numerical modelling over analytical solutions is that it can usually be done with far fewer approximations in order to study problems which include many of the complications present in the physical reality. For these, usually analytical models are so complicated that they cannot be solved by analytical means or they fail to correctly model the physical reality in the first place. In comparison with experiments, numerical modelling allows full flexibility in parameters and set-up since basically anything that can be modelled can also be studied even beyond practical or technical limitations. Moreover, the level of diagnostics can be arbitrarily accurate, since nearly any observable can be exposed.

The main drawback of numerical models is that they have to be executed on computers and, depending on the complexity of the problem, they need to run for a long time in order for them to produce meaningful results. Therefore, the goal of a good numerical model is to capture as much of the physical reality that is necessary to understand the problem under investigation. At the same time the algorithms should be sufficiently easy and fast to provide solutions within a reasonable amount of time. If implemented correctly, numerical models are able to capture the majority of physical effects and allow for a vast range of diagnostics and data analysis to understand the underlying mechanisms that pose intensity limitations.

There are several different ways of numerically modelling beam dynamics and intensity effects. These range from time-domain or frequency-domain Vlasov solvers (MOSES [1] and DELPHI [2]), circulant matrix models (BimBim [3]) or beam envelope tracking algorithms (HOMDYN [4]), for example. In these lectures, we will investigate and focus on macroparticle models. Examples of modern implementations of macroparticle models are ECLOUD [5] or HEADTAIL [6], which was among the first to treat two-stream effects in a strong–strong manner. As we will see, macroparticle models are the most natural way of mapping the model of a physical particle beam onto a computer system. They generally provide an easy and straightforward platform for implementing almost any known physical effect

https://doi.org/10.23730/CYRSP-2017-003.247

in beam dynamics without having to move to a great level of abstraction. In that sense, they are ideally suited to tackle most of the problems encountered in beam dynamics and collective effects and also to introduce some of the basic concepts of numerical methods.

1.1 Macroparticle models and computer systems

Macroparticle models, obviously, make use of macroparticles. A macroparticle is a numerical representation of an ensemble of physical particles clustered together within a representative region in space. This is sketched in Fig. 1. Typical systems of physical particles such as a particle beam in the Large Hadron Collider (LHC) at CERN, for example, consist of more than 1×10^{14} particles. These cannot be practically represented on modern computer systems due to memory limitations. To overcome this limitation, physical particles are grouped together into a single macroparticle in such a way that one is left with roughly of the order of 1×10^{7} macroparticles. This is the dimensionality of systems that can be easily solved on conventional computer systems. In clustering physical particles into a discrete set of macroparticles, one of course has to take care about noise issues. Many effects are now less averaged out and therefore enhanced (consider a single macroparticle representing a cluster of 1×10^{6} physical particles receiving a kick by some element or device—all particles now receive the same kick and move in exactly the same manner, where in reality there would be some finite spread in the motion of each individual physical particle). To make sure the results are still physically relevant and not dominated by numerical noise, one has to perform convergence studies. For this, a numerical parameter is varied, in our example the total number of macroparticles, which essentially is a measure of the granularity of our system, and the results are checked. Obviously, the results should not depend on the choice of the numerical parameter. In this case we say that the result is converged and we accept it as output from a numerical simulation. The results may still be wrong but the error should not originate from numerical noise issues but rather from a systematically wrong modelling of the physical reality.

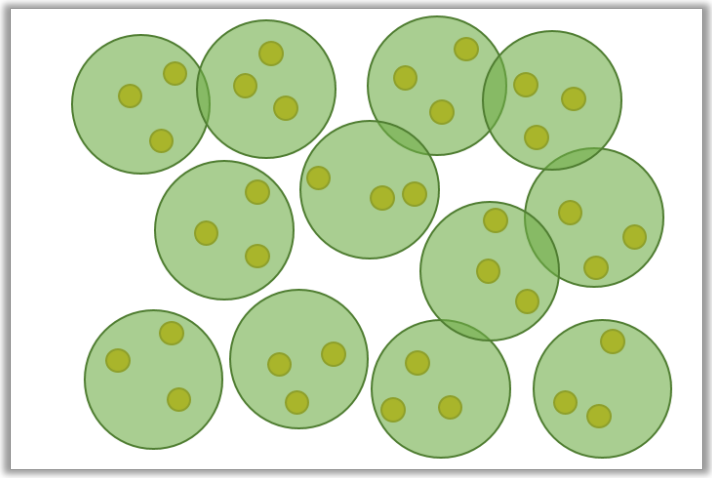


Fig. 1: Macroparticles (green) representing clusters of neighbouring physical particles (yellow)

Macroparticle models naturally map an accelerator–beam system onto modern computer systems. If we look at the main components of an accelerator–beam system we identify the physical particle beam and the accelerator with its elements and devices. If we think of some of the most fundamental components of computer hardware we can identify the main memory (RAM) and the CPU.

To fully describe the dynamics of a charged particle system, such as our physical particle beam, we need to know the generalized coordinates and the canonically conjugate momenta (i.e., the six phase-space variables) of each individual particle along with its charge and mass. In addition, we need to know the action of the different machine elements and devices on the particles, i.e., we need to know how

they modify the coordinates and momenta upon interaction of the particles with the machine elements. Numerically, the physical particles are represented by macroparticles and the machine elements are described by functions that update the coordinates and momenta of the macroparticles in a specific manner. Mapping this to the computer hardware, we recognize that the macroparticle system, which basically consists of a set of coordinates and momenta along with charges and masses for each macroparticle, can be represented by some allocated memory block in the main memory which contains and stores these numbers. The machine elements can be represented by functions that pass instructions to the CPU to update the numbers stored in this allocated memory block, thus generating an evolution of the coordinates and momenta of the macroparticle system and, hence, finally, describing beam dynamics. This correspondence is conceptually illustrated in Fig. 2.

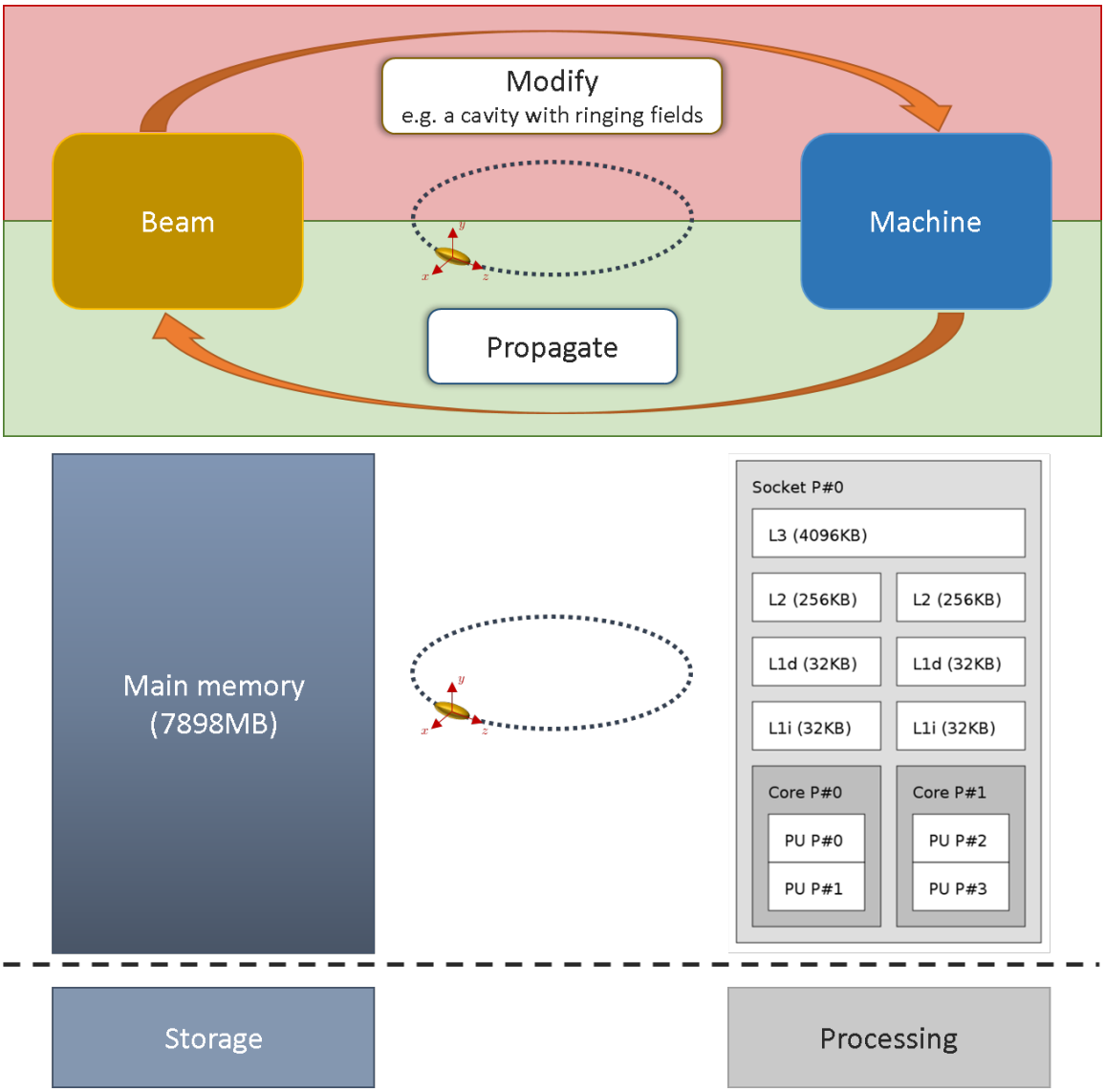


Fig. 2: The mapping of an accelerator–beam system onto the fundamental components of modern computer system hardware (display of the topological map of the system as obtained from *lstopo*).

1.2 Coordinate system and initialization

In a synchrotron the reference orbit curvature ρ and the reference momentum p_0 are related according to the synchronous condition via the dipole magnetic field B as

$$B\rho = \frac{p_0}{e}\,, \tag{1}$$

where e is the unit charge. In these lectures, we will choose generalized coordinates and canonically conjugate momenta to describe the dynamics of a charged particle system as illustrated in Fig. 3. These are, in the transverse plane, the horizontal and vertical positions x and y in metres with respect to the reference orbit of the machine and the respective angles x' and y'. In the longitudinal plane it is the position z in metres with respect to a position equivalent to the zero crossing of the RF fields (i.e., synchronous to an external RF clock) and the relative momentum offset δ with respect to the reference momentum. We name the phase space Γ; then this results in a set of six phase-space variables for every particle:

$$\left(\begin{pmatrix} x \\ x' \end{pmatrix}_i, \begin{pmatrix} y \\ y' \end{pmatrix}_i, \begin{pmatrix} z \\ \delta \end{pmatrix}_i\right) \in \Gamma, \quad i \in [1, \ldots, \text{particle number}]\,. \tag{2}$$

Fig. 3: The coordinate system used in these lectures

The first step in running a macroparticle simulation is to initialize a macroparticle system. For this we allocate a memory block in the main memory which is sufficiently large to accommodate all the relevant quantities of the macroparticle system, i.e., generalized coordinates and canonically conjugate momenta, charges and masses. Table 1 shows an example memory layout of a macroparticle system in the main memory where each of the six phase-space variables is allocated as an array of length equal to the macroparticle number which contains the corresponding values for every macroparticle. Then we fill this memory block randomly with numbers. Of course we do not do this entirely randomly. Typically, our macroparticle system will represent a particle bunch in a machine. As such, it should mimic the macroscopic statistical properties of the particle bunch such as the particle distribution function or the bunch size. A macroparticle system is therefore generated in a three-stage approach—first, we generate a distribution according to the particle distribution function, then we scale the distribution to match the bunch size and finally we distort the distribution without changing its r.m.s. size to match the local machine optics.

The particle bunch's intrinsic size is basically fully characterized by its emittance in the three planes, horizontal, vertical and longitudinal. The transverse and longitudinal emittances are defined as

$$\varepsilon_u = \gamma\beta\sqrt{\sigma_u^2\,\sigma_{u'}^2 - (\sigma_u\,\sigma_{u'})^2}\,, \quad u = x, y\,, \tag{3}$$

$$\varepsilon_z = 4\pi\frac{p_0}{e}\sqrt{\sigma_z^2\,\sigma_\delta^2 - (\sigma_z\sigma_\delta)^2}\,, \tag{4}$$

respectively, with

$$\sigma_b^2 = \left\langle (b - \langle b\rangle)^2 \right\rangle\,. \tag{5}$$

Table 1: Example memory layout of a macroparticle system in the main memory. Each of the six phase-space variables is allocated as an array of length equal to the macroparticle number N.

Count	0	1	2	...	N
x	...	...	...	...	...
x'	...	...	...	...	...
y	...	...	...	...	...
y'	...	...	...	...	...
z	...	...	...	...	...
δ	...	...	...	...	...

As a simple example we generate a Gaussian distribution in the horizontal plane with a given emittance. For this we simply need to create a Gaussian distribution using a Gaussian distribution random number generator, available in most standard libraries, in (x, x') with parameters [1]

$$\sigma_x = \sqrt{\frac{\varepsilon_x}{\gamma\beta}}, \tag{6}$$

$$\sigma_{x'} = \sqrt{\frac{\varepsilon_x}{\gamma\beta}}, \tag{7}$$

such that $\varepsilon_x = \gamma\beta\sigma_x\sigma_{x'}$. The resulting distribution is plotted in Fig. 4.

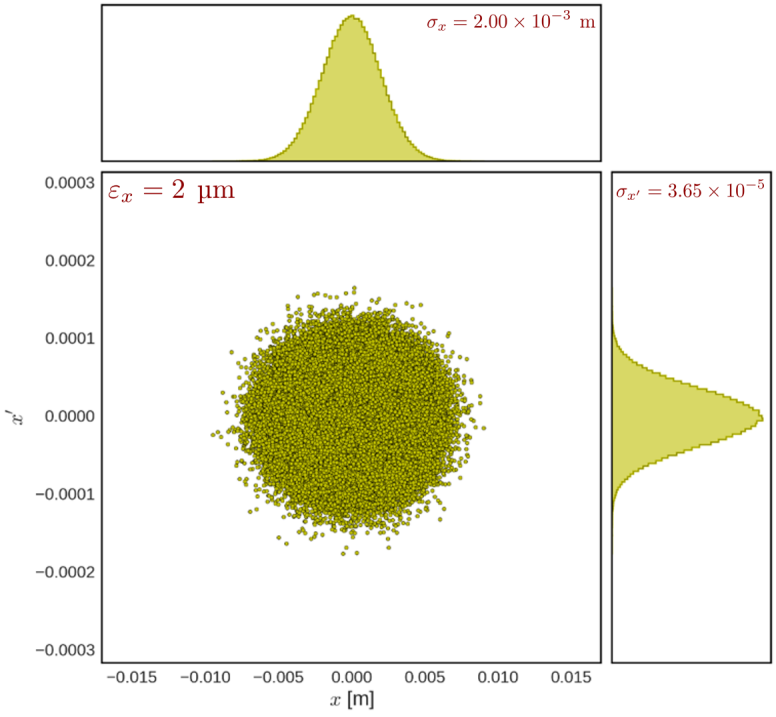


Fig. 4: Bi-Gaussian distribution in (x, x') with a normalized emittance of 2 µm

The distribution can now be matched to the local machine optics described by the optics functions (α_x, β_x) by scaling and correlating the coordinates and momenta according to

$$x_{\text{matched}} = \sqrt{\beta_x}\, x\,, \tag{8}$$

[1] The careful reader will realize an apparent mismatch in the dimensions in Eq. (8). In reality, this transformation emerges from the application of a matching section and there is another factor $\sqrt{1\ [\text{m}]}$ involved, which solves the dimension mismatch.

$$x'_{\text{matched}} = \frac{1}{\sqrt{\beta_x}}\, x' - \frac{\alpha_x}{\sqrt{\beta_x}}\, x\,. \tag{9}$$

We will not go into further details of creating non-linearly matched distributions such as for an RF bucket here.

Once all numbers have been generated, we are left with a numerical representation of the particle bunch as a macroparticle system. This manifests in the main memory as six individual arrays each with the length equal to the number of macroparticles in the macroparticle system. We will assume the charge and mass of all macroparticles to be identical such that they can each be described by two single floating point numbers. The particle bunch is now fully described by $(6 \times \text{macroparticle number} + 2)$ floating point numbers.

So far, we are able to generate a numerical representation of a particle bunch as a macroparticle system residing in the main memory in six arrays and two floating point numbers. No dynamics has been included yet. For this, we need to provide functions that modify the entries in each array in accordance with the machine elements that these functions are supposed to represent. This will be discussed in the following sections.

2 Modelling beam dynamics in absence of collective effects

The goal in the previous section was to abstract a physical particle system such that it can be numerically modelled. We implemented this by moving to a macroparticle system and representing it in the main memory as a set of arrays and numbers. Now we need to propagate this macroparticle system through the machine. For this, we need to identify machine elements and assign to them actions that they are to apply to the macroparticle system. This will be implemented as functions that update the entries of each of the respective arrays in accordance with the machine element that this function is supposed to represent.

As a first step we need to formalize this action so that we understand the beam dynamics and can come up with an algorithm that can be implemented on a computer. We can work in the framework of classical mechanics where we know that the entire dynamics of a particle system is fully contained in a single function, the Hamiltonian H, which is a function of the generalized coordinates and canonically conjugate momenta $[(x, x'), (y, y'), (z, \delta)]$ and an independent variable s along which the evolution is generated:

$$\left(\begin{pmatrix} x \\ x' \end{pmatrix}, \begin{pmatrix} y \\ y' \end{pmatrix}, \begin{pmatrix} z \\ \delta \end{pmatrix} \right) \in \Gamma, \quad H(x, x', y, y', z, \delta; s) \in f : \Gamma \to \mathbb{R}. \tag{10}$$

This takes place according to the Hamilton equations of motion

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}s} &= \frac{\partial H}{\partial x'}, & \frac{\mathrm{d}x'}{\mathrm{d}s} &= -\frac{\partial H}{\partial x}, \\ \frac{\mathrm{d}y}{\mathrm{d}s} &= \frac{\partial H}{\partial y'}, & \frac{\mathrm{d}y'}{\mathrm{d}s} &= -\frac{\partial H}{\partial y}, \\ \frac{\mathrm{d}z}{\mathrm{d}s} &= \frac{\partial H}{\partial \delta}, & \frac{\mathrm{d}\delta}{\mathrm{d}s} &= -\frac{\partial H}{\partial z}. \end{aligned} \tag{11}$$

We will now treat two simple Hamiltonians that generate betatron and synchrotron motion, which will allow us to propagate our macroparticle system along a circular accelerator in the absence of collective effects.

2.1 Transverse tracking

The basic elements in a circular accelerator that control the transverse motion are the dipoles to keep the beam on a circular orbit and the quadrupoles that provide focusing for off-orbit particles. The orbit

defined by the dipoles is given by the synchronous condition in Eq. (1). The focusing of the quadrupoles is described by the Hamiltonian

$$H(x,x') = \frac{1}{2}x'^2 + \frac{1}{2}K(s)x^2\,, \quad K(s) = K(s+C)\,. \tag{12}$$

The corresponding equation of motion is Hill's equation given as

$$x'' + K(s)x^2 = 0\,. \tag{13}$$

Equation (13) has the complication that it contains an s-dependent coefficient $K(s)$. Fortunately, this coefficient comes with the periodicity condition given in Eq. (12), which makes Eq. (13) actually solvable. An excellent treatment of this problem together with non-linear terms can be found in Ref. [7]. Here, we will focus on solving the linear problem.

The solution of any linear second order differential equation of the form (13) is uniquely determined by the initial values of x and its derivative x':

$$\begin{aligned} x(s) &= a\,x(s_0) + b\,x'(s_0), \\ x'(s) &= c\,x(s_0) + d\,x'(s_0). \end{aligned} \tag{14}$$

In matrix notation, this can be written as

$$\begin{pmatrix} x \\ x' \end{pmatrix}\Big|_s = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix}\Big|_{s_0} = M(s|s_0) \begin{pmatrix} x \\ x' \end{pmatrix}\Big|_{s_0}\,. \tag{15}$$

The matrix formulation is useful because it separates the properties of the general solution from those due to a specific initial condition. The matrix depends only on $K(s)$ and the length of the interval $(s - s_0)$. In addition, the matrix for any interval made up of subintervals is just the product of the matrices for the subintervals, that is,

$$M(s_2|s_0) = M(s_2|s_1)M(s_1|s_0). \tag{16}$$

Using the transfer matrix method together with the conditions of periodicity, stability and boundedness, one discovers that the eigenvalues of M are given as

$$\lambda = \exp\left(\pm \mathrm{i}\mu\right). \tag{17}$$

The matrix M satisfies (by construction, as it ultimately emerged from the Hamiltonian in Eq. (12))

$$M^T J M = J, \tag{18}$$

where J is the symplectic structure matrix

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{19}$$

Hence, the matrix M may be written in a form which exhibits the eigenvalues explicitly:

$$M = I\cos(\mu) + JA\sin(\mu), \tag{20}$$

where I is the identity matrix and A is a symmetric parameterized matrix

$$A = \begin{pmatrix} \gamma & \alpha \\ \alpha & \beta \end{pmatrix}. \tag{21}$$

The parameters (α, β, γ) are the Courant–Snyder parameters [8] and correspond to the optics functions. They are s-dependent in general and play a major role in determining the details of the motion. In particular, β determines the maximum local amplitude of transverse oscillations.

In applying Floquet's theorem, we insert a solution

$$x(s) = w(s)\exp\left(\mathrm{i}\psi(s)\right) \equiv \sqrt{2J_x\beta_x(s)}\cos(\psi_x(s) + \psi_{x,0}), \tag{22}$$

which again exhibits the eigenvalues of M explicitly, into (13) and find that the transfer matrix that propagates the coordinates and momenta (x, x') from one point s_0 with local optics function $(\alpha_0, \beta_0, \gamma_0)$ to another point s_1 with local optics function $(\alpha_1, \beta_1, \gamma_1)$ in the ring is written as

$$M(s_1|s_0) = \begin{pmatrix} \sqrt{\beta_1} & 0 \\ -\frac{\alpha_1}{\sqrt{\beta_1}} & \frac{1}{\sqrt{\beta_1}} \end{pmatrix} \begin{pmatrix} \cos(\Delta\mu_{0\to1}) & \sin(\Delta\mu_{0\to1}) \\ -\sin(\Delta\mu_{0\to1}) & \cos(\Delta\mu_{0\to1}) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_0}} & 0 \\ \frac{\alpha_0}{\sqrt{\beta_0}} & \sqrt{\beta_0} \end{pmatrix}. \tag{23}$$

Coming back to our original problem of numerically implementing linear transverse (betatron) motion, we can now see that this reduces basically to implementing a simple matrix multiplication function. The betatron motion is fully characterized by the machine optics functions. These need to be provided as an external input, either explicitly as a table obtained from an optics calculation (e.g. MAD-X) or, which is often done for simplification, using the smooth approximation where

$$\alpha = 0, \quad \beta = \frac{R}{Q} = \text{constant}\,. \tag{24}$$

Here, R is the ring radius and Q is the tune defined as the number of betatron oscillations per ring revolution:

$$Q = \frac{1}{2\pi}\oint \mu(s)\,\mathrm{d}s = \frac{1}{2\pi}\oint \frac{\mathrm{d}s}{\beta(s)}\,. \tag{25}$$

Once the optics functions are known, we can split the ring into segments and build the set of transfer matrices $M(s_j|s_i)$ that propagate macroparticles through subsequent segments (s_i, s_j) along the ring, as indicated for the two points s_0 and s_1 in Fig. 5. The numerical propagation of the macroparticle system through the machine is then performed by executing

$$\begin{aligned} \begin{pmatrix} x_k \\ x'_k \end{pmatrix}\Bigg|_{s_{i+1}} &= M_x(s_{i+1}|s_i) \begin{pmatrix} x_k \\ x'_k \end{pmatrix}\Bigg|_{s_i}, \\ \begin{pmatrix} y_k \\ y'_k \end{pmatrix}\Bigg|_{s_{i+1}} &= M_y(s_{i+1}|s_i) \begin{pmatrix} y_k \\ y'_k \end{pmatrix}\Bigg|_{s_i}, \\ i &\in [0, \ldots, \text{segment number} - 1]\,, \\ k &\in [0, \ldots, \text{macroparticle number} - 1]\,, \end{aligned} \tag{26}$$

repetitively for every segment and for every macroparticle in both the horizontal and the vertical planes.

In implementing linear, periodic and uncoupled betatron motion we simply need to generate and store the transfer matrices for every segment in the horizontal and the vertical planes, which amounts to a total of $(4 \times \text{segment number} \times 2)$ constant floating point numbers. The betatron propagation of the macroparticle system is computationally cheap and fast.

2.2 Longitudinal tracking

Longitudinal motion is a little more involved than the linear betatron motion described in the previous subsection. Longitudinal motion in circular accelerators, or synchrotron motion, has some peculiarities such as the presence of transition or the intrinsic non-linearities important particularly for hadron

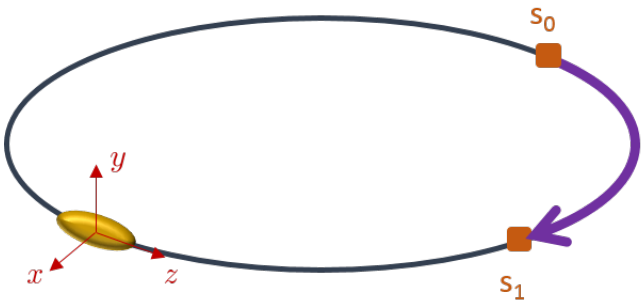


Fig. 5: A segment from a point s_0 to another point s_1 along the ring for which the betatron motion is described by the transfer matrix in Eq. (23).

machines where bunch lengths are comparable to the RF period of the accelerating cavities. A good introduction into the longitudinal dynamics can be found in Ref. [9].

Transition is a phenomenon that occurs in circular accelerators as a result of two competing effects. At low energies the relation between energy and momentum is nearly linear such that particles with slightly higher energy travel faster and hence for these particles the revolution period decreases, as one would naively expect. For high energies another effect becomes important, which results from the momentum compaction present in any circular machine with dispersion. What happens is that particles at high energies no longer gain in speed. However, due to dispersion along the machine they take longer paths and hence for these particles the revolution period increases. This effect is governed by the momentum compaction factor of the machine

$$\alpha_\mathrm{c} = \frac{1}{C} \oint \frac{D(s)}{\rho} \, \mathrm{d}s \,, \tag{27}$$

where C is the machine circumference, $D(s)$ the dispersion function and ρ the orbit curvature. The energy at which this behaviour switches, from lower to higher revolution periods with an increase in energy, is called the transition energy, defined by

$$\gamma_\mathrm{transition} = \frac{1}{\sqrt{\alpha_\mathrm{c}}} \,. \tag{28}$$

At transition itself all synchrotron motion is frozen.

Since a particle bunch in a circular machine naturally has a spread in energy it consequently also has a spread in revolution period. If simply left to drift, the particle bunch will eventually de-bunch along the machine. What is needed to keep particles bunched is some mechanism of focusing similar to the one that is provided in the transverse plane by the quadrupoles. This is accomplished in the longitudinal plane by the RF cavities. At each revolution, particles experience a kick from the integrated fields in the RF cavities. Depending on the delay of particles arriving at the RF cavities, they will gain more or less energy from the RF cavities. In that way the RF cavities keep the revolution periods close and thus provide focusing around what is called the synchronous phase φ_s.

Finally, the RF cavities also provide the necessary energy for acceleration. When ramping up the magnetic dipole fields, the revolution period of the entire particle bunch changes resulting in a shift of the synchronous phase. This leads to a net energy gain of the entire particle bunch and thus to acceleration.

The synchrotron motion together with all the resulting effects mentioned above are described by the Hamiltonian

$$H = -\frac{1}{2} \eta \beta c \, \delta^2 + \frac{e}{p_0 C} V_\mathrm{RF}(z) \,, \tag{29}$$

and V_{RF} usually takes the form

$$V_{\mathrm{RF}} = \sum_i \frac{V_i R}{h_i} \left(\frac{h_i z}{R} + \varphi_i \right) + \frac{\Delta E}{e} z \,. \tag{30}$$

Here, η is the slippage factor

$$\eta = \frac{1}{\gamma_{\mathrm{tr}}^2} - \frac{1}{\gamma^2} \,, \tag{31}$$

which we will discuss further below, $\beta = v/c$ is the velocity in units of the speed of light, p_0 is the reference momentum (1) and C is the ring circumference. Also, V is the integrated voltage along the RF cavities (this includes the transit-time factor), $h = \omega_{\mathrm{RF}}/\omega_0$ the harmonic number, which is the RF frequency in units of the revolution frequency, and ΔE the mean energy gain per turn.

If we assume a single harmonic RF system, from Eqs. (29) and (30) together with Eq. (11) the longitudinal equations of motion are derived as

$$\frac{\mathrm{d}z}{\mathrm{d}s} = -\eta \beta c \, \delta, \tag{32}$$

$$\frac{\mathrm{d}\delta}{\mathrm{d}s} = \frac{eV}{p_0 C} \left(\sin\left(\frac{hz}{R} \right) - \frac{\Delta E}{eV} \right) . \tag{33}$$

Equation (32) contains the slippage factor from Eq. (31) and exhibits the phenomenon of transition. Particles below the transition energy ($\eta < 0$) with positive momentum offset have a positive velocity, whereas above the transition energy ($\eta > 0$) a positive momentum offset leads to a negative velocity with respect to the direction of flight.

For small phases close to the synchronous phase, we may linearize the kick in Eq. (33) around the synchronous phase φ_{s} such that we obtain the linearized longitudinal equations of motion [2]

$$z'' + \underbrace{\frac{eV\eta h}{p_0 \beta c C R} \cos(\varphi_{\mathrm{s}})}_{\left(\frac{\omega_{\mathrm{s}}}{\beta c}\right)^2} z = 0 \,. \tag{34}$$

From Eq. (34), we identify the synchrotron tune as

$$Q_{\mathrm{s}} = \frac{\omega_{\mathrm{s}}}{\omega_0} = \sqrt{\frac{eV\eta h}{2\pi E_0 \beta^2} \cos(\varphi_{\mathrm{s}})} \,. \tag{35}$$

Defining

$$\alpha_{\mathrm{s}} = 0 \,, \quad \beta_z = \frac{|\eta| R}{Q_{\mathrm{s}}} = \text{constant} \,, \tag{36}$$

we can apply the same concept as for the linear betatron motion in the transverse plane, obtaining the one-turn transfer matrix

$$M = \begin{pmatrix} \sqrt{\beta_z} & 0 \\ 0 & \frac{1}{\sqrt{\beta_z}} \end{pmatrix} \begin{pmatrix} \cos(2\pi Q_{\mathrm{s}}) & \sin(2\pi Q_{\mathrm{s}}) \\ -\sin(2\pi Q_{\mathrm{s}}) & \cos(2\pi Q_{\mathrm{s}}) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_z}} & 0 \\ 0 & \sqrt{\beta_z} \end{pmatrix} , \tag{37}$$

and propagating the longitudinal phase-space variables (z, δ) of the macroparticle system by applying a simple matrix multiplication as in Eq. (26).

For modelling the general non-linear longitudinal motion we need to solve the system of equations (32) and (33) via a numerical integration algorithm. There are several numerical integration schemes

[2] Note that via this linearization, z is now the distance from the synchronous phase.

available (e.g., Runge–Kutta methods). In order to respect the Hamiltonian nature of the problem, it is important that we choose a symplectic integrator. We will not go into the details of symplectic integrators here but rather refer the reader to Refs. [10, 11]. An example of a reasonably simple and fast symplectic integrator of second order in the time step is the velocity-Verlet algorithm

$$\begin{aligned} z_{k,i+1/2} &= z_{k,i} - \frac{\eta C}{2}\,\delta_{k,i}, \\ \delta_{k,i+1} &= \delta_{k,i} + \frac{eV_{\mathrm{RF}}}{m\gamma\beta^2c^2}\sin\left(\frac{2\pi h}{C}\,z_{k,i+1/2}\right), \\ z_{k,i+1} &= z_{k,i+1/2} - \frac{\eta C}{2}\,\delta_{k,i+1}, \\ i &\in [0,\ldots,\text{turn number} - 1], \\ k &\in [0,\ldots,\text{macroparticle number} - 1]. \end{aligned} \tag{38}$$

The integration of the longitudinal equations of motion (32) and (33) for modelling the synchrotron motion is then applied once per turn for the full revolution as depicted in Fig. 6.

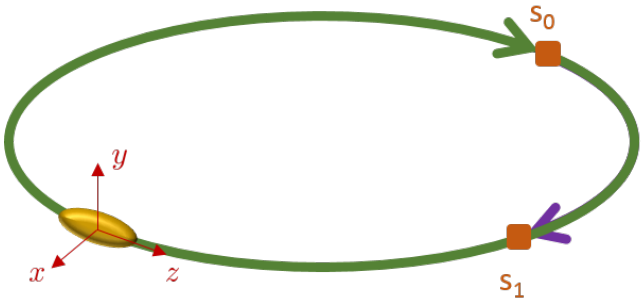


Fig. 6: The integration for the synchrotron motion according to Eqs. (38) is applied once per turn for the full ring

Once this algorithm has been implemented as a function, the numerical propagation of the macroparticle system through the machine is then performed by an iterative turn-by-turn application of Eqs. (38) for every macroparticle.

Figure 7 shows a simulation of a macroparticle bunch which is propagated through a ring applying both betatron and synchrotron motion. The horizontal (top) and longitudinal (bottom) phase spaces are shown. The non-linearity of the synchrotron motion leads to different tunes for different macroparticles depending on their longitudinal action. As a result, a filamentation of the longitudinal phase space takes place, which can be observed as the originally aligned colours spiral around the synchronous phase.

2.3 Modelling chromaticity and detuning with amplitude

We are now able to advance a macroparticle system through the machine by transverse and longitudinal tracking of the individual single macroparticles. The transverse tracking implemented so far models the linear betatron motion. For collective effects certain aspects of the non-linear transverse motion also have an important impact. These can appear in the form of chromaticity or detuning with amplitude, for example.

Chromaticity leads to a shift of the coherent beam spectrum and modifies the way the beam interacts with the machine impedance. This significantly impacts the behaviour of head–tail modes and instabilities as well as the transverse mode coupling instability (TMCI) (see Ref. [12]). Chromaticity results from different focusing of off-momentum particles. Thus, the quadrupole magnets, needed to provide the betatron focusing, introduce a natural chromaticity in the machine. Sextupole magnets

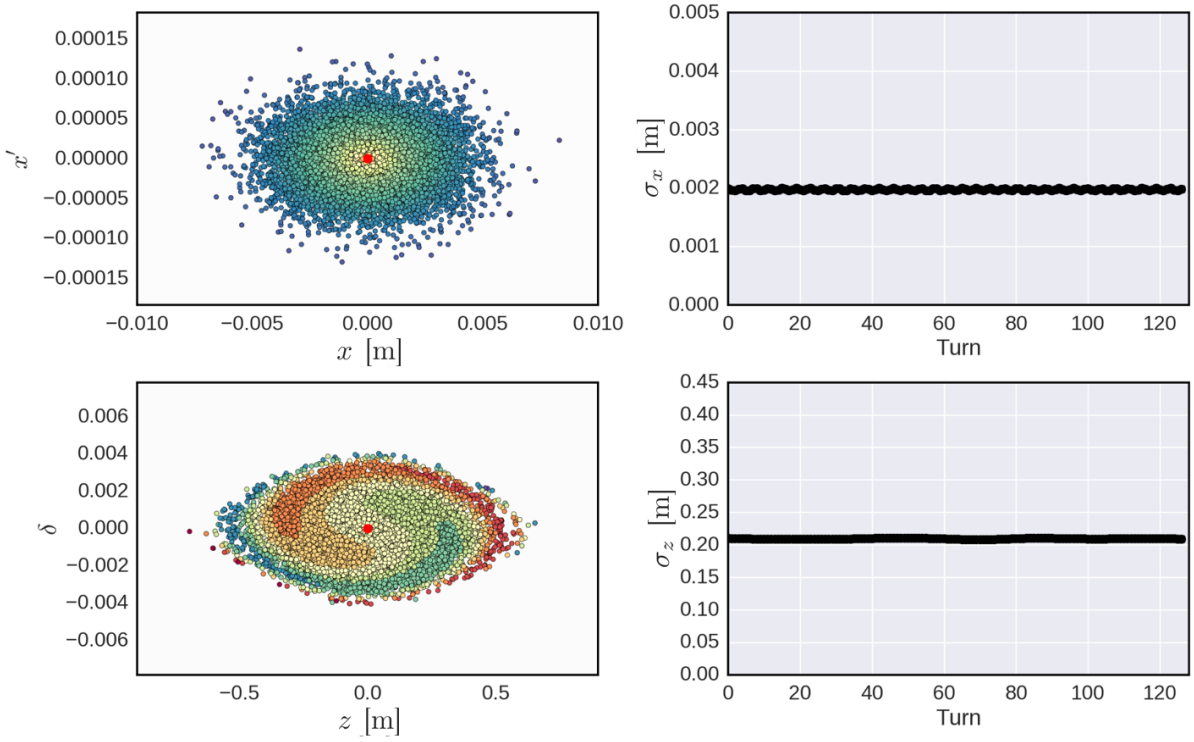


Fig. 7: A macroparticle bunch tracked through a ring applying betatron and synchrotron motion. The figure shows the horizontal (top left) and longitudinal (bottom left) phase spaces after several turns. Filamentation of the longitudinal phase space is clearly observable as the originally aligned colours spiral around the synchronous phase. The horizontal and longitudinal bunch sizes remain constant, as can be seen in the plots on the right, indicating that the bunch was well matched upon initialization to the machine optics.

placed in dispersive regions can then be used to control the chromaticity. Chromaticity is defined as

$$Q' = \frac{\mathrm{d}(\Delta Q)}{\mathrm{d}\delta}, \tag{39}$$

and can be easily calculated by writing down the (non-linear) Hamiltonian in action–angle variables as done in Ref. [7], for example. It may be written as

$$\begin{aligned} Q'_x &= -\frac{1}{4\pi} \oint \beta_x(s) \left(K_1(s) - K_2(s) D(s)\right) \mathrm{d}s, \\ Q'_y &= +\frac{1}{4\pi} \oint \beta_y(s) \left(K_1(s) - K_2(s) D(s)\right) \mathrm{d}s. \end{aligned} \tag{40}$$

Here, $K_i = B_i/(B\rho)$ are the effective (normalized) magnetic field strengths of the quadrupole $(i = 1)$ and sextupole $(i = 2)$ magnets, respectively.

Detuning with amplitude leads to a transverse tune spread within the beam, which can result in the suppression of instabilities via Landau damping. This is achieved for example with octupole magnets which provide a different focusing depending on the transverse offset of particles from the centre of the octupoles (which are usually placed on the reference orbit). Detuning with amplitude is described by the anharmonicities of the machine, which are defined as

$$\begin{pmatrix} \Delta Q_x \\ \Delta Q_y \end{pmatrix} = \begin{pmatrix} \alpha_{xx} & \alpha_{xy} \\ \alpha_{yx} & \alpha_{yy} \end{pmatrix} \begin{pmatrix} J_x \\ J_y \end{pmatrix}, \tag{41}$$

or consequently

$$\begin{aligned} \alpha_{xx} &= \frac{\mathrm{d}\Delta Q_x}{\mathrm{d}J_x}, \qquad \alpha_{xy} = \frac{\mathrm{d}\Delta Q_x}{\mathrm{d}J_y}, \\ \alpha_{yy} &= \frac{\mathrm{d}\Delta Q_y}{\mathrm{d}J_y}, \qquad \alpha_{yx} = \frac{\mathrm{d}\Delta Q_y}{\mathrm{d}J_x} \end{aligned} . \tag{42}$$

As for chromaticity, the anharmonicities can be easily calculated by writing down the (non-linear) Hamiltonian, now also containing octupole fields, in action–angle variables. They may be written as

$$\begin{aligned} \alpha_{xx} &= +\frac{3}{8\pi} \oint K_3(s)\beta_x(s)^2 \,\mathrm{d}s, \\ \alpha_{xy} &= -\frac{3}{4\pi} \oint K_3(s)\beta_x(s)\beta_y(s) \,\mathrm{d}s = \alpha_{yx}, \\ \alpha_{yy} &= +\frac{3}{8\pi} \oint K_3(s)\beta_y(s)^2 \,\mathrm{d}s. \end{aligned} \tag{43}$$

Here, $K_3 = B_3/(B\rho)$ is the effective (normalized) octupole field strength.

Chromaticity and detuning with amplitude can be implemented rather easily if we recollect that for a certain particle with momentum offset δ and transverse actions J_x and J_y, the detuning resulting from chromaticity and anharmonicities of the machine integrated over one turn is calculated as

$$\begin{aligned} \Delta Q_x &= Q'_x\delta + \alpha_{xx}J_x + \alpha_{xy}J_y \,, \\ \Delta Q_y &= Q'_y\delta + \alpha_{yx}J_x + \alpha_{yy}J_y \,. \end{aligned} \tag{44}$$

Hence, for a given segment (i, j) along the ring, the proportional detuning becomes

$$\begin{aligned} \Delta Q_{x,i\to j} &= \left(Q'_x\delta + \alpha_{xx}J_x + \alpha_{xy}J_y\right) \frac{\Delta\mu_{x,i\to j}}{2\pi\, Q_x}, \\ \Delta Q_{y,i\to j} &= \left(Q'_y\delta + \alpha_{yx}J_x + \alpha_{yy}J_y\right) \frac{\Delta\mu_{y,i\to j}}{2\pi\, Q_y}. \end{aligned} \tag{45}$$

We can pick up the transfer matrix in Eq. (23), which, when we include detuning effects from chromaticity and detuning with amplitude, becomes a separate transfer matrix for every individual single macroparticle:

$$\begin{aligned} M_k(s_1|s_0) &= \begin{pmatrix} \sqrt{\beta_1} & 0 \\ -\frac{\alpha_1}{\sqrt{\beta_1}} & \frac{1}{\sqrt{\beta_1}} \end{pmatrix} \begin{pmatrix} \cos(\Delta\mu_{k,0\to1}) & \sin(\Delta\mu_{k,0\to1}) \\ -\sin(\Delta\mu_{k,0\to1}) & \cos(\Delta\mu_{k,0\to1}) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_0}} & 0 \\ \frac{\alpha_0}{\sqrt{\beta_0}} & \sqrt{\beta_0} \end{pmatrix}, \\ \Delta\mu_{k,0\to1} &= \Delta\mu_{k,0\to1}(\delta_k, J_k) = \left(Q'\delta_k + \alpha J_k\right) \frac{\Delta\mu_{i\to j}}{Q}, \\ & k \in [0, \ldots, \text{macroparticle number} - 1]. \end{aligned} \tag{46}$$

Chromaticity and detuning with amplitude already lead to some macroscopic effects which are not yet collective effects (the potential does not depend on the particle distribution) but are nevertheless distinct for multiparticle systems. Examples are the decoherence from chromaticity or emittance blow-up due to filamentation from detuning with amplitude. These effects can be easily simulated with the tools we have implemented so far. An example of a simulation of decoherence and emittance blow-up in the presence of anharmonicities is shown in Fig. 8. It can be clearly seen how macroparticles with larger transverse actions have a smaller tune and thus start lagging behind in transverse phase space. This results in the characteristic spiralling of the macroparticle distribution in phase space, which translates into the observed decoherence and emittance blow-up. From the signal of the mean position in Fig. 8, we can compute the coherent spectrum, which is plotted in Fig. 9. The spectrum clearly features a finite spread around the tune, which in this case was at 20.13. Because we are running a macroparticle simulation we

have full access to all six phase-space variables of every macroparticle at any time. Equipped with the data for each individual macroparticle we are now even able to produce the incoherent spectrum. For this we perform a fast Fourier transform (FFT) analysis for each macroparticle to obtain their individual tunes in the horizontal and the vertical planes. We plot the obtained points in tune space (Q_x, Q_y) for every macroparticle to obtain the tune footprint of the macroparticle system. This is shown in Fig. 10. The tune footprint is an essential piece of information for the characterization of Landau damping, which will not be treated in further detail here. The interested reader is referred to Refs. [13–15].

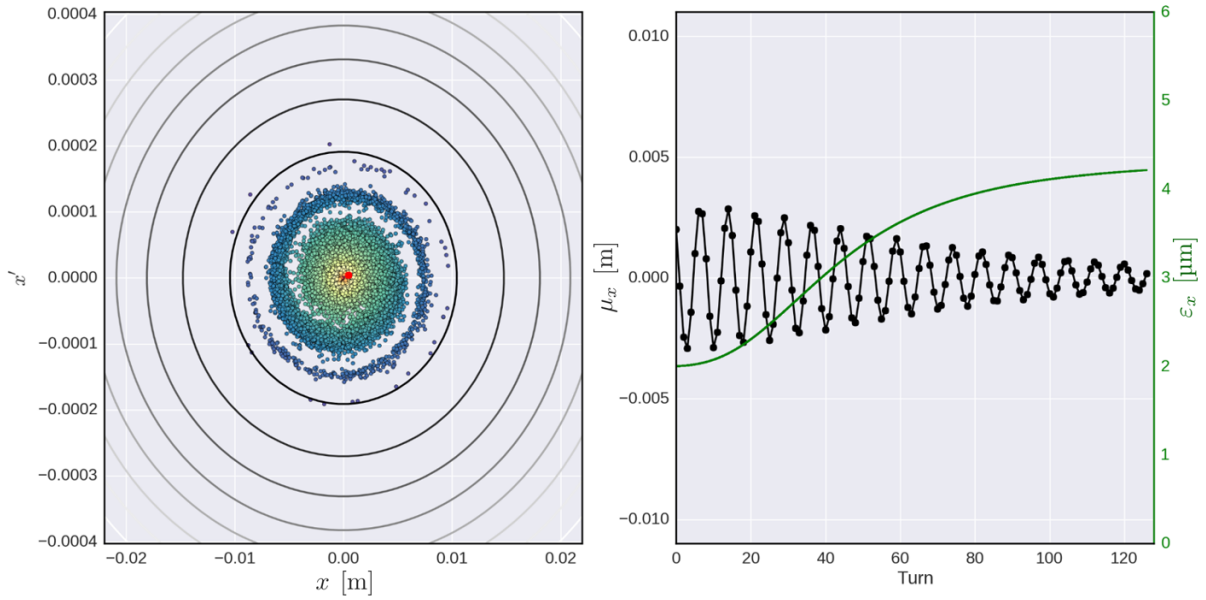


Fig. 8: Filamentation in the transverse phase space as a result of anharmonicities when the bunch enters with an initial offset (left). The mean position decoheres accompanied by an emittance blow-up, as can be seen in the plot on the right.

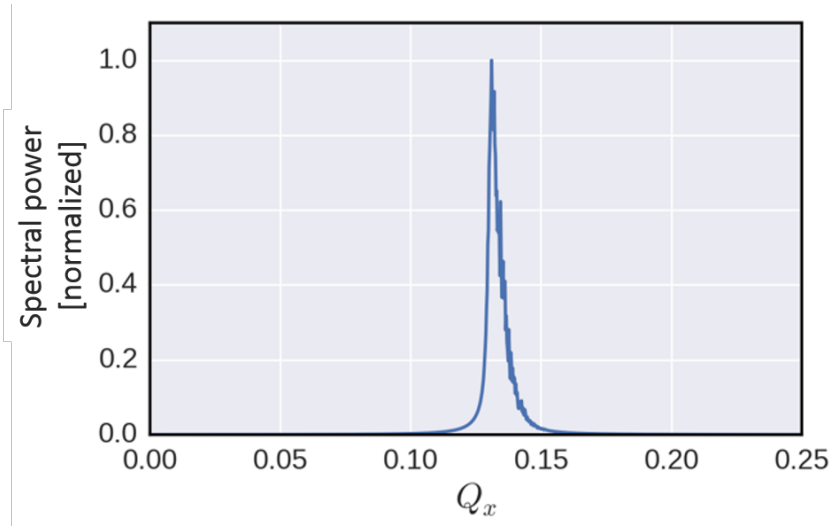


Fig. 9: The coherent spectrum obtained from the signal of the mean position. The spectrum features a finite spread around the tune (20.13).

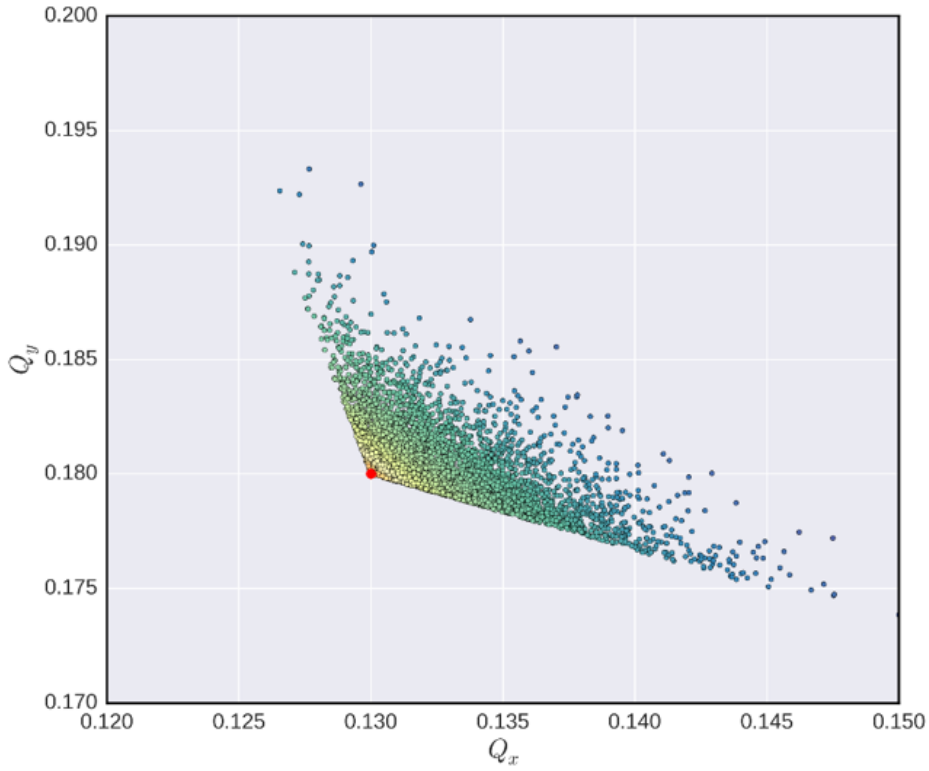


Fig. 10: The incoherent (fractional tune) spectrum obtained from the signal of each macroparticle. The nominal tune in this case was at $(20.13, 20.18)$. The darker points correspond to macroparticles with larger actions. The triangular shape of the tune footprint is characteristic for the detuning generated by octupoles, which were used for this simulation.

3 Modelling beam dynamics in the presence of wake fields

We have learned how to initialize a macroparticle system in accordance with the macroscopic statistical properties of a physical particle bunch and how to track this macroparticle system both transversely and longitudinally through a circular machine in absence of collective effects. So far, all propagation of the macroparticle system took place independently, i.e., macroparticles were tracked individually and independently from other macroparticles. Adding collective effects, tracking can no longer be performed in this manner. Collective effects add dependences to the macroparticle system such that macroparticles have to be tracked taking into account all other macroparticles present in the macroparticle system.

Physically, collective effects originate from machine elements and devices that store electromagnetic fields which have been excited by particles travelling through these elements or devices. The net electromagnetic field excited will depend on the exact distribution of the particles travelling through an element or device. Examples of these types of collective effects are the resistive wall effect coming from the resistivity of the walls of a vacuum chamber or trapped modes excited in a cavity-like structure. In all cases, particles passing through these objects leave back electromagnetic fields which then affect trailing particles. To model these effects one employs the concept of wake fields (or impedances).

Wake fields are the electromagnetic fields left behind a source particle travelling through an object and felt by a trailing or target particle. A such, they are functions of the distance between the source and the target particle. The concept of wake fields is illustrated in Fig. 11. Formally, a wake field is the electromagnetic response function of an object. This can be computed independently for any individual object either by analytical means for simple objects or by numerical means for more sophisticated structures. The wake function is an intrinsic property of any such object.

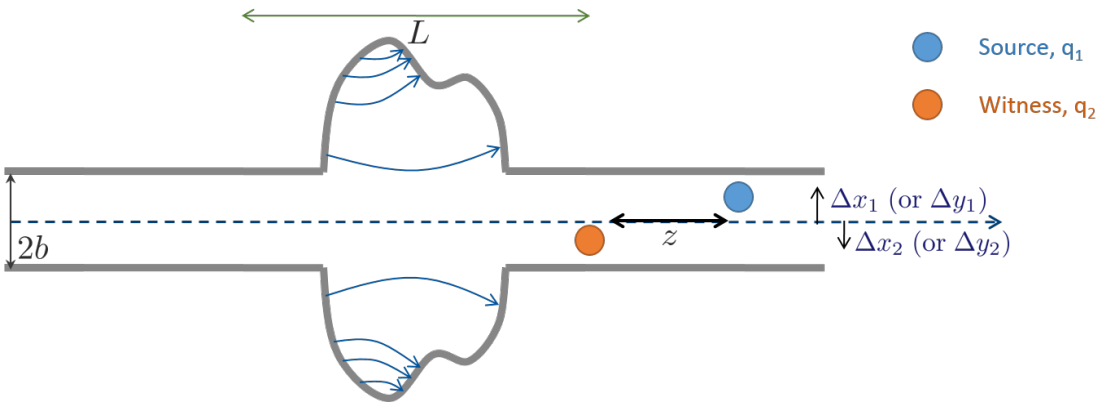


Fig. 11: The concept of wake fields. Wake fields are the integrated force from the electromagnetic fields left by a source particle (blue) and felt by a trailing or target particle (orange).

In the formalism to follow, we will restrict ourselves to the horizontal plane, but it is equally valid for both the horizontal and the vertical planes. Both planes must be included of course when one wishes to add coupling effects. The wake function w is defined via the integrated force along the effective length of a given object:

$$\int_{-L/2}^{L/2} F(z,s)\,\mathrm{d}s = qq_\mathrm{t}w(z). \tag{47}$$

Here, q and q_t are the charges of the source and target particles, respectively. The wake function in general is a function of the transverse positions of both the source and the target particles as well as of the distance between the two, so that

$$w(z) \to w(x, x_\mathrm{s}, z - z_\mathrm{s}). \tag{48}$$

Here, x and z denote the positions of the target particle. With the linearity of Maxwell's equations in the fields, the superposition principle holds and the integrated force generated not by a single particle but by a particle distribution travelling through an object is given as the convolution

$$\begin{aligned}\int_{-L/2}^{L/2} F(x,z,s)\,\mathrm{d}s &= qq_\mathrm{t}\int \rho(x_\mathrm{s},z_\mathrm{s})\,w(x,x_\mathrm{s},z-z_\mathrm{s})\,\mathrm{d}x_\mathrm{s}\,\mathrm{d}z_\mathrm{s}\\ &\equiv -\nabla V(x,z).\end{aligned} \tag{49}$$

We are now ready to write down the Hamiltonian for the simple betatron motion together with wake fields:

$$H = \frac{1}{2}x'^2 + \frac{1}{2}K(s)x^2 + \sum_k \frac{e^2}{m\gamma\beta^2c^2\,C}\iint \rho(x_\mathrm{s},z_\mathrm{s})\,w(x,x_\mathrm{s},z-z_\mathrm{s}-kC)\,\mathrm{d}x_\mathrm{s}\,\mathrm{d}z_\mathrm{s} + H_s(z,\delta) \tag{50a}$$

$$= \ldots + \sum_k \frac{e^2}{m\gamma\beta^2c^2\,C}\iint \rho(x_\mathrm{s},z_\mathrm{s})\sum_{mn} x^n x_\mathrm{s}^m\,W_{mn}(z-z_\mathrm{s}-kC)\,\mathrm{d}x_\mathrm{s}\,\mathrm{d}z_\mathrm{s} + H_\mathrm{s}(z,\delta) \tag{50b}$$

$$= \ldots + \sum_k \frac{e^2}{m\gamma\beta^2c^2\,C}\sum_{mn} x^n \int \lambda_m(z_\mathrm{s})\,W_{mn}(z-z_\mathrm{s}-kC)\,\mathrm{d}z_\mathrm{s} + H_\mathrm{s}(z,\delta), \tag{50c}$$

where

$$\lambda_m(z_\mathrm{s}) = \int \rho(x_\mathrm{s},z_\mathrm{s})x_\mathrm{s}^m\,\mathrm{d}x_\mathrm{s} \tag{51}$$

is the transverse moment of order m of the source distribution at z_s. We included multiturn wakes in the Hamiltonian above, where k is the turn number and C the ring circumference. Since the wake function is a function of the longitudinal coordinates we of course needed to include the longitudinal dynamics via the Hamiltonian $H_\mathrm{s}(z,\delta)$ to obtain a complete and self-consistent picture. In Eq. (50b), we made a Taylor expansion around the reference orbit and obtained the wake function of order mn. Here, m gives the order of the transverse moment of the source distribution and n gives the order of the resulting kick from the wake applied to the target particles. For most objects in practice, the wake functions that play a role are

$$\begin{aligned} &W_{01} : \text{constant wake},\\ &W_{11} : \text{dipole (driving) wake},\\ &W_{02} : \text{quadrupole (detuning) wake}. \end{aligned} \tag{52}$$

For these wakes, and omitting multiturn effects, the Hamiltonian simplifies to

$$H = \frac{1}{2}x'^2 \tag{53a}$$

$$+ A\left(\int [\rho(z_\mathrm{s})]\, W_{01}(z-z_\mathrm{s})\,\mathrm{d}z_\mathrm{s} + \int [\rho(z_\mathrm{s})\,\langle x_\mathrm{s}\rangle\,(z_\mathrm{s})]\, W_{11}(z-z_\mathrm{s})\,\mathrm{d}z_\mathrm{s}\right) x \tag{53b}$$

$$+ \left(\frac{1}{2}K(s) + A\int [\rho(z_\mathrm{s})]\, W_{02}(z-z_\mathrm{s})\right) x^2 + H_\mathrm{s}(z,\delta), \tag{53c}$$

where we set $A = e^2/(m\gamma\beta^2c^2C)$ and $\langle x_\mathrm{s}\rangle\,(z_\mathrm{s})$ is the mean horizontal position of the source distribution located at z_s. From this set of equations we can readily identify the effect of the different types of wakes on the beam dynamics. The term (53b) is linear in x and thus describes dipolar kicks changing the particle orbit. We can see that a constant wake produces a change of orbit depending on the source charge, whereas a dipolar wake creates an orbit distortion that depends on the mean source orbit offset. This feature of the dipolar wake is actually what enables it to drive beam instabilities, which is why it is sometimes also called the driving wake. The term (53c) produces a change in tune depending on the source charge. For this reason quadrupolar wakes are sometimes also called detuning wakes. We will discuss the different wakes and their impact on beam dynamics further below in parallel with simulations. But first we will move to the numerical implementation of wake field effects.

Going back to the definition of the wake fields in Eq. (49), we see that for a macroparticle system, which is essentially a collection of discrete individual macroparticles, the wake field kick, i.e., the change in momentum, from a constant, dipole or quadrupole wake field for an individual macroparticle is easily calculated via the sum over all macroparticles:

$$\Delta x'_k = \begin{cases} -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j W_{01}(z_k - z_j), \\ -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j W_{11}(z_k - z_j)\, x_j, \\ -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j W_{02}(z_k - z_j)\, x_k, \end{cases} \tag{54}$$

$$j,k \in [0,\ldots,\text{macroparticle number} - 1].$$

If we denote by N the macroparticle number, this results in N^2 operations and in particular in N^2 wake function evaluations, which can be computationally very expensive.

The computation can be made much faster by a simple numerical trick, which is similar to the concept of macroparticle models itself. As long as the wake function is sufficiently smooth and does not vary strongly within a given interval $[z_i, z_k]$, we can further discretize a macroparticle system into a set of longitudinal slices. We assume the wake function to be constant within a given slice. In this case, it

is sufficient to evaluate the wake function only once for all macroparticles within this slice. Instead of having to do a macroparticle-to-macroparticle evaluation we now only need to perform a slice-to-slice evaluation of the wake function. Equations (54) thus become

$$\Delta x'_k[i] = \begin{cases} -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j N[j]\, W_{01}(z[i]-z[j]), \\ -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j N[j]\, W_{11}(z[i]-z[j])\, \langle x\rangle[j], \\ -\dfrac{e^2}{m\gamma\beta^2c^2}\displaystyle\sum_j N[j]\, W_{02}(z[i]-z[j])\, x_k[i], \end{cases} \tag{55}$$

$$i,j \in [0,\ldots,\text{slice number} - 1],$$
$$k \in [0,\ldots,\text{macroparticle number} - 1].$$

Here, $\Delta x'_k[i]$ is the change in momentum for a macroparticle in slice i, $N[j]$ is the number of macroparticles in slice j, $z[i]$ is the centre longitudinal position of slice i, $\langle x\rangle[j]$ is the mean horizontal offset of slice j and $x_k[i]$ is the horizontal position of a macroparticle in slice i.

If we denote by n the slice number, the number of evaluations now reduces to n^2. If we consider that typically we have of the order of several thousand macroparticles per slice, this means that the number of evaluations is reduced by six orders of magnitude compared to the macroparticle-to-macroparticle evaluation. Furthermore, if the discretization of the macroparticle system is done longitudinally uniformly such that all slices have the same width, the number of slice-to-slice distances, and therefore the necessary wake field evaluations, reduces to $2 \times n - 1$. The wake function can be pre-calculated for all those distances and the values can be stored in an array. In this case the numerical algorithm to be performed for applying wake field kicks to a macroparticle system would look as illustrated in Fig. 12.

Figure 12 assumes a macroparticle bunch encountering a constant horizontal wake function W_{Cx} (i.e., a W_{01}-type wake function in Eq. (55)). Axis A) shows the macroparticle bunch and axis B) the wake function as it will be seen by the individual slices. Axis C) plots the number of macroparticles in each slice and axis D) finally depicts the wake field kicks that will be applied to each slice. The longitudinal slices are rendered on all axes as the coloured bars, where each coloured bar corresponds to a slice. The colours in axes C) and D) match in the sense that each coloured slice on axis C) will receive the wake kick from the corresponding coloured slice on axis D).

The formula at the bottom of Fig. 12 demonstrates how a function that models wake field kicks would have to be implemented. The number of macroparticles per slice N for the n slices and the wake function W_{Cx} for each of the $2 \times n - 1$ slice-to-slice distances from $[-L, +L]$ are pre-computed and stored in two arrays. The wake field kicks are then computed according to this formula (or, similarly, to Eq. (55)) for all macroparticles in a slice by updating all the angles of those macroparticles.

We will now go through some examples of numerically modelled wake field effects.

3.1 Constant wake fields

Constant wake fields are generated by structures with left–right or top–bottom symmetry. They create a closed orbit distortion similar to the potential well distortion that exists in the longitudinal plane. The Hamiltonian which includes constant wake fields W_{Cx} is given as

$$H = \frac{1}{2}\, x_i'^2 + \frac{1}{2}K(s)\, x_i^2 - \frac{e^2}{m\gamma\beta^2c^2C}\sum_{j=0}^{n-1} N(z_j)\, W_{Cx}(z_i - z_j)\, x_i + H_{\mathrm{s}}(z,\delta). \tag{56}$$

Here, x_i is a macroparticle in slice i, $N(z_j)$ is the number of macroparticles in slice j, z_j is the centre longitudinal position of slice j and n is the number of slices. The wake field term in the Hamiltonian is linear in x_i. Thus, it generates an orbit shift (like dipoles would).

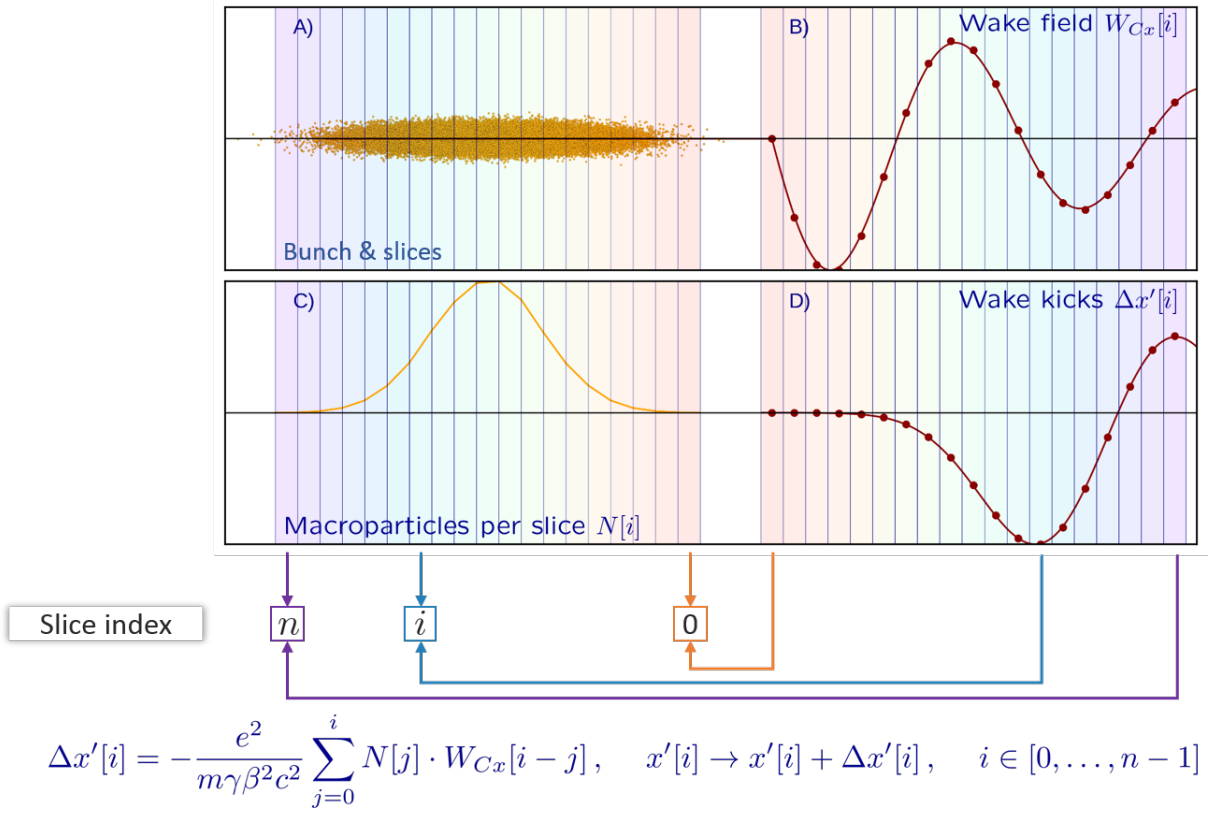


Fig. 12: Sketch of the algorithm to compute and apply wake field kicks. The coloured bars correspond to, and identify via their colour, the individual longitudinal slices. A) shows the macroparticle bunch. Axis B) shows the wake function (marker value) as it will be seen by the colour-corresponding slices. Axis C) plots the number of macroparticles in each slice and axis D) finally depicts the wake field kicks (marker value) that will be applied to each colour-corresponding slice. The arrows highlight which wake field kick will be applied to which slice. The wake field kick itself is evaluated for each slice according to the formula rendered below the plot.

From the Hamiltonian in Eq. (56), we can evaluate, apart from the betatron motion, which we already solved in Section 3, the corresponding additional kick from the wake fields as

$$\Delta x_i' = \underbrace{-\frac{e^2}{m\gamma\beta^2c^2}\overbrace{\sum_{j=0}^{n-1} N(z_j)\, W_{Cx}(z_i - z_j)}^{\text{Slice-dependent orbit shift (if line density does not change)}}\,.}_{\text{Dipolar term} \to \text{orbit kick}} \tag{57}$$

We can see from Eqs. (56) or (57) that a constant wake field generates a slice-dependent orbit shift. This results in a closed orbit distortion.

Equipped with the possibility of initializing and tracking macroparticle systems and having now also included wake field kicks, we can run a simulation that includes constant wake field effects. Figure 13 shows an example of such a simulation. A macroparticle bunch is initialized on the design orbit. The constant wake field shifts the orbit along the bunch. As a result, the bunch oscillates around the distorted orbit. Depending on the line density of the bunch and the shape of the constant wake field, this distortion may look different. In the example above, the distortion is pronounced towards the tail of the bunch. The bunch takes on a banana-like shape and the tail of the bunch flaps around the distorted orbit. For more information on the effects of constant wakes, the reader is referred to Ref. [16].

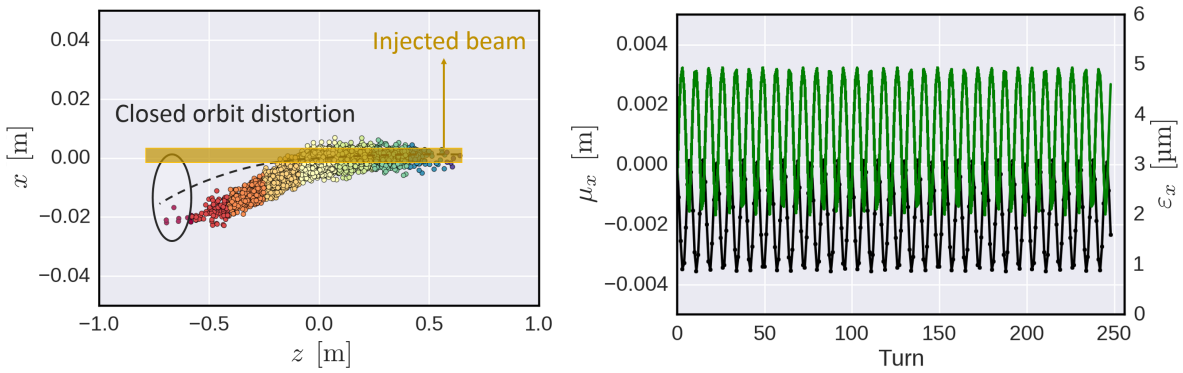


Fig. 13: Macroparticle simulation of the effects of a constant wake field. The initialization of the macroparticle bunch is outlined by the yellow rectangle. The distorted orbit is sketched by the dashed line. As a consequence of the constant wake field the tail of the macroparticle bunch will oscillate around the distorted orbit. The plot on the right shows the resulting oscillation in the mean bunch position and emittance. The motion remains bounded and the beam is stable.

3.2 Dipole wake fields

Dipole wake fields are the type of wake fields that generate instabilities. The kind of instability depends on the machine configuration. We will discuss three different forms of instabilities with simulations below. The Hamiltonian which includes dipole wake fields W_{Dx} is given as

$$H = \frac{1}{2}\,x_i'^2 + \frac{1}{2}K(s)\,x_i^2 - \frac{e^2}{m\gamma\beta^2c^2C}\sum_{j=0}^{n-1} N(z_j)\,\langle x_j\rangle\,W_{Dx}(z_i - z_j)\,x_i + H_{\mathrm{s}}(z,\delta). \tag{58}$$

Here, x_i is a macroparticle in slice i, $N(z_j)$ is the number of macroparticles in slice j, z_j is the centre longitudinal position of slice j and n is the number of slices. The wake field term in the Hamiltonian is linear in x_i. Thus, it generates an orbit shift (like dipoles would). However, in contrast to the constant wake field case we see that the Hamiltonian now contains an additional dependency on $\langle x_j\rangle$. Thus, different slices within the bunch get coupled. One can already imagine that this coupling of slices can lead to peculiar effects.

From the Hamiltonian in Eq. (58), we can evaluate, again apart from the betatron motion, the corresponding additional kick from the wake fields as

$$\Delta x_i' = \underbrace{-\frac{e^2}{m\gamma\beta^2c^2}\overbrace{\sum_{j=0}^{n-1} N(z_j)\,\langle x_j\rangle\,W_{Dx}(z_i - z_j)}^{\text{Slice-dependent orbit shift}}}_{\text{Dipolar term} \rightarrow \text{orbit kick}}. \tag{59}$$

We can see from Eqs. (58) or (59) that a dipole wake field generates a slice-dependent orbit shift. This shift in orbit is now not only dependent on the charge of a given slice but it is dependent on the weighted mean horizontal offset of the slice. Thus, the dynamics of the individual slices enters back into the equation. One can imagine that there are stationary solutions to the system described by the Hamiltonian in Eq. (58) and characterized by the kick in Eq. (59). Indeed, such solutions can be found analytically for simple cases by employing the Vlasov equation in a perturbation approach, as done in Ref. [12].

3.2.1 Beam break-up instability

Having learned how to implement dipolar wake fields, we can study the beam dynamics numerically. As the first case, we will investigate the behaviour of a bunch subject to a dipolar wake field in the absence of

synchrotron motion. In this configuration, the head of the bunch will generate dipolar wake fields, which will shift the orbit of the following slices in the bunch. This will trigger a mean dipole motion of these slices, which consequently will generate stronger dipolar wake fields for the slices in the bunch still to follow. This process builds up turn by turn causing the tail of the bunch to oscillate increasingly violently and rendering the bunch unstable. This type of instability is the beam break-up instability and can occur in linear accelerators or in circular accelerators at transition where the synchrotron motion is frozen. The beam is inherently unstable and will grow exponentially if no additional means of stabilization is put in place. Figure 14 shows a simulation of a macroparticle bunch subject to a dipolar wake field. The synchrotron motion was turned off. The build-up of the oscillation towards the tail of the bunch is clearly visible. This is accompanied by a fast exponential growth of the motion and the emittance of the bunch.

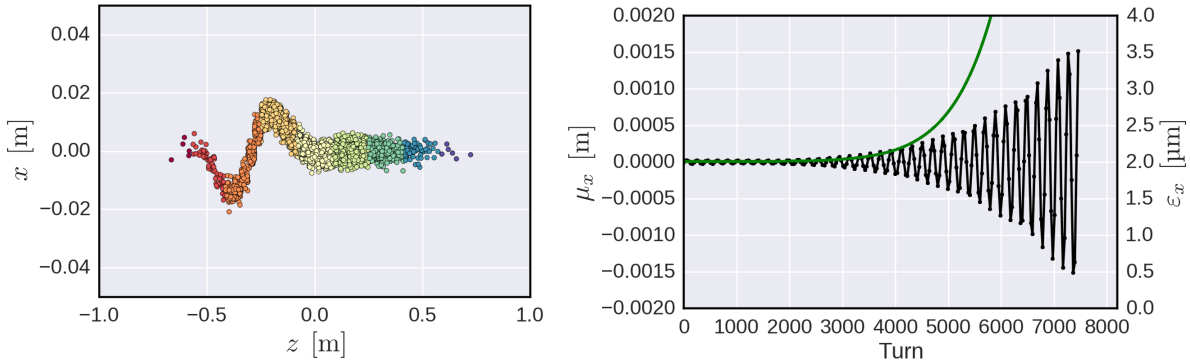


Fig. 14: Macroparticle simulation of the effects of a dipolar wake field. The synchrotron motion was turned off. The build-up of the oscillation towards the tail of the bunch is clearly visible in the left-hand plot. The exponential growth of the mean position and emittance of the bunch is shown in the right-hand plot.

3.2.2 Transverse mode coupling instability

We repeat the simulation from above investigating the behaviour of a bunch subject to a dipolar wake field; however, now we include synchrotron motion. We keep the chromaticity at zero. In this configuration, the head of the bunch will still generate dipolar wake fields, which will shift the orbit of the following slices in the bunch, triggering a mean dipole motion of these slices. Due to the synchrotron motion, eventually, the head and tail slices will exchange their positions. This will prevent the motion of the tail slices from building up coherently. In fact, in absence of chromaticity, gradually, this motion that builds up within a certain fraction of a synchrotron period gets washed out again at a later stage. Synchrotron motion acts as a stabilizing mechanism and the larger the synchrotron tune the more effective this washing-out of the coherent dipole motion is. This is true as long as the beam intensity remains below the TMCI threshold. At this point, the beam intensity is so high that the dipole motion builds up coherently before it can be cleared away via the synchrotron motion. A particularly violent instability emerges from this, which in many aspects is similar to the beam break-up instability discussed above. Figure 15 shows an example of a simulation of a TMCI. The macroparticle bunch is heavily distorted and the motion and the emittance of the bunch grow exponentially.

We can run several simulations at different intensities to gain more insight into this mechanism. For each simulation, at a given intensity, we do a spectral analysis of the coherent motion. We obtain spectra similar to the ones shown in Fig. 16. Below the TMCI threshold we can distinguish several spikes in the spectrum which correspond to different azimuthal and radial modes which arise from the interaction of the macroparticle bunch with the dipolar wake fields in combination with the synchrotron motion. These are in fact the stationary solutions, which were already mentioned earlier and discussed in more detail in Ref. [12]. What can be observed is that, as the intensity increases, certain modes tend to approach each other. This is indicated in the left-hand plot in Fig. 16. At a certain intensity these

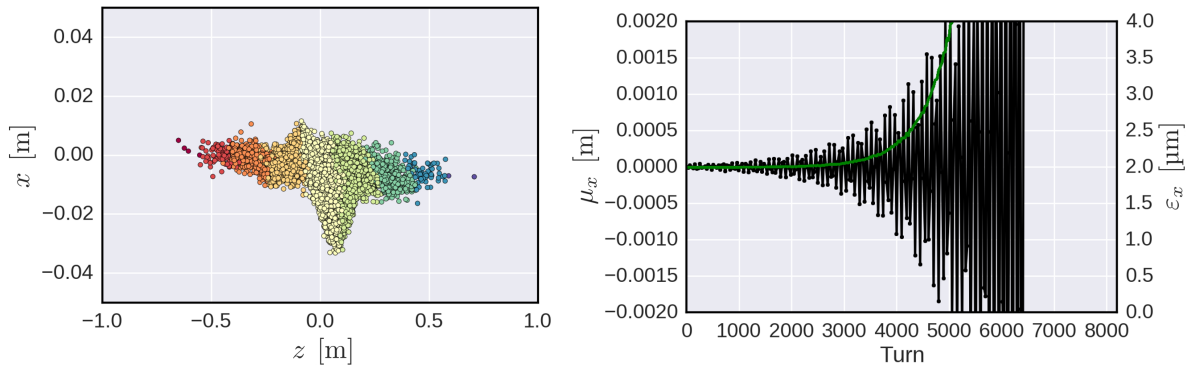


Fig. 15: Macroparticle simulation of the effects of a dipolar wake field. The synchrotron motion in turned on. The bunch intensity is above the TMCI threshold. The macroparticle bunch is heavily distorted. The exponential growth of the mean position and emittance of the bunch is shown in the right-hand plot.

modes merge or couple, as shown in the right-hand plot in Fig. 16. This is exactly the TMCI threshold at which the beam becomes violently unstable. The TMCI is a threshold effect. The beam is entirely stable below the intensity threshold and becomes inherently unstable above this threshold. Therefore, the TMCI threshold is often a hard limit for most machines.

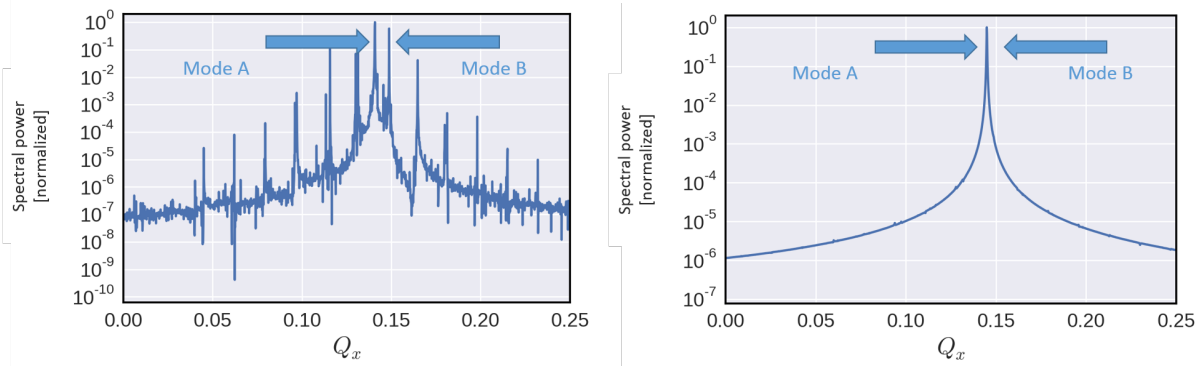


Fig. 16: Bunch spectra obtained from the mean position for a simulation below (left) and above (right) the intensity threshold for TMCI. Several spikes are distinguishable below the threshold. Two of the spikes (modes A and B) approach each other as the intensity increases until they finally merge as shown in the right-hand plot. At this point the bunch becomes violently unstable.

One of the strategies employed to raise the TMCI threshold is to increase the synchrotron tune. This was implemented recently in the CERN Super Proton Synchrotron (SPS) [17]. In changing the machine optics, the transition energy could be lowered such that the injection energy of the beam actually ended up farther away from the transition energy. This leads to an increase in the synchrotron tune. Figure 17 shows two plots in the way they are typically presented to visualize TMCI. The plots feature the bunch spectra (mode number on the y axis vs. marker colour and size) laid out against the bunch intensity. The mode number is given as the betatron tune shift in units of the synchrotron tune $m = (Q_x - Q_{x0})/Q_s$. The left-hand plot in Fig. 17 displays the coherent tune shifts with intensity and highlights a mode coupling of modes –2 and –3 taking place at an intensity of 1.5×10^{11} ppb. Increasing the synchrotron tune leads to a larger separation of the individual mode lines causing them to couple only at a later stage at 4×10^{11} ppb, as illustrated in the right-hand plot in Fig. 17.

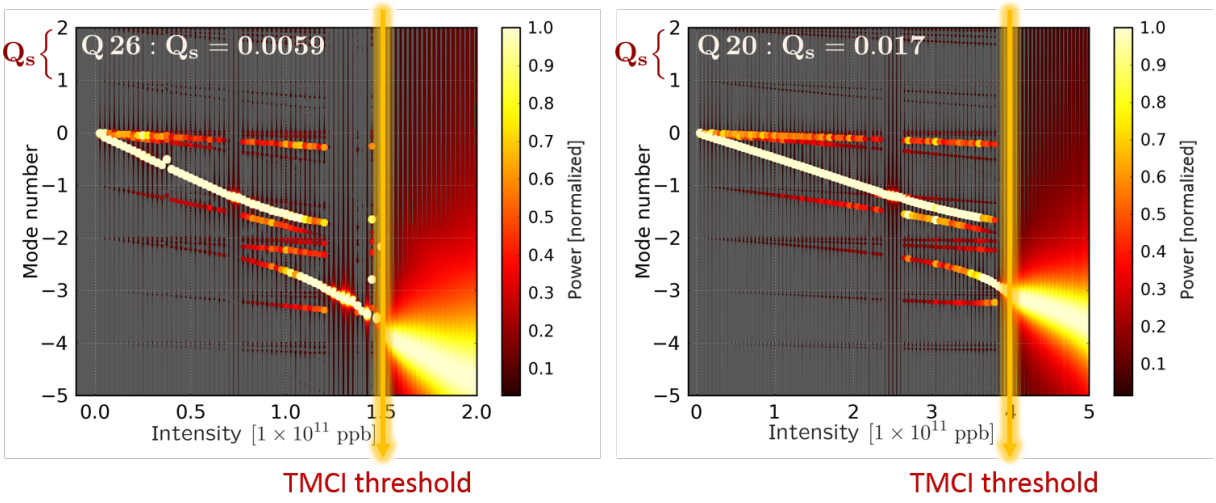


Fig. 17: TMCI spectra in the SPS for the two different optics (Q26 & Q20). The bunch spectra are plotted against the intensity. The spectral power is indicated by the markers and scales with the marker size and colour according to the legend. Mode coupling occurs between modes –2 and –3 at an intensity of $1.5 \times 1-^{11}$ ppb in the left-hand plot for which $Q_\mathrm{s} = 0.0059$ and at an intensity of 4×10^{11} ppb in the right-hand plot for which $Q_\mathrm{s} = 0.017$.

3.2.3 Head–tail instability

Finally, we study the case of a bunch subject to a dipolar wake field including synchrotron motion and now also including chromaticity. In this configuration, the head of the bunch will generate dipolar wake fields shifting the orbit of the following slices in the bunch while the synchrotron motion will periodically exchange head and tail slices within the bunch. As long as we are below the TMCI threshold, the bunch would be stable due to the synchrotron motion clearing away the dipole motion before it can build up coherently. This is true at zero chromaticity. Chromaticity, however, correlates the transverse motion of slices with their longitudinal position and thus incorporates a form of temporal interrelationship between the transverse and the longitudinal motions. This introduces a synchronicity mechanism which can lead to the synchrotron motion no longer clearing away the coherent dipole motion, but, instead, may cause it to actually seed a new form of coherent dipole motion. Finite chromaticity in combination with synchrotron motion (of course, otherwise chromaticity has no effect) and dipolar wake fields, in this case, generate head–tail instabilities. These are usually much less violent, i.e., they feature slower rise times, than the TMCI discussed in the previous section. On the other hand, head–tail instabilities are not a threshold effect. At finite chromaticity a bunch is inherently unstable, which would lead to an exponential growth of the head–tail mode if no additional means of stabilization is put in place. Head–tail modes are, again, stationary solutions of the accelerator–beam system. The imaginary part of their complex tune shift is given essentially by the overlap of the mode spectrum with the impedance. At zero chromaticity this overlap vanishes. At finite chromaticity certain modes are damped while others grow. Head–tail modes feature distinct patterns when observed in a pickup.

Figure 18 shows simulations of a head–tail instability for two sets of chromaticities compared to measurements in the LHC. The modes clearly exhibit their time-stationary nature. The different head–tail modes can be characterized by the number of nodes seen along the waveform. This number in fact corresponds to the radial mode number when performing the analytical analysis.

In summary, dipole wake fields generate instabilities that can be effectively investigated by numerical methods using macroparticle models. We looked at three different types of instabilities, which can be classified as shown in Table 2.

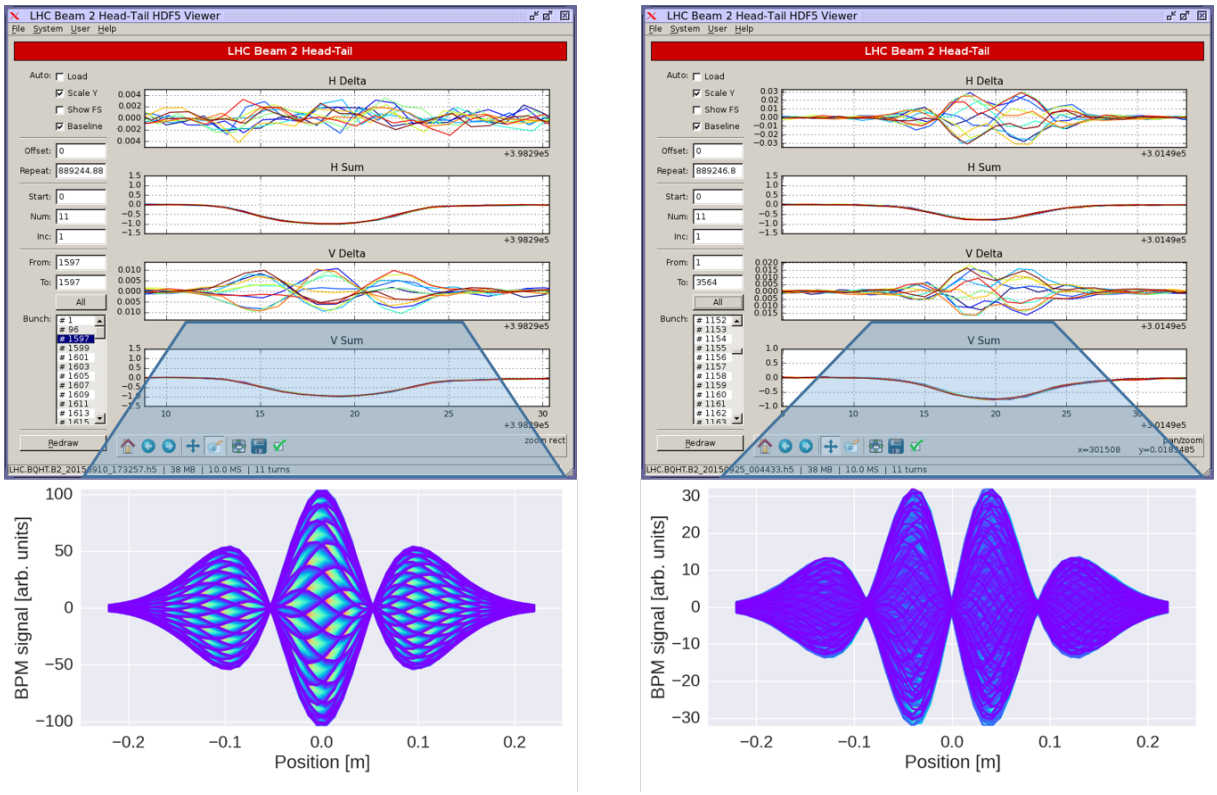


Fig. 18: Macroparticle simulation of a head–tail instability compared with measurements in the LHC. The left-hand plot shows a (radial) mode 2 instability occurring at a chromaticity of $Q' = 10$. The right-hand plot shows a (radial) mode 3 instability occurring at a chromaticity of $Q' = 15$.

Table 2: Classification of different types of instabilities generated by dipolar wake fields along with their conditions.

	Synchrotron motion	Chromaticity	Threshold effect	Rise time
Beam break-up	No	Irrelevant	No	Fast
TMCI	Yes	No	Yes	Fast
Head–tail instability	Yes	Yes	No	Slow

3.3 Quadrupole wake fields

Quadrupole wake fields are the highest order wake fields in the target coordinates that are commonly still considered. They do not usually generate instabilities. The Hamiltonian which includes quadrupole wake fields W_{Qx} is given as

$$H = \frac{1}{2}\,x_i'^2 + \frac{1}{2}K(s)\,x_i^2 - \frac{e^2}{m\gamma\beta^2c^2C}\sum_{j=0}^{n-1} N(z_j)\,W_{Qx}(z_i - z_j)\,x_i^2 + H_\mathrm{s}(z,\delta). \tag{60}$$

Here, x_i is a macroparticle in slice i, $N(z_j)$ is the number of macroparticles in slice j, z_j is the centre longitudinal position of slice j and n is the number of slices. The wake field term in the Hamiltonian is quadratic in x. Thus, it generates a tune shift (like quadrupoles would).

From the Hamiltonian in Eq. (60), we can evaluate, again apart from the betatron motion, the corresponding additional kick from the wake fields as

$$\Delta x_i' = \underbrace{-\frac{e^2}{m\gamma\beta^2c^2}\overbrace{\sum_{j=0}^{n-1} N(z_j)\, W_{Qx}(z_i - z_j)}^{\substack{\text{Slice-dependent tune shift} \\ \text{(if line density does not change)}}}\; x_i}_{\text{Quadrupolar term} \rightarrow \text{tune kick}}. \tag{61}$$

We can see from Eqs. (60) or (61) that a quadrupole wake field generates a slice-dependent tune shift. This results in an additional coherent tune shift, which in combination with dipole wake fields may expedite or hinder mode coupling. In addition it introduces a tune spread within the bunch combined with all the consequences such as potential resonance crossing or in principle even Landau damping.

Having learned how to implement quadrupolar wake fields, we can study the beam dynamics numerically. We find that a bunch subject to a quadrupolar wake field remains stable. We can perform a spectral analysis of each individual macroparticle to obtain the tune footprint of the macroparticle system. The resulting footprint is shown in Fig. 19. It is clearly visible how the footprint is grouped according to the slices of the bunch. Each coloured marker is actually a group of macroparticles within a given slice. The slice-dependent tune shift reflects the convolution with the quadrupolar wake field according to Eq. (61).

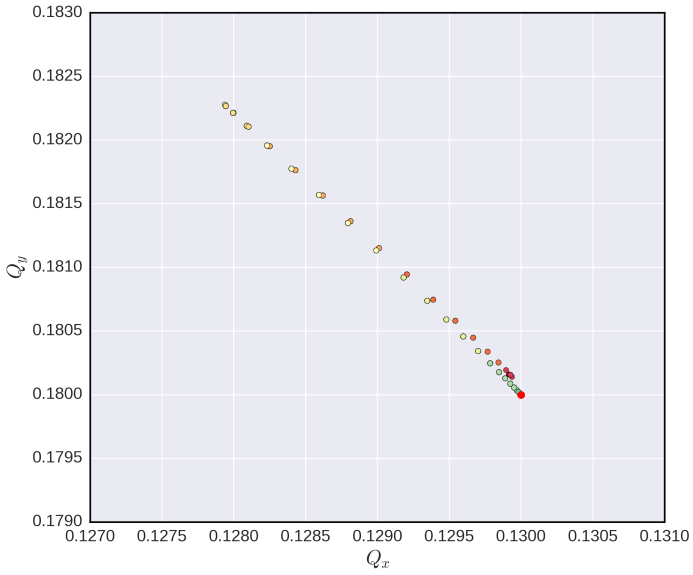
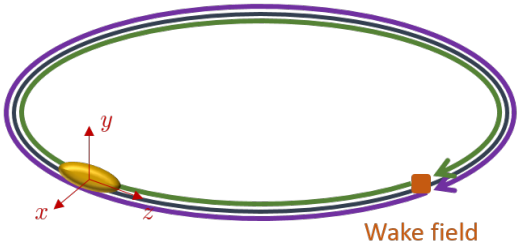


Fig. 19: Tune footprint of a macroparticle bunch subject to a quadrupolar wake field

This concludes our investigation of the numerical modelling of wake fields. This was the first truly collective effect we implemented which had to be treated taking the full macroparticle system self-consistently into account. We introduced wake fields as the electromagnetic response function of an object and learned that the net wake field kick is obtained via the convolution of the wake function with the macroparticle distribution. We studied constant, dipole and quadrupole wake fields with simulations and learned about the different types of instabilities. In the final part of the lectures we will examine another type of collective effect. This will consist of a second macroparticle system to interact with our macroparticle bunch. Being a dynamical system itself, this macroparticle system cannot be modelled via an electromagnetic response function as done for a rigid object via the wake fields. This will make the problem quite a bit more involved.

4 Modelling beam dynamics together with electron clouds

In the previous sections we learned how to initialize a macroparticle system representing a physical particle beam and how to track this macroparticle system through a circular machine transversely and longitudinally given the machine optics and RF configuration. Finally, we learned how to include and treat collective effects via wake fields to interact self-consistently with this macroparticle system. A feature of wake fields—being the electromagnetic response function of an object—is that they are a stationary property of the object that generates them. Unless the object changes its geometry or electromagnetic properties the wake function will not change. It can be computed once for an object and can then be used turn after turn. Figure 20 summarizes the numerical treatment of beam dynamics under the influence of collective effects via macroparticle models developed so far.



$$\left.\begin{pmatrix} x_i \\ x_i' \end{pmatrix}\right|_{k+1} = \mathcal{M}_i \left.\begin{pmatrix} x_i \\ x_i' \end{pmatrix}\right|_{k}$$

$$\left.\begin{pmatrix} z_i \\ \delta_i \end{pmatrix}\right|_{k+1} = \mathcal{I}\left[\left.\begin{pmatrix} z_i \\ \delta_i \end{pmatrix}\right|_{k}\right]$$

$$\left.(x_i')\right|_{k+1} = \left.(x_i')\right|_{k} + \mathcal{WK}$$

1. Initialize a macroparticle distribution with a given emittance
2. Update transverse coordinates and momenta according to the linear periodic transfer map – adjust the individual phase advance according to chromaticity and detuning with amplitude
3. Update the longitudinal coordinates and momenta according to the leap-frog integration scheme
4. Update momenta only (apply kicks) according to wake field generated kicks
5. Repeat turn-by-turn…

Fig. 20: Figurative summary of beam dynamics components treated so far. These include macroparticle system initialization, single-macroparticle tracking in the transverse and the longitudinal planes and collective interaction among macroparticles via wake fields.

In this section we will investigate another form of collective effects introduced not by a static object but by either the charged particle beam itself or also by another charged particle system. This can be a second beam which brings about the beam–beam interaction. This is a phenomenon that needs to be studied around the collision point of particle colliders. Instead of another beam it can also be a different form of charged particle distribution even with different particle species and energies such as for example electron clouds. We will use the example of the electron cloud throughout the rest of this section to illustrate the numerical concepts and methods used to treat these types of collective effects.

Electron clouds are a phenomenon that develops in accelerators of high-intensity positively charged beams. Primary electrons get released through residual gas ionization or synchrotron radiation. These electrons are accelerated in the electromagnetic fields generated by the positively charged circulating beam. At some point they will hit the vacuum chamber walls and depending on their energy and impact angle as well as on the surface properties—characterized by the secondary electron yield (SEY)—they will release a certain amount of secondary electrons. These will again be accelerated by the circulating beam and will also hit the vacuum chamber walls releasing yet more secondary electrons. Depending on the configuration of the circulating beam (intensity, bunch spacing etc) this can lead to an avalanche effect in generating an increasing amount of electrons, which we call multipacting. This generation of electrons will saturate when the electron density has reached a value such that the resulting space-charge

force repels secondary electrons sufficiently to suppress any further multipacting. At the end of this electron cloud build-up process one is left with stationary distributions of electrons at different locations around the ring. We call these stationary distributions electron clouds. Of course, one can imagine that these electron clouds are able to generate fields that can substantially perturb the motion of the circulating beam.

Electron clouds contain their own dynamics and macroscopic properties. In contrast to the static objects treated via wake fields in the previous section, electron clouds dynamically change their macroscopic properties. As such, if one were to calculate the electromagnetic response function of electron clouds, this would continuously change. In addition, the electromagnetic fields generated by electron clouds are highly non-linear, so the linear treatment using dipole and quadrupole wake fields would no longer be appropriate and higher order wake fields would have to be included. Most of the advantages gained from the wake field concept would no longer apply for electron clouds. Other strategies are required for an efficient numerical treatment of the interaction of charged particle beams with electron clouds.

4.1 Electron clouds as additional macroparticle systems

To numerically treat the electron cloud effects, the electron cloud has to be modelled as a macroparticle system in itself. This makes beam dynamics simulations with electron clouds much more involved, as one now needs to treat not only the macroparticle system resembling the physical particle beam but also one or several macroparticle systems resembling the electron clouds.

The same reasoning that was applied earlier for a physical particle beam also applies for the electron clouds. We can handle the electron clouds numerically by clustering the physical electrons into a discrete set of macroparticles while at the same time taking care about noise issues. We need to know the generalized coordinates and the canonically conjugate momenta of each individual macroparticle along with its charge and mass. For reasons that will be discussed further below, the electron cloud macroparticle system can be treated purely transversely. As such, it will be fully characterized by the generalized coordinates, given as the horizontal and vertical positions (x, y) in metres with respect to the centre of the vacuum chamber (usually) and the canonically conjugate momenta, given as the corresponding horizontal and vertical velocities $(\dot{x}, \dot{y})$. Initialization will usually take place as import of a distribution from an electron cloud build-up simulation. The electron cloud build-up simulations are treated in detail in Refs. [18, 19] and will not be further discussed here. If a distribution from an electron cloud build-up simulation is not available, one often chooses to start from a distribution that uniformly fills the vacuum chamber (or a part of it) with a given electron density.

As done earlier for the charged particle beam, the electron cloud macroparticle system can be represented by some allocated memory block in the main memory which is sufficiently large to accommodate all the relevant quantities of the macroparticle system, i.e., generalized coordinates and canonically conjugate momenta, charges and masses. Table 3 shows an example memory layout of an electron cloud macroparticle system in the main memory where each of the four phase-space variables is allocated as an array of length equal to the macroparticle number which contains the corresponding values for every macroparticle.

Table 3: Example memory layout of an electron cloud macroparticle system in the main memory. Each of the four phase-space variables is allocated as an array of length equal to the macroparticle number N.

Count	0	1	2	...	N
x	...	...	...	...	...
$\dot{x}$	...	...	...	...	...
y	...	...	...	...	...
$\dot{y}$	...	...	...	...	...

The action of the electron cloud on a charged particle beam will apply via the electromagnetic fields generated by the electron cloud and impacting the charged particle beam. On the other hand, the charged particle beam acts back onto the electron cloud as it also generates electromagnetic fields which in turn will cause electrons to be accelerated towards the positively charged particle beam. It follows that we will now have to handle the dynamics of the two macroparticle systems self-consistently.

During the beam passage, the electrons are accelerated, which causes the electron cloud to reform and, therefore, changes its macroscopic distribution. The electromagnetic fields generated by the electron cloud are thus strongly modified and consequently highly dynamic already during a single bunch passage. The resulting electromagnetic fields in the vicinity of the passing beam will perturb the beam motion and eventually lead to a distortion of the particle distribution within the beam. Then, again, the electromagnetic fields generated by the passing beam are changed, which in turn affects the electron dynamics and electron cloud reformation (and thus also the electromagnetic fields generated by the electron cloud) during the beam passage.

In frozen models, the distortion of the particle distribution within the beam is not taken into account and the electromagnetic fields are computed only for one beam passage. The electromagnetic fields generated by the electron cloud during this beam passage are stored as a field map and are then re-applied to the beam at every passage. Fully self-consistent models, on the other hand, do a full re-computation of the electromagnetic fields taking into account the current distributions of both the electron cloud and the charged particle beam at every iteration. It is clear that this becomes a strongly coupled system of many subsystems where each of the subsystems already contains highly complex dynamics of their own.

The motions of the electrons and the charged particles in the beam take place at very different time-scales due to their very different energy scales (the electrons are nearly at rest while the beam is moving at relativistic speeds). The electron motion takes place within fractions of the bunch length whereas the charged particle motion takes place within fractions of a revolution period. We will discuss the modelling and implementation of the electron cloud and charged particle beam interaction in the next sections.

4.2 Numerical treatment and computation of fields

The modelling of the accelerator–beam system now needs to include electron clouds as part of the accelerator, with these, at the same time, being independent macroparticle systems with their own dynamics. Figure 21 shows a sketch of a model machine layout. At each interaction the electromagnetic fields need to be computed for both the electron cloud and the charged particle beam. These fields then need to be applied to the two macroparticle systems and the macroparticles need to be propagated. The computation of the electromagnetic fields for the, in principle, arbitrary macroparticle distributions is, in general, a non-trivial task. Fortunately, we can employ a few basic assumptions which will simplify this exercise.

The electron cloud is assumed to be distributed over a considerable length L along the vacuum chamber much larger than its transverse dimensions b such that is has a low aspect ratio $A = b/L \ll 1$. In this case we can assume the electric fields of the electron cloud to be purely transverse. Since the electrons are also nearly at rest just before the beam passage we have for the electron velocity $v \ll c$ and the magnetic fields of the electron cloud can be practically neglected. The charged particle beam is assumed to be close to relativistic. For these types of beams we know that the electric field in the laboratory frame is strongly enhanced in the direction perpendicular to the direction of flight, i.e. it is also almost purely transverse. As mentioned already we can now see that the electromagnetic interaction between the electron cloud and the charged particle beam can be treated purely transversely and as an electrostatic problem.

Hence, what we need to do is to numerically solve the 2D Poisson equation at every encounter. Using the electric fields obtained from the solution, we then need to kick and propagate both the electrons and the charged particles in the beam. Since the electron motion happens very fast—on the time-scale of

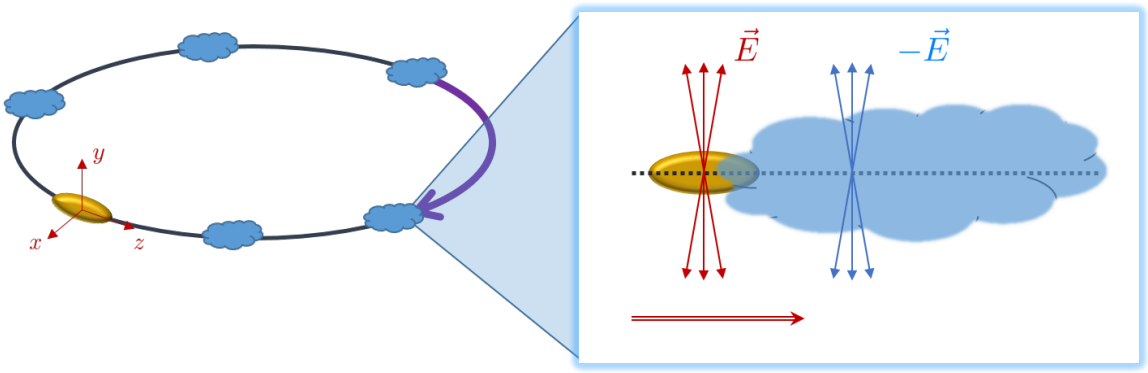


Fig. 21: Machine layout for a machine containing electron clouds. The electron clouds are placed at several locations along the ring. At each location, an electron cloud with charged particle beam interaction takes place.

fractions of a bunch length—the electrons receive a kick and then have to be propagated within this time frame, while the charged particles in the beam receive just the kick and are propagated through the ring via the betatron motion. For this reason, a macroparticle bunch is discretized into a set of longitudinal slices, as was done for the wake field interactions. The slice width will determine the maximum time frame in which the electrons will be linearly propagated. This situation is depicted in Fig. 22.

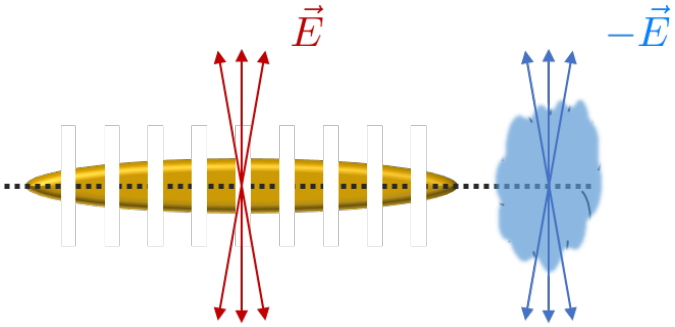


Fig. 22: Discretization of the incoming charged particle beam into a set of longitudinal slices. The interaction takes place slice-by-slice and the electrons are propagated on the time-scale of a single slice.

The motion of the charged particle beam on the other hand takes place on a much larger time-scale. Hence, the impact of the electron cloud is treated as an integrated effective force and the total kick is applied after the passage of a slice through the electron cloud. As such, since the kicks from each segment of the electron cloud will just linearly add up, the electron cloud can be longitudinally projected and treated for each slice in a 2D plane.

The overview of the numerical treatment of beam dynamics under the influence of electron clouds is now slightly adapted in comparison with Fig. 20 and is pictured in Fig. 23.

For each slice at each iteration the following steps are performed. First, one needs to compute the electric fields of the respective slice and of the projected electron cloud by numerically solving the Poisson equation

$$\begin{aligned} \Delta\phi_{\mathrm{e}^-}(x,y) &= -\frac{\rho_{\mathrm{e}^-}(x,y)}{\varepsilon_0}, \\ \Delta\phi_{\mathrm{p}^+}(x,y) &= -\frac{\rho_{\mathrm{p}^+}(x,y)}{\varepsilon_0}, \end{aligned} \tag{62}$$

where ϕ_{e^-} and ϕ_{p^+} and ρ_{e^-} and ρ_{p^+} are the potentials and densities of the electrons and the charged particles of the beam, respectively. There are several means of numerically solving the 2D Poisson

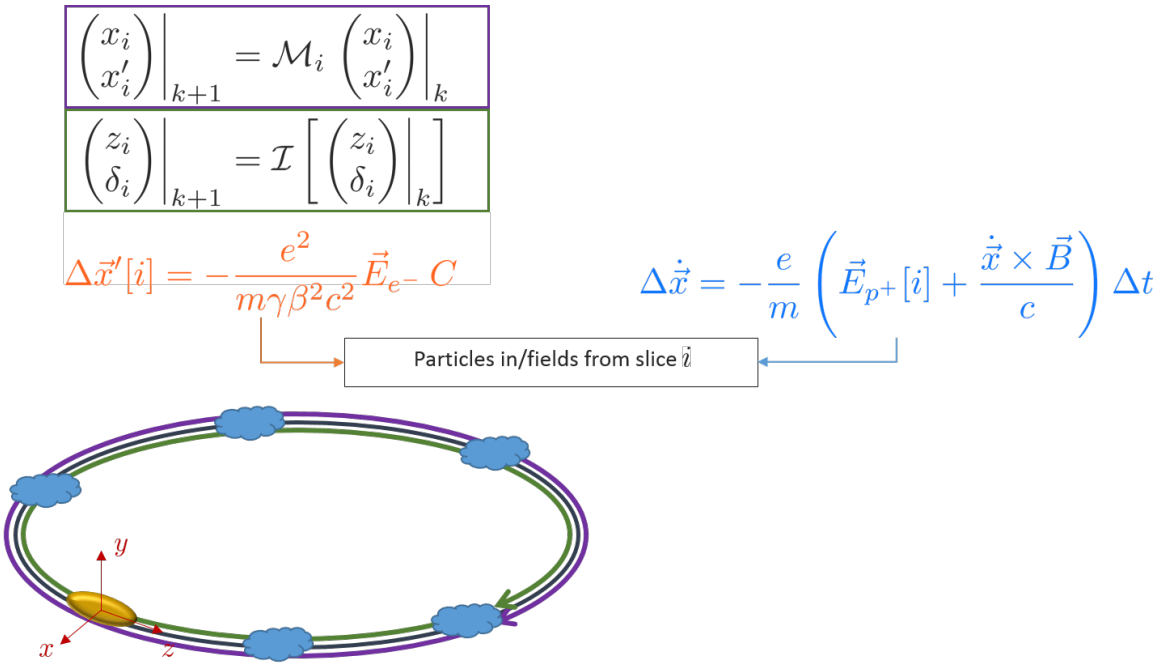


Fig. 23: Figurative summary of beam dynamics with electron clouds. The single- macroparticle tracking in the transverse and the longitudinal planes is included as earlier. In addition, several electron clouds are placed along the ring. The momenta of the electrons and the charged particles in the beam are updated according to the self-generated electric fields and any potential external magnetic field.

equation. One of the rather widely used methods for solving electron cloud problems is the particle-in-cell (PIC) algorithm—this will be discussed in more detail in the Appendices.

Having obtained the electric field of the electron cloud, we apply the corresponding kick to all macroparticles in the slices as already indicated in Fig. 23,

$$\begin{aligned} \Delta x_k'[i] &= \frac{e^2}{m\gamma\beta^2c^2}\,E_x(x_k,y_k)[i]\,L, \\ \Delta y_k'[i] &= \frac{e^2}{m\gamma\beta^2c^2}\,E_y(x_k,y_k)[i]\,L, \\ i &\in [0,\ldots,\text{slice number} - 1], \\ k &\in [0,\ldots,\text{macroparticle number} - 1]. \end{aligned} \tag{63}$$

Here, L is the length of the electron cloud and E_x and E_y the generated horizontal and vertical electric fields during the passage of slice i.

In the final step we need to update the velocities and positions of the electrons in the electron cloud. This is a multiscale problem and requires careful treatment. As already seen in Fig. 23, we will need to solve the system of equations

$$\begin{aligned} \dot{\vec{x}} &= \vec{v}, \\ \dot{\vec{v}} &= \frac{e}{m}\left(\vec{E}_{\mathrm{p}^+} + \frac{\vec{v}\times\vec{B}_{\text{ext.}}}{c}\right), \end{aligned} \tag{64}$$

where now, to be general, we set $\vec{x}$, $\vec{v}$, $\vec{E}$, $\vec{B}$ to be vectors. Looking at Eq. (64), we realize that the equations of motion contain a magnetic field term.

The majority of segments of a circular accelerator actually consist of regions encompassed by magnetic fields. If these regions are occupied by electron clouds, which is usually the case, then there will be a significant impact of these external magnetic fields on the electron dynamics, which needs to be taken into account. To numerically solve the set of equations in (64), these are discretized in time to

$$\begin{aligned} \frac{\mathbf{x}_{k+1}-\mathbf{x}_k}{\Delta t} &= \mathbf{v}_{k+1}, \\ \frac{\mathbf{v}_{k+1}-\mathbf{v}_k}{\Delta t} &= \frac{e}{m}\left(\mathbf{E}_k + \frac{(\mathbf{v}_{k+1}+\mathbf{v}_k)\times\mathbf{B}_k}{2c}\right). \end{aligned} \tag{65}$$

From Eq. (65), the multiscale nature of the problem now becomes evident (slow guiding drift from electric fields with fast cyclotron motion from magnetic fields). For a correct treatment of the electron motion, the cyclotron motion needs to be well resolved. The cyclotron period can be as short as fractions of a slice length such that a single time step has to be split into even smaller subintervals. The special, apparently implicit form of the velocity equation in Eq. (65), using the average velocity at a time $k+1/2$, ensures that the algorithm conserves the phase-space volume, i.e. it is volume-preserving [3]. This is important as electrons can perform several cyclotron oscillations during one slice passage and the motion is required to remain stable. Because it is implicit, the numerical solution of Eq. (65) is rather expensive. Fortunately, there is another slightly adapted algorithm that is very well suited to treat these kinds of problems. This is the Boris algorithm which is borrowed from plasma physics where it is the de facto standard for particle pushing (see Refs. [20, 21]). While Eq. (65) appears to be implicit, it can actually be made manifest explicit by writing

$$\mathbf{v}^- = \mathbf{v}_k + \frac{e}{m}\,\mathbf{E}_k\,\frac{\Delta t}{2}, \tag{66a}$$

$$\frac{\mathbf{v}^+ - \mathbf{v}^-}{\Delta t} = \frac{e}{2mc}\left(\left(\mathbf{v}^+ + \mathbf{v}^-\right)\times\mathbf{B}_k\right), \tag{66b}$$

$$\mathbf{v}_{k+1} = \mathbf{v}^+ + \frac{e}{m}\,\mathbf{E}_k\,\frac{\Delta t}{2}. \tag{66c}$$

Equations (66a) and (66c) eliminate the electric field. Then Eq. (66b) can in fact be solved by basic geometric means. In assuming this form, the algorithm—although it is not strictly symplectic—becomes explicit (efficiency), time centred (local accuracy) and the energy error is globally bound (preservation of phase-space volume).

With this final step in updating the velocities and positions of the electrons, we have completed the cycle for one slice and can now take on the following slice. By propagating the electrons in this way, in a slice-by-slice manner, during the beam passage, electrons will be attracted towards the passing beam. We call this process the pinch of the electron cloud, as electrons are pinched towards the beam. A schematic of this slice-by-slice interaction during a beam passage is shown in Fig. 24.

Depending on the configuration of the external magnetic fields the electrons will form different structures in electron density and exhibit different locations where the electrons will be concentrated into hot spots. Fig. 25 shows electron cloud pinches for a field-free region compared to a dipole and a quadrupole magnetic field region. The simulation uses an LHC-type vacuum chamber. Regions of high electron density are light coloured. The different patterns in electron cloud density are clearly visible. In the field-free region the electrons are free to propagate towards the passing beam. In regions with magnetic fields the electrons are trapped to circulate around and follow the magnetic field lines. In dipoles, this leads to the formation of the characteristic stripes of high electron density. In quadrupoles, the magnetic field can even form a magnetic bottle causing electrons to be reflected and preventing further impact into the vacuum chamber but forcing them around the beam. Figure 25 also highlights the multipacting that takes place on the inner surface of the vacuum chamber.

[3] In fact, Eq. (65) resembles a leap-frog scheme, which is one of the most simple symplectic integrators. It is not strictly a leap-frog scheme, though, because in a leap-frog scheme the accelerating force cannot depend on velocity.

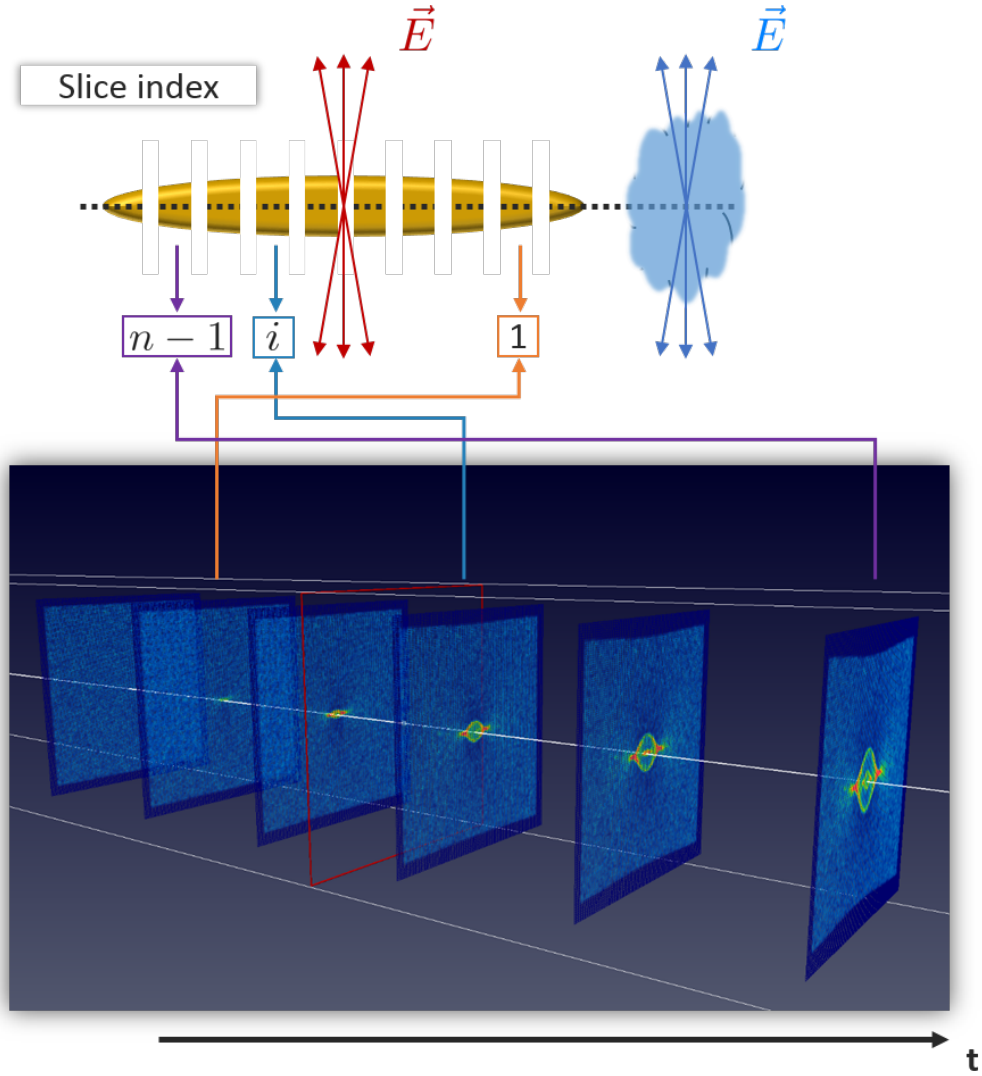
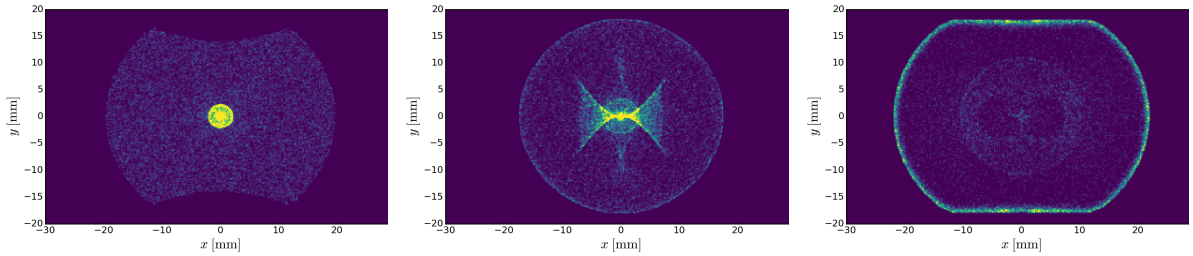


Fig. 24: The slice-by-slice interaction of an electron cloud with a passing bunch. Each plane pictures a snapshot of the current projected electron density distribution on the grid which is used for the field computation.

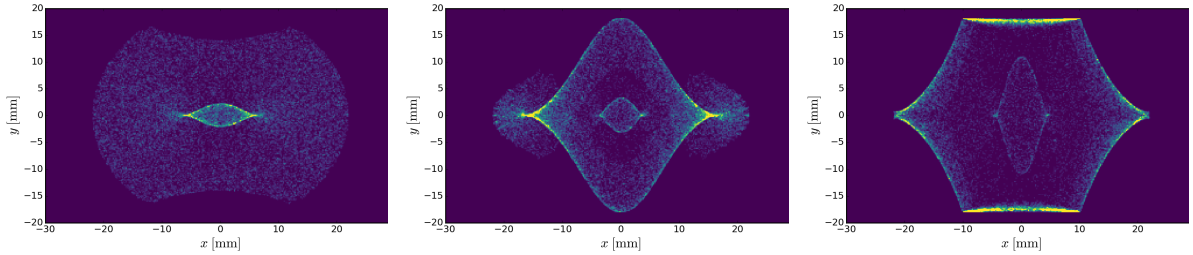
4.3 Combining build-up and instability simulations

Fully self-consistent electron cloud simulations require both build-up and instability simulation. Build-up simulations accurately model the electron dynamics and usually use a coarser model for the passing beam, making a weak–strong approximation, i.e., the beam is rigid and remains unaffected by the electron clouds. As we have just learned, in reality, this is not true and after several revolutions the beam may have significantly changed its original profile, which in turn changes the originally simulated build-up process and hence the stationary electron cloud distributions that evolve from this. The instability simulation, on the other hand, in addition to the electron dynamics also takes the detailed beam dynamics into account. Therefore, the correct approach would be to combine the two simulations in a way as outlined in Fig. 26.

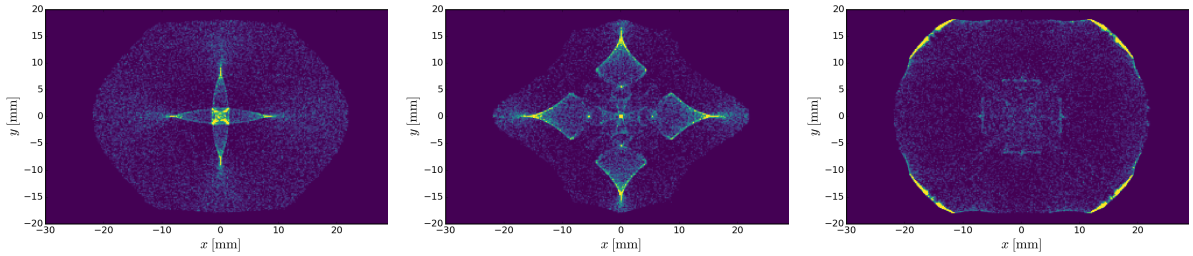
The problem is that the two simulations cover entirely different time-scales. The electron cloud build-up is usually completed within a single turn but requires multiple bunches for the multipacting to take place. The electron cloud instability, although it is relatively fast, can require several thousands of turns to grow. During this time, for a self-consistent description, multiple electron clouds together with multiple bunches would need to be stored and propagated at every time step. This cannot be easily



(a) Field-free region



(b) Dipole magnet



(c) Quadrupole magnet

Fig. 25: Electron cloud pinch during the passage of a charged particle beam within different magnetic field configurations. Sketched is the electron density evolution within an LHC-type beam screen. High-density regions are coloured light. The multipacting which takes place upon impact on the boundaries is well recognizable.

done in an efficient way and would require highly optimized code with parallel capabilities. Efforts are ongoing to overcome this limitation. To date, most of the instability simulations are carried out by assuming a given stationary electron cloud distribution at every bunch passage.

4.4 Application of electron cloud instability simulations

Electron clouds pose serious intensity limitations especially for circular machines with high brightness and closely spaced bunches. They deposit a considerable amount of heat onto the beam screens, which becomes a big issue especially for superconducting machines like the LHC. They reduce the dynamic aperture, which leads to beam quality degradation and poor lifetime. Finally, electron clouds also generate coherent instabilities. Whereas the heat loads can be evaluated by means of build-up simulations, the incoherent losses and coherent instabilities require instability simulations. These are mostly used to determine threshold values on the electron cloud density before beams go unstable due to electron clouds. Figure 27 shows an example of a typical electron cloud instability simulation. Electron clouds are initialized at different densities. For each density a separate simulation is run to evaluate the interaction of the electron cloud with the circulating beam. The emittance evolution of the circulating beam is observed for each case and the density threshold is set to the lowest value for which an exponential

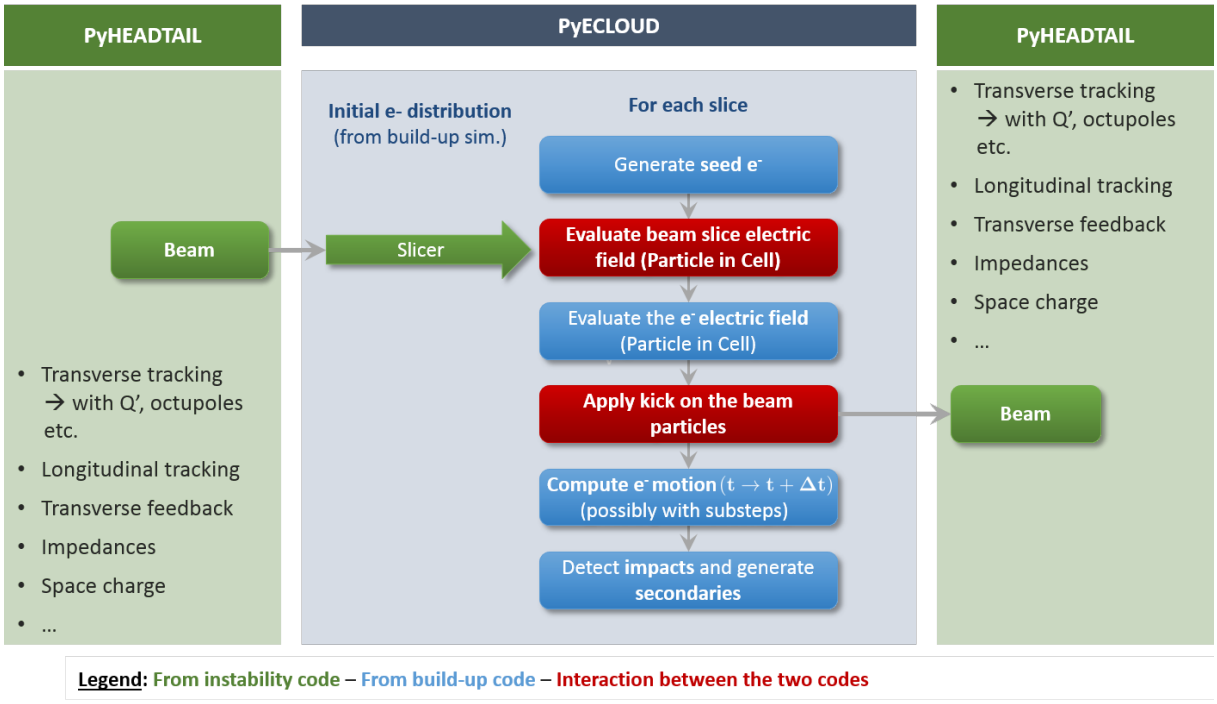


Fig. 26: Scheme for a simulation strategy combining build-up (e.g. PyECLOUD) and instability (e.g. PyHEADTAIL) simulations. Of course this requires a good and modular code design.

growth of the emittance is observed for the first time. As the density is increased further the emittance grows more rapidly. For high densities the emittance tends to saturate. This is understood to be a combination of several effects linked to the strong growth in the beam size. In combination with the strong non-linearities of the electron cloud fields, the large beam accumulates a large tune spread causing it to strongly decohere and also at some point to be Landau damped, which is understood to be one of the reasons why the beam stabilizes. On the other hand, the large beam reduces the strength of the electron cloud pinch, which also renders the beam more stable. This can be seen for the orange curve in Fig. 27. Figure 28 shows the corresponding BPM signal that would be observed. There is a strong intrabunch motion that looks similar to what would be expected from a TMCI. After more turns this initially coherent signal decoheres, as can be seen more clearly in the 3D representation of the BPM signal in Fig. 28.

The instability threshold density can be compared to the multipacting threshold density expected for a given SEY. From this the beam stability can be inferred as well as its dependence on different SEYs. This technique was used for example to identify the most critical vacuum chambers in the SPS and prioritize them accordingly for potential amorphous carbon coating which can significantly lower the SEY of the inner surface of the vacuum chamber.

Another application which combines the build-up with the instability simulations is the investigation of single-bunch instabilities for multibunch batches in the LHC. This was studied in detail in Ref. [22]. From build-up simulations strong electron clouds are expected for the injection of batches with multiple bunches at a bunch spacing of 25 ns. Observations with the pickups from the transverse damper feature strong transverse instabilities towards the end of the batches. This is usually a clear indication for electron cloud activity, which is increased towards the end of a batch. To reproduce and confirm this effect with simulations, a build-up simulation is run for the injection of a single batch. The electron distribution is recorded just before each bunch's passage. This distribution is then imported into an instability simulation which is then performed for each of the corresponding bunches, respectively. Figure 29 shows the computed electron cloud line density as obtained by the build-up simulation. The red dots indicate the line density of the snapshots of the distribution loaded into the instability simulation.

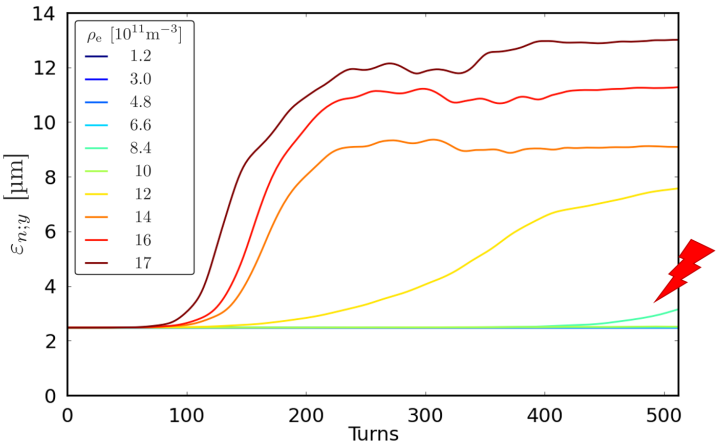


Fig. 27: Electron cloud instability simulation to determine the density threshold. The curves show the transverse emittance evolution for different electron cloud densities. The threshold is the lowest density for which an exponential growth of the emittance is observed over the given number of turns (here at $\rho_e \approx 8.5 \times 10^{11}$ m^{-3}).

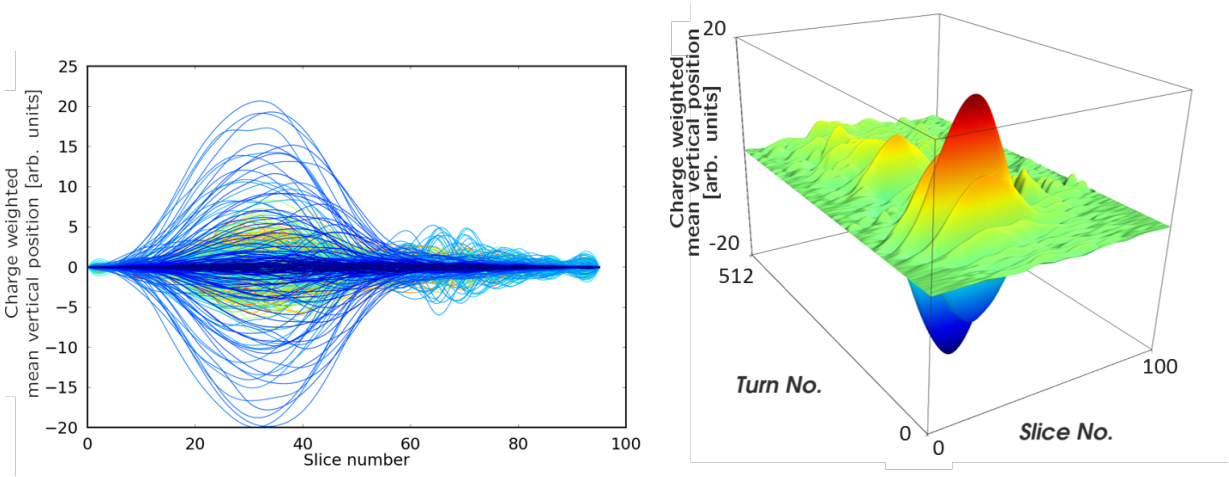


Fig. 28: The charge-weighted transverse offset as would be measured in a BPM pickup is shown for the case of $\rho_e \approx 14 \times 10^{11}$m^{-3}. A strong coherent motion can be seen towards the tail of the bunch similar to what is observed for a TMCI. The motion decoheres, combined with an emittance blow-up after approximately 128 turns.

Figure 30 shows the comparison of the measurements and the simulations of the observed electron cloud instability. The instability occurs for exactly the same bunches in the measurement as in the simulations and the overall behaviour of the beam is well captured. This helped a lot in the understanding of both the the electron cloud build-up and instabilities in the LHC.

Appendix A: Brief introduction to the particle-in-cell algorithm

This section introduces briefly some of the basic concepts of the PIC algorithm. Essentially, it follows the same philosophy that we already encountered when introducing macroparticle models themselves. The physical particle systems were significantly reduced in the number of degrees of freedom by moving to macroparticle systems. This was done by clustering physical particles within a representative region in space into single macroparticles to obtain a numerical representation of the original physical particle system as a macroparticle system. The PIC algorithm now, instead of physical systems of discrete particles, targets physical systems of continuous fields. In the framework of the algorithm, the physical space

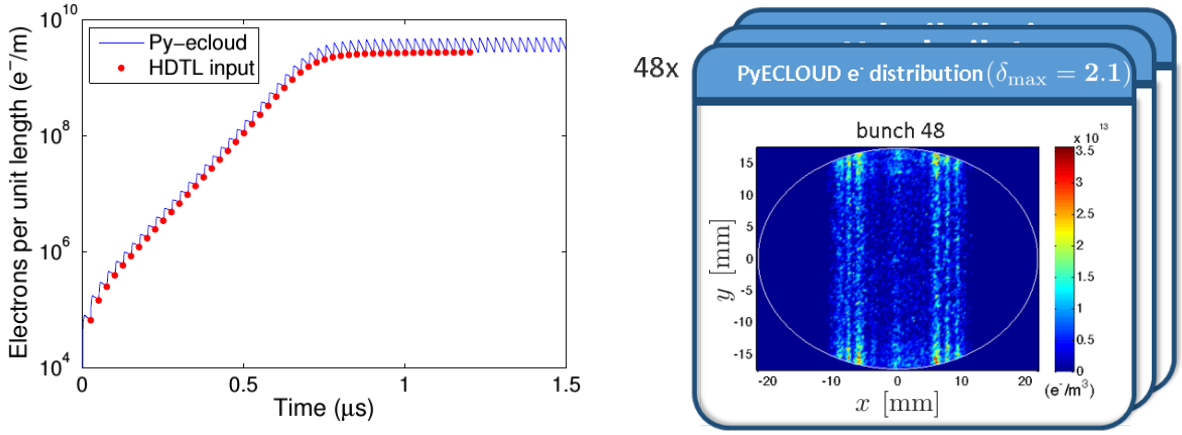


Fig. 29: Snapshots of the electron distributions are taken from a build-up simulation (blue) just before each bunch passage. The snapshots are then imported into an instability simulation (red) for each of the 48 corresponding bunches, respectively.

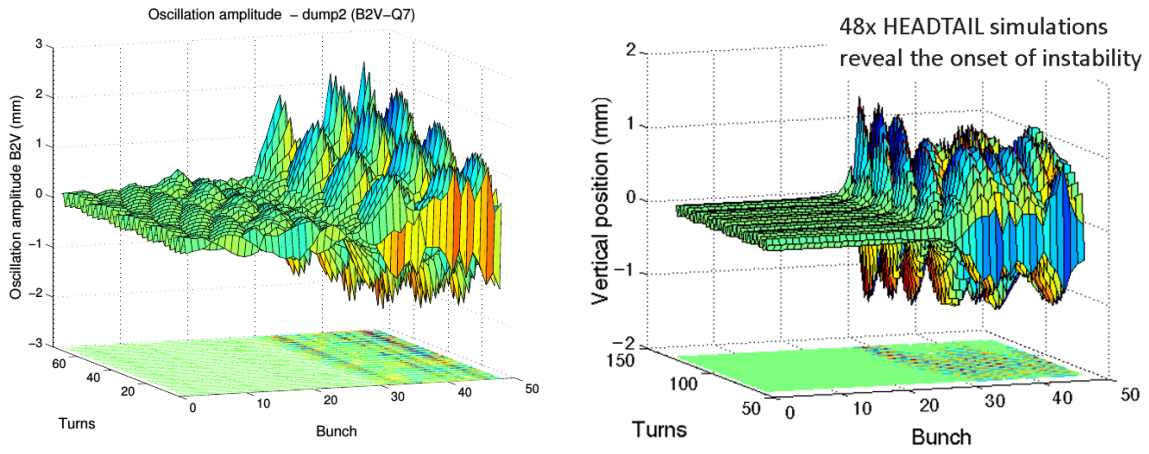
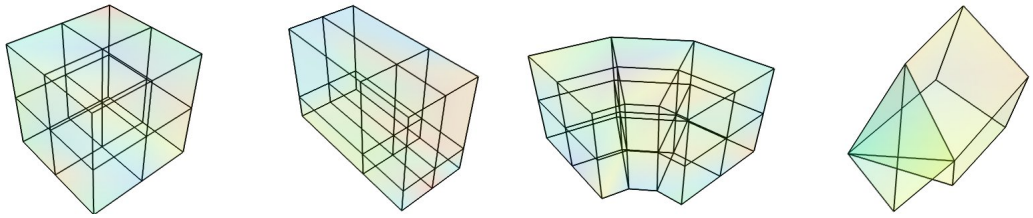


Fig. 30: Transverse oscillations measured with the pickups of the transverse feedback for 73 turns just before the beam dump (left). Bunch-by-bunch oscillations in the vertical plane as obtained from an instability simulation (right). In both cases the second half of the batch is rendered unstable by the electron cloud.

on which the fields are defined is now discretized into a set of regularly or irregularly distributed space points which is imbued with a topology by assigning to each space point the connectivity information to other space points. The set of space points together with the connectivity information forms a mesh. The space points are then called mesh nodes and the areas surrounded by connected sets of mesh nodes form mesh cells. Figure A.1 shows some examples of different discretizations of space with different types of meshes. The choice of the mesh type depends strongly on the problem to be analysed.

Computationally often even macroparticle systems are too big to efficiently handle the particle-to-particle interactions induced, for example, by electrostatic fields. One way to deal with this, then, is to perform a further step of discretization according to the PIC algorithm. In this particular case, of course, a computational gain is only made if the resulting number of space points is less than the number of macroparticles. It is then interesting to think about the gain that one can expect from computing the electrostatic forces particle-to-particle in comparison with computing them via a mesh. The electrostatic fields at the macroparticle positions $\mathbf{x}_k$ can be computed in a rather straightforward manner by summing



(a) Uniform (b) Rectangular (c) Structured (d) Unstructured

Fig. A.1: Different basic types of mesh commonly used to computationally represent domains in space (taken from Ref. [24]).

all forces from all other macroparticle positions $\mathbf{x}_l$,

$$\Delta\mathbf{x}_k = \sum_{l\neq k} \frac{e^2}{2\pi\varepsilon_0}\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}, \qquad k \in [0, \ldots, \text{macroparticle number}]. \tag{A1}$$

This sum needs to be evaluated for every macroparticle, which makes the algorithm scale as

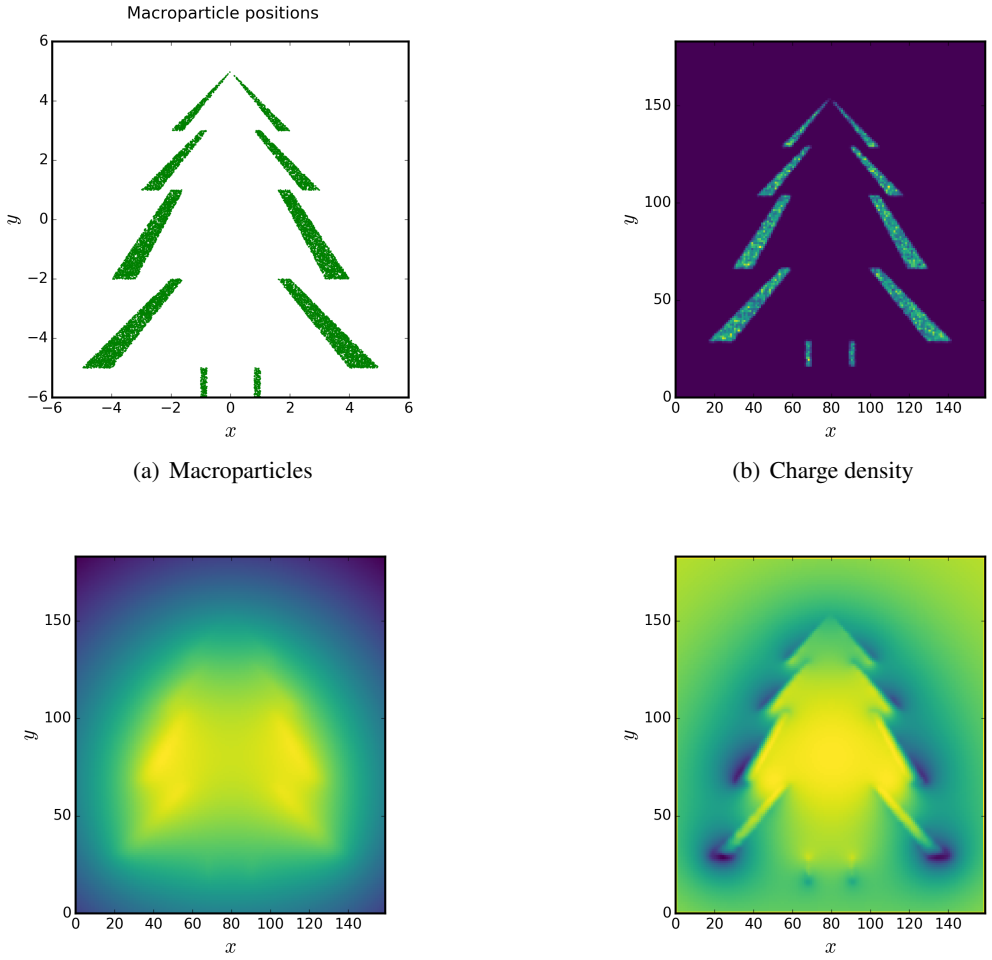
$$\frac{1}{2}\left(N^2_{\text{macroparticle}} - N_{\text{macroparticle}}\right).$$

Here, $N_{\text{macroparticle}}$ is the macroparticle number, which can be of the order of a few millions. Introducing mesh nodes and computing the forces on the mesh nodes instead reduces the number of operations according to the number of mesh nodes $N_{\text{mesh nodes}}$ proportionally. If we assume a macroparticle system of 1×10^6 macroparticles and compare this to a mesh of dimensions 128×128 then the number of operations reduces from $\sim 500 \times 10^9$ to $\sim 0.1 \times 10^9$ when moving from a particle-to-particle (p2p)- to a mesh node-to-mesh node (m2m)-based computation. Of course, for the latter there will be some overhead from the scattering and gathering of the macroparticles to the mesh nodes and vice versa. However, even this large reduction factor is often not enough to achieve satisfactory results, especially as the dimensions of the mesh increase. In addition, sometimes regions in space need to be computed that do not contain any macroparticles but still provide an important contribution to the electromagnetic fields. Here, the goal is not to reduce the number of mesh nodes compared to the number of macroparticles but rather to obtain a well-resolved discrete sampling of the regions of interest. And, finally, introducing boundary conditions makes the situation a lot more complicated. To satisfy the boundary conditions for arbitrarily shaped boundaries, in principle, an infinite number of image charges would need to be placed iteratively. For these reasons, other more efficient methods are used instead for solving the field equations.

The PIC algorithm basically consists of three main steps.

- Macroparticles are scattered to the grid. This is done by depositing charges via an interpolation algorithm from the macroparticle positions to the locations of the mesh nodes.
- A potential is computed by solving the field equations on the mesh. There are many different ways for solving partial differential equations (PDEs) on a mesh and we will mention just a few below. Once the potential is known, the differential on the mesh yields the force fields.
- The force fields are then gathered from the mesh node locations back to the macroparticle positions. This must be done via the same interpolation that was used to scatter macroparticles to the mesh, otherwise the conservation of momentum will be violated.

We will discuss these steps briefly below using the example of solving the 2D Poisson equation on a rectangular mesh with dimensions $N_x \times N_y$ within a bounded domain Ω and some given (Dirichlet)



(a) Macroparticles

(b) Charge density

(c) Electrostatic potential

(d) Magnitude of the electric field

Fig. A.2: The different stages of the PIC algorithm, here for computation of the magnitude of the electrostatic fields from a given charge distribution in open space.

boundary conditions on $\partial\Omega$

$$\begin{aligned} \Delta\varphi &= -\frac{\rho}{\varepsilon_0} \qquad \text{in } \Omega, \\ \varphi &= b \qquad\quad\ \ \text{on } \partial\Omega. \end{aligned} \tag{A2}$$

An excellent reference that explains all parts of the algorithm in detail is Ref. [23]. As an example, we may consider Fig. A.2, which shows the results of the different stages of the space-charge calculation for a rather arbitrary given charge distribution.

A.5 Macroparticle and field interpolation

To solve the Poisson equation on the mesh one first needs to know the charge density on the mesh. For this, the macroparticle charges are distributed (scattered) to the mesh nodes. There are several ways of doing this. One way is to assign the full macroparticle charge to the nearest mesh node. This is the nearest grid point (NGP) scheme, which is computationally very efficient but lacks resolution and suffers from high noise. A more refined variant is to assign weighted fractions of the macroparticle charge to all the surrounding mesh nodes within a given cell. This results in a linear interpolation of the macroparticle charge density to the mesh and is the cloud-in-cell (CIC) scheme. It offers a good compromise between computational effort and resolution and is often employed in many standard PIC codes. Higher order

interpolation schemes can be systematically constructed depending on the constraints (i.e. long-range constraint, smoothness constraint or momentum conservation constraint) imposed on them. Generally, these interpolation schemes are characterized by a shape function S and the charge weighting is then generally given as (see Ref. [23], Section 5-3-4)

$$W_i(x) = \int_{x_i-h/2}^{x_i+h/2} S(x-x')\,\mathrm{d}x' = \int \Pi\left(\frac{x'}{h}\right) S(x-x')\,\mathrm{d}x', \tag{A3}$$

with h the mesh spacing and Π the 'top-hat' function

$$\Pi(x) = \begin{cases} 0, & |x| > \frac{1}{2}, \\ \frac{1}{2}, & |x| = \frac{1}{2}, \\ 1, & |x| < \frac{1}{2}. \end{cases}$$

We will look at an example for a 2D CIC scheme charge assignment on a regular rectangular mesh with constant mesh spacing h_x and h_y. The shape function for this scheme is given as

$$S(x) = \frac{1}{H}\,\Pi\left(\frac{x}{H}\right).$$

The charge of each macroparticle is distributed to the four neighbouring nodes of the mesh according to the CIC scheme to obtain a discrete representation of the charge distribution. In particular, for a macroparticle at the location (x, y) carrying charge q, the indices of the neighbouring grid nodes can be easily computed by means of a floor division (see Fig. A.3)

$$i = \text{floor}\left[\frac{x-x_0}{h_x}\right], \quad j = \text{floor}\left[\frac{y-y_0}{h_y}\right], \tag{A4}$$

where x_0, y_0 is the bottom left corner of the grid. Then the fractional normalized distances are given as (see Fig. A.3)

$$f_x = \frac{x-x_0}{h_x} - i, \quad f_y = \frac{y-y_0}{h_y} - j, \tag{A5}$$

and the charge density matrix is updated as

$$\begin{aligned} \tilde{\rho}_{i,j} &= \rho_{i,j} + \frac{q}{h_x\,h_y}\,(1-f_x)(1-f_y), \\ \tilde{\rho}_{i+1,j} &= \rho_{i+1,j} + \frac{q}{h_x\,h_y}\,f_x(1-f_y), \\ \tilde{\rho}_{i,j+1} &= \rho_{i,j+1} + \frac{q}{h_x\,h_y}\,(1-f_x)f_y, \\ \tilde{\rho}_{i+1,j+1} &= \rho_{i+1,j+1} + \frac{q}{h_x\,h_y}\,f_x\,f_y. \end{aligned} \tag{A6}$$

Once the fields have been obtained on the mesh, the inverse operation is performed in the same manner as the fields are gathered back to the macroparticle positions:

$$\mathbf{E}(x,y) = \mathbf{E}_{i,j}\,(1-f_x)(1-f_y) + \mathbf{E}_{i+1,j}\,f_x(1-f_y) + \mathbf{E}_{i,j+1}\,(1-f_x)f_y + \mathbf{E}_{i+1,j+1}\,f_x f_y. \tag{A7}$$

A.6 Field computation

Poisson solvers on discretized grids exist in many variants.

Finite-difference methods are used on structured grids where the continuous domain in space is replaced by a discrete set of points on a grid on which the electric and magnetic fields are computed.

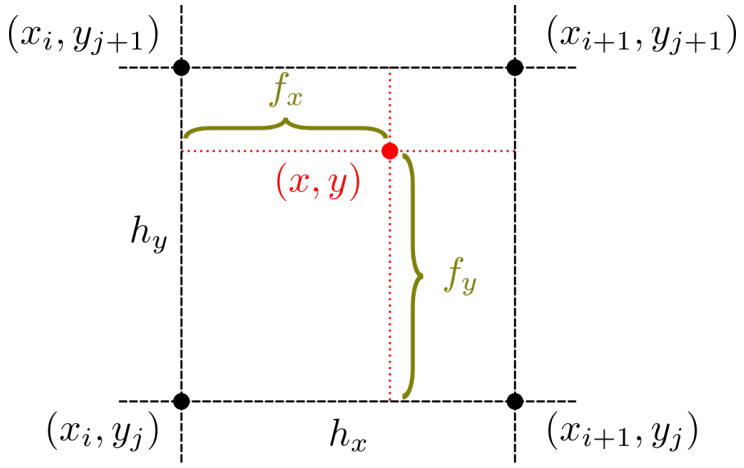


Fig. A.3: Nodes and distances involved in the macroparticle scatter and field gather stages of the space charge field calculation.

Derivatives are then approximated with differences between neighbouring grid point values. Thus, the differential Poisson equation is turned into an algebraic equation. We assume again a 2D rectangular grid, not necessarily uniform, that discretizes a domain Ω in the x and in the y directions into N_x and N_y subintervals, respectively, as

$$\left[h_{x,0},\, h_{x,1},\, \ldots,\, h_{x,N_x}\right], \quad \left[h_{y,0},\, h_{y,1},\, \ldots,\, h_{y,N_y}\right]. \tag{A8}$$

Furthermore, we introduce what is known in the finite integration technique (FIT), which has been introduced by Weiland (see Ref. [25]), as the mesh spacing (edges) on the dual grid

$$\hat{h}_{x,i} = \begin{cases} \dfrac{h_{x,i-1} + h_{x,i}}{2}, & 0 < i < N_x, \\ \dfrac{h_{x,i}}{2}, & i = 0, N_x \end{cases} \tag{A9}$$

and equally for the y direction. The discretization of the second-order derivative with second-order finite differences in the general case gives

$$\frac{\partial^2 \varphi(x_i, y_j)}{\partial x^2} \approx \frac{\varphi(x_{i-1}, y_i)}{\hat{h}_{x,i} h_{x,i-1}} + \frac{2\varphi(x_i, y_i)}{h_{x,i-1} h_{x,i}} + \frac{\varphi(x_{i+1}, y_i)}{\hat{h}_{x,i} h_{x,i}}. \tag{A10}$$

Consequently, it follows that the discretization of the Poisson equation with second-order finite differences on the above-described non-equidistant grid leads to the following system of equations:

$$\begin{aligned} &\hat{h}_j \left(\frac{1}{h_{i-1}} \varphi_{i-1,j} - \left(\frac{1}{h_{i-1}} + \frac{1}{h_i} \right) \varphi_{i,j} + \frac{1}{h_i} \varphi_{i+1,j} \right) \\ &+ \hat{h}_i \left(\frac{1}{h_{j-1}} \varphi_{i,j-1} - \left(\frac{1}{h_{j-1}} + \frac{1}{h_j} \right) \varphi_{i,j} + \frac{1}{h_j} \varphi_{i,j+1} \right) \\ = &- \hat{h}_i \hat{h}_j \frac{\rho_{i,j}}{\varepsilon_0}, \end{aligned} \tag{A11}$$

where we now set $\varphi(x_i, y_j) = \varphi_{i,j}$ and $h_{x,i} = h_i$. The same system of equations is obtained with the application of the FIT. More details along with a generalization to higher dimensions and the treatment of boundaries—also refined, using the Shortley–Weller star—can be found in Ref. [26].

The grid spacings define the coefficients of the potential. These can be assembled in a matrix such that Eq. (A11) can be written in matrix form as

$$\boldsymbol{A}\vec{\varphi} = -\frac{1}{\varepsilon_0}\vec{\rho}. \tag{A12}$$

Here, $\boldsymbol{A}$ is the discrete Laplace operator, which is an $N_x\,N_y \times N_x\,N_y$ (usually sparse) matrix, and $\vec{\varphi}$ and $\vec{\rho}$ are vectors with entries ordered along the grid nodes according to the lexicographical rule

$$\{v_{1,1}, v_{2,1}, \ldots, v_{N_x,1}, v_{1,2}, v_{2,2}, \ldots, v_{N_x,2}, \ldots, v_{1,N_y}, v_{2,N_y}, \ldots, v_{N_x,N_y}\}, \quad v = \varphi, \rho.$$

By solving this system of linear equations it is possible to obtain the electrostatic potential at the nodes of the grid. The electric fields at the internal nodes of the grid can then be computed easily using the central difference formula

$$\begin{aligned}
(E_x)_{i,j} &= -\left(\frac{1}{2h_i}\varphi_{i+1,j} + \left(\frac{1}{2h_{i-1}} - \frac{1}{2h_i}\right)\varphi_i - \frac{1}{2h_{i-1}}\varphi_{i-1,j}\right), \\
(E_y)_{i,j} &= -\left(\frac{1}{2h_j}\varphi_{i,j+1} + \left(\frac{1}{2h_{j-1}} - \frac{1}{2h_j}\right)\varphi_j - \frac{1}{2h_{j-1}}\varphi_{i,j-1}\right).
\end{aligned} \tag{A13}$$

For boundary nodes, a forward difference formula is adopted.

There are many different types of solvers available to solve Eq. (A11) and the optimal solver to be employed depends mostly on the structure of the Laplace matrix $\boldsymbol{A}$. If the matrix is very sparse and if the geometry and the grid do not change in time, a LU decomposition (where 'LU' stands for 'lower upper', and also called LU factorization, see e.g., Ref. [27]), which is a direct solver, can be very efficient. This decomposition needs to be done only once for the Laplace matrix $\boldsymbol{A}$ and the potential can be found from there on for any charge density by forward and backward substitution.

Iterative solvers can become interesting for very large scale problems. They use less memory and are often well suited for parallelization. Iterative solvers start with an initial guess solution and continue iterations until a stopping criterion is satisfied (typically that the error or residual is less than a given tolerance). They then return the final guess solution. For the fast iterative solution of the system there are many alternative algorithms depending on the properties of the Laplace matrix $\boldsymbol{A}$ (primarily symmetry and definiteness). The most straightforward iterative methods are the relaxation-type methods. Typical examples are the Jacobi, Gauss–Seidel and SOR (successive over relaxation) algorithms. These classical iteration methods, also known as stationary methods, however, lack speed in comparison with the more modern non-stationary methods such as the Krylov-subspace methods like CG (conjugate gradient), BiCG (bi-conjugate gradient) and its derivatives, the minimal residual algorithms like the GMRES (generalized minimal residual) or hybrid methods like the BiCGSTAB (bi-conjugate gradient stabilized) [28].

Another often employed and very efficient way of solving the Poisson equation for open boundaries, in particular, employs Green's functions in combination with the fast Fourier transform (FFT) algorithm. This method becomes interesting for situations where boundaries are far away from any source charge distributions up to the limit where open boundaries can be assumed. The 2D electrostatic potential of a point charge in open space is simply given by the Green's function of the 2D Laplacian

$$G(x - x_0, y - y_0) = \frac{1}{2\pi}\left((x - x_0)^2 + (y - y_0)^2\right)^{1/2}. \tag{A14}$$

Then the electrostatic potential can be computed at any point (i, j) on the grid by means of the sum

$$\begin{gathered}
\varphi_{ij} = \frac{e}{\varepsilon_0}\sum_{m,n\neq i,j} \rho_{i,j}\, G(x_i - x_m, y_j - y_n), \\
i, j, m, n \in [0, \ldots, \text{grid point number} - 1].
\end{gathered} \tag{A15}$$

Equation (A15), which runs over all grid points, is computationally often already significantly cheaper compared to Eq. (A1), which, instead, runs over all macroparticles (both are of order $\mathcal{O}(N^2)$, where N is the grid point number or the macroparticle number, respectively). Nevertheless, also the number of grid points can be high and the computation of the potential can be made yet more efficient by using another algorithm. This makes use of the convolution theorem, which states that a convolution in the time domain reduces to a product in the frequency domain, where Eq. (A15), instead, reads

$$\begin{aligned} &\hat{\varphi}_{ij} = \frac{e^2}{2\pi\varepsilon_0}\,\hat{\rho}_{i,j}\cdot\hat{G}_{i,j}, \\ &i,j \in [0,\ldots,\text{grid point number}-1]. \end{aligned} \tag{A16}$$

Equation (A16) scales linearly with the number of grid points (i.e. $\mathcal{O}(N)$). A Fourier transform still needs to be performed on the charge density ρ and the Green's function G as well as an inverse Fourier transform to obtain the potential φ in the time domain. This produces some overhead, which, finally, when employing the FFT, leads to a scaling of the algorithm as $\mathcal{O}(N\log N)$.

The algorithm produces the correct potential for open boundaries by exploiting certain properties of the FFT algorithm. To implement the algorithm correctly, therefore, the computational domain needs to be extended to double its original size along each dimension, i.e. in 2D the area is fourfold. The charge density is left as it is in its original quadrant and set to zero in all other quadrants. The Green's function is mirrored from one quadrant to another—this is required due to the particular way the FFT algorithm works. The multiplication is then done in the frequency domain and the inverse FFT is made to obtain the potential on the extended grid. The potential in the original quadrant is the potential generated by the given charge distribution in an open space whereas the potential in all other quadrants is unphysical, but also not of interest and can, hence, be discarded. This algorithm was first proposed by Hockney [23] and has since then been used for many problems in accelerator physics, e.g. [29].

As a final remark, we mention that using the Green's function for a point charge in open space can be problematic for grids featuring large aspect ratios. The Green's function for a point charge in this case does not well combine with the charge interpolation to the grid points using the commonly employed CIC method. While the Green's function takes the perspective of point-centred charges, the CIC method assumes uniformly distributed charges over one grid cell. Combining the two perspectives will lead to a poor representation of the forces within the grid cells. Especially, along the dimension sampled at lower frequency, the short-range nature of the point-charge Green's function will cause the resulting forces to decay rapidly, when the averaged force within one grid cell would actually not do so. For this reason, a more robust way of computing the forces is to use the integrated Green's function, which essentially is the point-charge Green's function integrated over one grid cell [29]. In 2D the integrated Green's function reads

$$G(x,y) = -\frac{1}{4\pi}\left(-3\,xy + xy\,\log(x^2+y^2) + x^2\arctan\left(\frac{y}{x}\right) + y^2\arctan\left(\frac{x}{y}\right)\right). \tag{A17}$$

Other means of solving the Poisson equation (or more complex PDEs) include finite-element methods which are used preferably on unstructured grids for complicated geometries where the continuous domain is divided into a discrete mesh of elements. The Poisson equation is treated as an eigenvalue problem and initially a trial solution is calculated using basis functions that are localized in each element. The final solution is then obtained by optimization until the required accuracy is reached. We will not discuss this method further here. There is a lot of literature available for the interested reader to embark further into these types of numerical methods.

Acknowledgments

I would like to thank L.R. Carver, G. Iadarola, G. Rumolo and M. Schenk for their help and their valuable input as well as for checking the material. Also, I would like to thank L. Mether, G. Rumolo and M. Schenk for proofreading this document.

References

[1] Y.H. Chin, User's guide for new MOSES version 2.0, CERN/LEP-TH/88-05 (1988).

[2] N. Mounet, DELPHI: an analytic Vlasov solver for impedance-driven modes, HSC Section Meeting, May 2014.

[3] X. Buffat, Transverse beams stability studies at the Large Hadron Collider, CERN-THESIS-2014-246 (2015).

[4] M. Ferrario, HOMDYN (higher order modes dynamics), Workshop on High Average Power & High Brightness Beams, Working Group A, Injector Design, November 2004.

[5] G. Rumolo and F. Zimmermann, Practical user guide for ECLOUD, SL-Note-2002-016-AP (2002).

[6] G. Rumolo and F. Zimmermann, Practical user guide for HEADTAIL, SL-Note-2002-036-AP (2002).

[7] R.D. Ruth, Single particle dynamics in circular accelerators, SLAC-PUB-4104 (1986).

[8] E.D. Courant and H.S. Snyder, *Ann. Phys.* **3** (1958) 1. http://dx.doi.org/10.1016/0003-4916(58)90012-5

[9] S.Y. Lee, *Accelerator Physics*, 3rd ed. (London: World Scientific, 2011). http://dx.doi.org/10.1142/8335.

[10] W. Herr, *Tools for Non-linear Dynamics* (CAS, Poland, 2015).

[11] E. Forest, *Beam Dynamics, A New Attitude and Framework* (Sydney: Harwood, 1998).

[12] A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators*, 1st ed. (Wiley-VCH, 1993).

[13] S. Berg and F. Ruggiero, Landau damping with two-dimensional betatron tune spread, CERN-SL-AP-96-71-AP (1996).

[14] S. Berg and F. Ruggiero, Stability diagrams for Landau damping, LHC Project Report 121 (1997).

[15] W. Herr, in Proceedings of the CAS-CERN Accelerator School: Advanced Accelerator Physics Course, Trondheim, Norway, 18-29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp. 377-404, http://dx.doi.org/10.5170/CERN-2014-009.377

[16] C. Zannini *et al.*, Effects of an asymmetric chamber on the beam coupling impedance, Proc. IPAC2012, 2012.

[17] H. Bartosik, Beam dynamics and optics studies for the LHC injectors upgrade, CERN-THESIS-2013-257 (2013).

[18] G. Iadarola, Electron cloud studies for CERN particle accelerators and simulation code development, CERN-THESIS-2014-047 (2014).

[19] G. Rumolo, Electron cloud, these proceedings, 2016.

[20] C. Birdsall and A. Langdon, *Plasma Physics Via Computer Simulation* (McGraw-Hill, New York, 1985).

[21] H. Qin *et al.*, *Phys. Plasmas* **20** (2013) 084503.

[22] H. Bartosik *et al.*, Benchmarking HEADTAIL with electron cloud instabilities observed in the LHC, Proc. Fifth Electron-cloud Workshop, 2012.

[23] R.W. Hockney and J.W. Eastwood, *Computer Simulation using Particles* (CRC Press, 1988). http://dx.doi.org/10.1887/0852743920

[24] Kitware Inc., The Visualization Toolkit (VTK) data model, 2016, http://www.vtk.org/data-model/.

[25] T. Weiland, *Archiv für Elektronik und Übertragungstechnik* **31** (1977) 116.

[26] A. Markoviḱ, Numerical computation of space-charge fields of electron bunches in a beam pipe of elliptical shape, TESLA Report 2005-21 (2005).

[27] X.S. Li, *ACM Trans. Math. Software* **31**(3) (2005) 302. http://dx.doi.org/10.1145/1089014.1089017

[28] U. van Rienen, *Numerical Methods in Computational Electrodynamics. Linear Systems in Practical Application*, Vol. 12 of *Lecture Notes in Computational Science and Engineering* (Springer, Berlin Heidelberg, 2001). http://dx.doi.org/10.1007/978-3-642-56802-2

[29] J. Qiang *et al.*, *J. Comput. Phys.* **198** (2004) 278. http://dx.doi.org/10.1016/j.jcp.2004.01.008

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Intrabeam Scattering: Anatomy of the Theory

M. Martini
CERN, Geneva, Switzerland

Abstract

This paper starts with an introduction to some elements of physical kinetics relevant to microscopic interactions in gas or plasma systems. The aim is to provide the necessary background for understanding charged particle beams. Emphasis is placed on the important role played by collisions in plasmas. We then give a detailed, albeit non-exhaustive, review of intrabeam scattering (IBS), which consists of the study of diffusion effects caused by multiple Coulomb scattering on charged particle beams, in both the transverse and the longitudinal beam dimensions. Finally, applications to the large hadron collider 7 TeV stored proton beam and the 'Extra Low Energy Antiproton' (ELENA) 100 keV decelerated antiproton beam are used to illustrate the behaviour of the IBS growth rates, for a high-energy storage ring well above the transition energy and an ultra-low-energy decelerator ring below the transition energy.

Keywords

Plasma kinetic; IBS paradigms; Piwinski framework; Bjorken-Mtingwa approach; IBS applications.

1 Introduction

Consider a system of N particles described by a 'density distribution' in a $6N$-dimensional phase space. In this space, the whole collection of N particles is represented by a single point. Under a stability condition, the particles occupy a finite volume in phase space. The density function behaves like an incompressible fluid when observed along the phase space trajectories, provided that all forces driving the system are *conservative* (i.e. derived from a potential): this is the crux of the matter. In practice, instead of using the full $6N$-dimensional space, it is more convenient to view the phase trajectories of all particles in the same six-dimensional phase space, which allows for better visualization of the particle density distribution. In the process of this reduction, however, difficulties could arise with interactions between the particles. In the presence of inter-particle collisions driven by *non-conservative* forces, the phase space volume containing the N particles does not remain incompressible; some particles may escape from it and others may enter it. *Liouville's formula*, modified by adding a 'collision term', becomes the *collision Boltzmann's equation*, which is still an active area of research. The reader may find it helpful to review the topic of particle collisions in plasmas and ionized gases, which are basically quasi-stationary, when we consider the case of a charged particle beam in a storage ring.

In what way is intrabeam scattering (IBS) in charged particle beams different from Coulomb scattering of plasma or gas particles in a 'closed box'? In storage rings, due to the curvature of the orbit, IBS can increase the beam density in phase space as a whole (i.e. transverse beam emittances and momentum spread can grow simultaneously) to above the ring transition energy, because of the coupling between the radial and longitudinal motions. While revisiting the methods used in IBS, our interest was drawn to the logical structure at the core of the subject, so we have decided to base our study on the theoretical framework of A. Piwinsky [1]. His original paper gives a clear presentation of the kinematics of the classical interaction process involved in multiple Coulomb scattering in charged particle beams, assuming weak-focusing accelerator or storage rings. His methodical approach makes all his papers fairly easy to grasp. Equipped with this framework, the essential features of IBS can be understood quite well. The alternative approach of J. Bjorken and S. Mtingwa [2] is based on the scattering matrix

https://doi.org/10.23730/CYRSP-2017-003.291

formalism from quantum electrodynamics, and the theory is applicable to strong-focusing machines. Both of these approaches are presented here, to provide a thorough grounding for the study of many aspects of IBS. However, Monte Carlo simulations of IBS based on a binary collision model [3–5] are not covered in this paper; nor are IBS models that implement the linear coupling between transverse betatron oscillations [6, 7].

2 Overview of the kinetics of gases and plasmas

The classical mechanics branch of dynamics is sometimes divided into kinematics and kinetics. Kinematics is the study of the properties of motion—i.e. the masses or forces which may be involved—without considering the causes, whereas kinetics aims to explain the change in motion as a result of the forces and torques applied. Kinetics refers also to the study of gases and plasmas, relating their macroscopic properties (such as pressure and temperature) to a microscopic model of which the constituents are many small particles. A plasma is a fully or partially ionized gas, i.e. an ensemble of particles made up of electrons, ions and neutrals moving under the influence of electromagnetic or gravity forces and particle interactions such as ionization or Coulomb scattering. Gases and plasmas are mostly non-relativistic. Unlike plasmas, a *plasma beam* is a directed stream of charged particles, generally relativistic, in which the motion depends on applied external fields; in a plasma beam the individual particles make small angles with the beam axis, and the particle energy spread is small relative to the total energy (typical of beams in accelerators and storage rings).

2.1 The Liouville (collisionless Boltzmann) equation

2.1.1 The many-particle phase space joint probability density

The dynamic state of a plasma is completely determined once given the instantaneous positions $\mathbf{r}_i(t)$ and momenta $\mathbf{p}_i(t)$ at time t for all particles in the $6N$-dimensional phase space $\{\mathbf{r} = \mathbf{r}_i(t),\ \mathbf{p} = \mathbf{p}_i(t) : i = 1, \ldots, N\}$, called Γ-space, where N is the number of plasma particles. The state of a plasma is a single point moving along a $6N$-dimensional trajectory $\{\mathbf{r}(t), \mathbf{p}(t)\}$ as time evolves; Γ-space is a many-particle phase space for the whole system of particles, often used in statistical mechanics and kinetic theory [8]. To simplify the notation, we assume that all particles in a plasma are of the same species (e.g. protons, electrons or ions). In a Cartesian frame we have $\{\mathbf{r}_i(t) = x_i\hat{\mathbf{x}} + y_i\hat{\mathbf{y}} + z_i\hat{\mathbf{z}},\ \mathbf{p}_i(t) = p_{x,i}\hat{\mathbf{x}} + p_{y,i}\hat{\mathbf{y}} + p_{z,i}\hat{\mathbf{z}}\}$ where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are unit vectors along the x-, y- and z-axes. The representative single point $M(t) = \prod_{i=1}^{N}\{\mathbf{r}_i(t), \mathbf{p}_i(t)\}$ of the N particles in Γ-space is called the *microstate*. Such a detailed model requires knowledge of $6N$ functions of time with initial conditions that are known only to a certain degree of precision and whose complete description cannot be achieved. When N is very large, say $N = 6 \times 10^{23}$ for one atom-gram of hydrogen, it makes sense to describe the plasma statistically instead.

To this end, we conceptualize a large number $\mathcal{N}$ of independent replicas of the same microstate of an N-particle system, each virtual replica being described by a different representative point $M(t)$ in Γ-space ($\mathcal{N}$ does not have to equal N). The abstract set of identical systems is called a statistical *ensemble*. Let $\mathrm{d}^{6N}\mathcal{N}(\mathbf{r}, \mathbf{p}, t)$ denote the number of representative points inside a infinitesimal phase space volume element $\mathrm{d}\Gamma = \mathrm{d}^{3N}\mathbf{r}\,\mathrm{d}^{3N}\mathbf{p} = \prod_{i=1}^{N}\mathrm{d}^3\mathbf{r}_i\,\mathrm{d}^3\mathbf{p}_i$ about a point $(\mathbf{r}, \mathbf{p})$ at time t, i.e. in the range $\{(\mathbf{r}, \mathbf{r} + \mathrm{d}\mathbf{r}), (\mathbf{p}, \mathbf{p} + \mathrm{d}\mathbf{p})\}$ about $(\mathbf{r}, \mathbf{p})$. Therefore, a normalized phase space probability density function $\rho(\mathbf{r}, \mathbf{p}, t)$ can be formally specified by

$$\rho(\mathbf{r}, \mathbf{p}, t) = \lim_{\mathcal{N}\to\infty} \frac{\mathrm{d}^{6N}\mathcal{N}(\mathbf{r}, \mathbf{p}, t)}{\mathcal{N}\,\mathrm{d}\Gamma}, \qquad \int \mathrm{d}\Gamma\,\rho(\mathbf{r}, \mathbf{p}, t) = 1\,, \tag{1}$$

where $\mathrm{d}\Gamma$ is a $6N$-dimensional hypercube of 'hypersides' $\mathrm{d}^{3N}\mathbf{r}$ and $\mathrm{d}^{3N}\mathbf{p}$.

Comment: As $\rho^{-1}(\mathbf{r}, \mathbf{p}, t)$ has the dimension of a reciprocal 'action', $[(\mathrm{J\,s})^{-1}]$, and because ρ^{-1} is proportional to $\mathrm{d}\Gamma = (\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{p})^{3N}$ it is evocative to attach a Planck constant $h = 2\pi\hbar$ to each pair $\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{p}$ so that $\mathrm{d}\Gamma = (\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{p}/2\pi\hbar)^{3N}$, where $h^{3N} \approx \Delta\mathbf{r}\Delta\mathbf{p}$ is the smallest possible phase space cell size by

virtue of the Heisenberg uncertainty relation. However, this is ill-suited to the framework of classical physics!

This means that $\rho(\mathbf{r},\mathbf{p},t)\,\mathrm{d}\Gamma$ is the probability of finding a non-specific microstate in the volume $\mathrm{d}\Gamma = \mathrm{d}^{3N}\mathbf{r}\,\mathrm{d}^{3N}\mathbf{p}$ about $(\mathbf{r},\mathbf{p})$ at time t; that is, $\rho(\mathbf{r},\mathbf{p},t)\,\mathrm{d}^{3N}\mathbf{r}\,\mathrm{d}^{3N}\mathbf{p}$ is the probability that particle 1 is in the volume $\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{p}_1$ about the point $(\mathbf{r}_1,\mathbf{p}_1)$, particle 2 is in the volume $\mathrm{d}^3\mathbf{r}_2\,\mathrm{d}^3\mathbf{p}_2$ about $(\mathbf{r}_2,\mathbf{p}_2)$, and so on, up to particle N located in the volume $\mathrm{d}^3\mathbf{r}_N\,\mathrm{d}^3\mathbf{p}_N$ about $(\mathbf{r}_N,\mathbf{p}_N)$ at time t. Then ρ is the N-particle joint probability density function for the plasma system. It is assumed that the density of microstates in Γ-space does not change too fast from one volume element to the next, so that $\rho(\mathbf{r},\mathbf{p},t)$ can be regarded as a continuous function. For a finite number $\mathcal{N}$, a so-called coarse-grained density is obtained which disregards the variations of ρ below some small resolution in Γ-space: $\rho(\mathbf{r},\mathbf{p},t) = \mathrm{d}^{6N}\mathcal{N}(\mathbf{r},\mathbf{p},t)/\mathcal{N}\,\mathrm{d}\Gamma$.

Properly normalized, $\rho(\mathbf{r},\mathbf{p},t)$ can be used to compute *macroscopic* values for various functions $\mathcal{O}(\mathbf{r},\mathbf{p})$:

$$\langle\mathcal{O}\rangle = \int \mathrm{d}\Gamma\,\rho(\mathbf{r},\mathbf{p},t)\mathcal{O}(\mathbf{r},\mathbf{p})\;. \tag{2}$$

A key property of an ensemble is that the microstate trajectories never intersect in Γ-space, because each trajectory is uniquely specified by $6N$ initial conditions $\{\mathbf{r}(0),\mathbf{p}(0)\}$. (In case trajectories might be crossed, a phase space configuration situated at the intersection would have multiple trajectories; but this is forbidden in classical physics!). Figure 1 shows two snapshots of the state of an ensemble of $\mathrm{d}^{6N}\mathcal{N}(\mathbf{r},\mathbf{p},t)$ microstate points in Γ-space at times t and $t+\mathrm{d}t$. It illustrates the evolution over the time interval $\mathrm{d}t$ of these microstates which occupy a tiny volume element $\mathrm{d}\Gamma(t) = \prod_{i=1}^{N}\mathrm{d}^3\mathbf{r}_i\,\mathrm{d}^3\mathbf{p}_i$ with border $C(t)$ around $(\mathbf{r},\mathbf{p})$ at time t. Meanwhile, the volume element $\mathrm{d}\Gamma(t)$ can become distorted in shape as a consequence of the particle motion and will occupy a new volume element $\mathrm{d}\Gamma(t+\mathrm{d}t)$ with border $C(t+\mathrm{d}t)$ at time $t+\mathrm{d}t$. Likewise, each microstate represented by a point $M(t) = \prod_{i=1}^{N}\{\mathbf{r}_i(t),\mathbf{p}_i(t)\}$ transforms gradually, to first order in time, into another point $M'(t+\mathrm{d}t)$:

$$M'(t+\mathrm{d}t) = \prod_{i=1}^{N}\left\{\mathbf{r}_i'(t+\mathrm{d}t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\,\mathrm{d}t,\;\mathbf{p}_i'(t+\mathrm{d}t) = \mathbf{p}_i(t) + \dot{\mathbf{p}}_i(t)\,\mathrm{d}t\right\}. \tag{3}$$

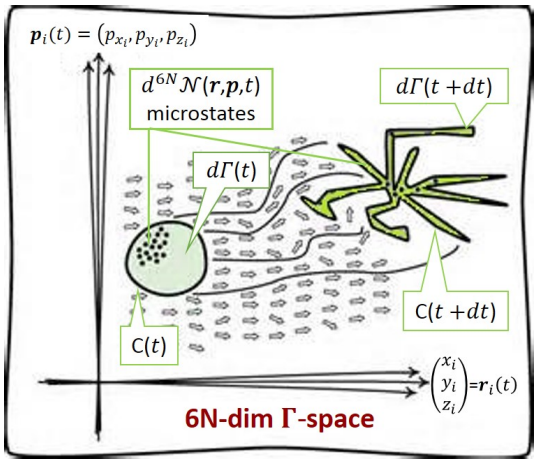


Fig. 1: Depiction of the motion during the time interval $\mathrm{d}t$ of a volume element $\mathrm{d}\Gamma(t) = \mathrm{d}^3\mathbf{r}(t)\,\mathrm{d}^3\mathbf{v}(t)$ about $(\mathbf{r}(t),\mathbf{v}(t))$ in Γ-space.

2.1.2 The single-particle phase space probability density

A simpler way to describe a plasma of N identical particles is to use the phase space for a single particle, called μ-space, in contrast to the Γ-space for the overall particle ensemble; this is the method typically

used in plasma physics, with Cartesian coordinates $\mathbf{r}$ and velocities $\mathbf{v}$ in place of the momenta $\mathbf{p}$, since plasma motion is often non-relativistic. When a particle moves, its representative point $\mu(t)$ traces out a trajectory in six-dimensional phase space such that at each instant the dynamic state of the N plasma particles is represented by N points in the phase space [9]. Likewise, a one-particle probability density function $\rho_1(\mathbf{r},\mathbf{v},t)$ can be defined such that $\mathrm{d}^6N(\mathbf{r},\mathbf{v},t)$ is the number of plasma particles contained in an infinitesimal phase volume element $\mathrm{d}\mu = \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ (a hypercube of sides $\mathrm{d}^3\mathbf{r}$ and $\mathrm{d}^3\mathbf{v}$) about one unspecified representative 'point particle' $\mu(t) = (\mathbf{r},\mathbf{v}) = \{(x,y,z),(v_x,v_y,v_z)\}$ among N in μ-space:

$$\rho_1(\mathbf{r},\mathbf{v},t) = \lim_{N\to\infty}\frac{\mathrm{d}^6N(\mathbf{r},\mathbf{v},t)}{N\,\mathrm{d}\mu}, \qquad \int \mathrm{d}\mu\,\rho_1(\mathbf{r},\mathbf{v},t) = 1\,. \tag{4}$$

Comment: Here, with $\rho_1(\mathbf{r},\mathbf{v},t)$, every 'phase' volume element $(\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{v})^3$ could be replaced by $\mathrm{d}\mu = (\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{v}/2\pi\hbar)^3$.

Like Eq. (2), the distribution function $\rho_1(\mathbf{r},\mathbf{v},t)$ enables one to compute macroscopic values of functions $\mathcal{O}(\mathbf{r},\mathbf{v})$:

$$\langle\mathcal{O}\rangle = \int \mathrm{d}\mu\,\rho_1(\mathbf{r},\mathbf{v},t)\mathcal{O}(\mathbf{r},\mathbf{v})\,. \tag{5}$$

For instance, let us formulate the horizontal beam emittance in terms of the single-particle emittance given by the *Courant–Snyder invariant*

$$\begin{aligned}\varepsilon_x &= \gamma_x x_\beta^2 + 2\alpha_x x_\beta x_\beta' + \beta_x x_\beta'^2 = \frac{x_\beta^2}{\beta_x} + \beta_x\left(x_\beta' - \frac{\beta_x'}{2\beta_x}x_\beta\right)^2,\\ x_\beta &= x - D_x\frac{\Delta p}{p_0}, \quad x_\beta' = x' - D_x'\frac{\Delta p}{p_0}\,,\end{aligned} \tag{6}$$

where $2\alpha_x = -\beta_x',\ \gamma_x = (1+\alpha_x^2)/\beta_x,\ D_x$ is the momentum dispersion function and $x' = p_x/|\mathbf{p}|$. To write down the overall horizontal beam emittance, we switch back to the momentum variable (since a charge particle beam is mostly relativistic), $\mathbf{p} = m\gamma\mathbf{v}$ with $\gamma = (1-\mathbf{v}^2/c^2)^{-1/2}$. Hence, upon averaging Eq. (6) over N particles we obtain, using the notation $\langle\cdots\rangle = \sum_{i=1}^N(\cdots)/N$,

$$\begin{aligned}\varepsilon_x &= \int \mathrm{d}\mu\,\rho_1(x,p,t)\left(\frac{x_\beta^2}{\beta_x} + \beta_x\left[x_\beta' - \frac{\beta_x'}{2\beta_x}x_\beta\right]^2\right)\\ &= \left\langle\frac{x_\beta^2}{\beta_x}\right\rangle\left(1+\frac{\beta_x'^2}{4}\right) - \beta_x'\langle x_\beta' x_\beta\rangle + \beta_x\langle x_\beta'^2\rangle\,.\end{aligned} \tag{7}$$

The equation that governs the evolution of the phase space probability density function under specified initial conditions is generally known as the *Boltzmann equation*. Suppose that each particle of the plasma is subjected to an external force $\mathbf{F}(t)$. In the absence of particle interactions, a particle with coordinates around $\{\mathbf{r},\mathbf{v}\}$ at time t will be found after a time interval $\mathrm{d}t$ around the new coordinates $\{\mathbf{r}',\mathbf{v}'\}$ so that, to first order in t, we obtain

$$\mathbf{r}'(t+\mathrm{d}t) = \mathbf{r}(t) + \mathbf{v}(t)\,\mathrm{d}t\,, \qquad \mathbf{v}'(t+\mathrm{d}t) = \mathbf{v}(t) + \mathbf{a}(t)\,\mathrm{d}t\,, \tag{8}$$

where $\mathbf{v} = \dot{\mathbf{r}}$ and $\mathbf{a} = \dot{\mathbf{v}} = \mathbf{F}/m$ (with $\mathbf{a}$ being the acceleration of the particle and m its mass).

All the particles inside the phase space volume element $\mathrm{d}\mu(t) = \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ with border $C(t)$ about $(\mathbf{r},\mathbf{v})$ at time t will occupy a new volume element $\mathrm{d}^3\mathbf{r}'\mathrm{d}^3\mathbf{v}'$ with border $C'(t+\mathrm{d}t)$ about $(\mathbf{r}',\mathbf{v}')$ after the time interval $\mathrm{d}t$. Figure 2 displays the state of $\mathrm{d}^6N(\mathbf{r},\mathbf{v},t)$ particles of plasma in μ-space at times t and $t+\mathrm{d}t$. Since we are considering the same particles at t and at $t+\mathrm{d}t$, the following equality holds, provided there are no inter-particle collisions or dissipative forces leading to non-conservation of phase space volume:

$$\begin{aligned}&\rho_1(\mathbf{r}',\mathbf{v}',t+\mathrm{d}t)\,\mathrm{d}^3\mathbf{r}'\,\mathrm{d}^3\mathbf{v}'\\ &\quad\equiv \rho_1(\mathbf{r},\mathbf{v},t)\,\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\,.\end{aligned} \tag{9}$$

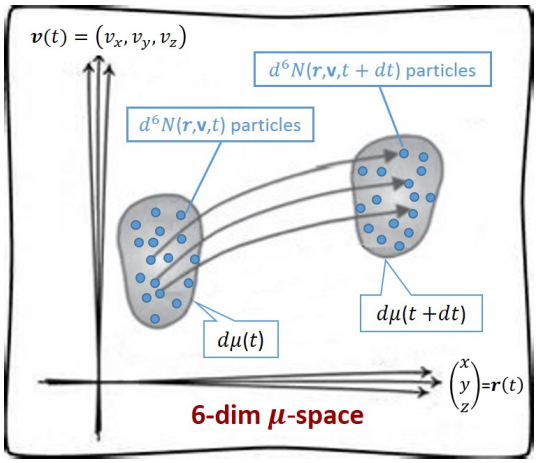


Fig. 2: Depiction of the motion during the time interval $\mathrm{d}t$ of a volume element $\mathrm{d}\mu(t) = \mathrm{d}^3\mathbf{r}(t)\,\mathrm{d}^3\mathbf{v}(t)$ about $(\mathbf{r}(t), \mathbf{v}(t))$ in μ-space.

The shape of the volume $\mathrm{d}\mu(t) = \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ may change into $\mathrm{d}\mu(t+\mathrm{d}t) = \mathrm{d}^3\mathbf{r}'\,\mathrm{d}^3\mathbf{v}'$ during the particle motion over the time interval $\mathrm{d}t$. The new phase volume $\mathrm{d}\mu(t+\mathrm{d}t)$ and the initial one $\mathrm{d}\mu(t)$ are related by the determinant of the 6×6 Jacobian matrix J of the transformation (8):

$$\mathrm{d}^3\mathbf{r}'\mathrm{d}^3\mathbf{v}' = |\det \mathrm{J}|\,\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}, \tag{10}$$

where

$$J = \frac{\partial(\mathbf{r}', \mathbf{v}')}{\partial(\mathbf{r}, \mathbf{v})} = \begin{pmatrix} \partial x'/\partial x & \partial y'/\partial x & \cdots & \partial v_z'/\partial x \\ \partial x'/\partial y & \partial y'/\partial y & \cdots & \partial v_z'/\partial y \\ \vdots & \vdots & \ddots & \vdots \\ \partial x'/\partial v_z & \partial y'/\partial v_z & \cdots & \partial v_z'/\partial v_z \end{pmatrix}. \tag{11}$$

We split the external force $\mathbf{F}$ into (i) a *velocity-independent* force $\tilde{\mathbf{F}}$ (e.g. an external force constraining the particles inside a box or an external electric field force $e\mathbf{E}$ accelerating charged particles) and (ii) a *velocity-dependent* force, customarily the Lorentz force due to an external magnetic field $\mathbf{B}$ confining the particles inside a torus. Thus, leaving $\tilde{\mathbf{F}}$ not fully specified at this stage, we get $\mathbf{F} = \tilde{\mathbf{F}} + e(\mathbf{v} \times \mathbf{B})$. Using Eq. (8) the partial derivatives in the Jacobian matrix can be written in the form

$$\begin{aligned} &\frac{\partial x'_{u=x,y,z}}{\partial x_{w=x,y,z}} = \delta_{u,w}, \qquad \frac{\partial v'_{u=x,y,z}}{\partial x_{w=x,y,z}} = \frac{\mathrm{d}t}{m}\frac{\partial \tilde{F}_u}{\partial x_w} = \frac{\mathrm{d}t}{m}\begin{pmatrix} \partial\tilde{F}_x/\partial x & \partial\tilde{F}_y/\partial x & \partial\tilde{F}_z/\partial x \\ \partial\tilde{F}_x/\partial y & \partial\tilde{F}_y/\partial y & \partial\tilde{F}_z/\partial y \\ \partial\tilde{F}_x/\partial z & \partial\tilde{F}_y/\partial z & \partial\tilde{F}_z/\partial z \end{pmatrix}, \\ &\frac{\partial x'_{u=x,y,z}}{\partial v_{w=x,y,z}} = \delta_{u,w}\,\mathrm{d}t, \quad \frac{\partial v'_{u=x,y,z}}{\partial v_{w=x,y,z}} = \delta_{u,w} + \frac{e\,\mathrm{d}t}{m}\frac{\partial(\mathbf{v}\times\mathbf{B})_{\hat{\mathbf{u}}=\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}}}{\partial v_w}, \end{aligned} \tag{12}$$

where $\delta_{u,w}$ is the Kronecker delta with the subscripts u and w alternately representing the coordinates x, y and z. For clarity, the Lorentz force term in the fourth part of (12) is explicitly written below as a submatrix of the Jacobian:

$$\begin{aligned} \frac{\partial(\mathbf{v}\times\mathbf{B})_{\hat{\mathbf{u}}=\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}}}{\partial v_{w=x,y,z}} &= \frac{\partial}{\partial v_w}\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ v_x & v_y & v_z \\ B_x & B_y & B_z \end{vmatrix}_{\hat{\mathbf{u}}} \\ &= \frac{\partial}{\partial v_{w=x,y,z}}\Big[\big(v_yB_z - v_zB_y\big)\hat{\mathbf{x}} + \big(v_zB_x - v_xB_z\big)\hat{\mathbf{y}} \\ &\quad + \big(v_xB_y - v_yB_x\big)\hat{\mathbf{z}}\Big]_{\hat{\mathbf{u}}=\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}} \end{aligned}$$

$$= \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} .$$

The full Jacobian matrix with its determinant $|\det \mathrm{J}|$ (up to second order in $\mathrm{d}t$) are shown in Eqs. (13) and (14) below:

$$J = \begin{pmatrix} 1 & 0 & 0 & (\mathrm{d}t/m)\,\partial\tilde{F}_x/\partial x & (\mathrm{d}t/m)\,\partial\tilde{F}_y/\partial x & (\mathrm{d}t/m)\,\partial\tilde{F}_z/\partial x \\ 0 & 1 & 0 & (\mathrm{d}t/m)\,\partial\tilde{F}_x/\partial y & (\mathrm{d}t/m)\,\partial\tilde{F}_y/\partial y & (\mathrm{d}t/m)\,\partial\tilde{F}_z/\partial y \\ 0 & 0 & 1 & (\mathrm{d}t/m)\,\partial\tilde{F}_x/\partial z & (\mathrm{d}t/m)\,\partial\tilde{F}_y/\partial z & (\mathrm{d}t/m)\,\partial\tilde{F}_z/\partial z \\ \mathrm{d}t & 0 & 0 & 1 & -B_z\,e\,\mathrm{d}t/m & B_y\,e\,\mathrm{d}t/m \\ 0 & \mathrm{d}t & 0 & B_z\,e\,\mathrm{d}t/m & 1 & -B_x\,e\,\mathrm{d}t/m \\ 0 & 0 & \mathrm{d}t & -B_y\,e\,\mathrm{d}t/m & B_x\,e\,\mathrm{d}t/m & 1 \end{pmatrix} , \tag{13}$$

$$|\det \mathrm{J}| = 1 + \mathrm{dt}^2(\mathrm{e}^2/\mathrm{m}^2)(\mathrm{B}_\mathrm{x}^2 + \mathrm{B}_\mathrm{y}^2 + \mathrm{B}_\mathrm{z}^2) - \big(\partial\tilde{\mathrm{F}}_\mathrm{x}/\partial\mathrm{x} + (\partial\tilde{\mathrm{F}}_\mathrm{y}/\partial\mathrm{y})(\partial\tilde{\mathrm{F}}_\mathrm{z}/\partial\mathrm{z})\big)/\mathrm{m} + \mathcal{O}(\mathrm{dt}^3) . \tag{14}$$

It follows that $|\det \mathrm{J}| = 1$ to first order in $\mathrm{d}t$, and so (10) and (9) become

$$(\mathrm{d}^3\mathbf{r}'\,\mathrm{d}^3\mathbf{v}')(t+\mathrm{d}t) = (d^3\mathbf{r}\,\mathrm{d}^3\mathbf{v})(t), \qquad \big[\rho_1(\mathbf{r}',\mathbf{v}',t+\mathrm{d}t) - \rho_1(\mathbf{r},\mathbf{v},t)\big]\,\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v} = 0\,; \tag{15}$$

that is, volume elements in μ-space are invariants. This is *Liouville's theorem*, which states that the phase space probability density $\rho_1(\mathbf{r},\mathbf{v},t)$ behaves like an incompressible fluid. Using (8), Liouville's theorem (the second equation in (15)) can be rewritten as

$$\rho_1(\mathbf{r}+\mathbf{v}\,\mathrm{d}t,\ \mathbf{v}+\mathbf{a}\,\mathrm{d}t,\ t+\mathrm{d}t) = \rho_1(\mathbf{r},\mathbf{v},t). \tag{16}$$

With these results, the incompressibility condition of Liouville's theorem can be cast in differential form by expanding the left-hand side of (16) in a Taylor series for ρ_1 with respect to the variables $(\mathbf{r},\mathbf{v},t)$, yielding to first order in $\mathrm{d}t$,

$$\begin{aligned} &\rho_1(\mathbf{r}+\mathbf{v}\,\mathrm{d}t,\ \mathbf{v}+\mathbf{a}\,\mathrm{d}t,\ t+\mathrm{d}t) \\ &\quad= \rho_1(\mathbf{r},\mathbf{v},t) + \left[\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\cdot\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial\mathbf{r}} + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\cdot\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial\mathbf{v}}\right]\mathrm{d}t \\ &\quad= \rho_1(\mathbf{r},\mathbf{v},t) + \left[\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \mathbf{v}\cdot\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial\mathbf{r}} + \mathbf{a}\cdot\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial\mathbf{v}}\right]\mathrm{d}t \\ &\quad= \rho_1(\mathbf{r},\mathbf{v},t) + \left[\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \mathbf{v}\cdot\nabla_r\rho_1(\mathbf{r},\mathbf{v},t) + \mathbf{a}\cdot\nabla_v\rho_1(\mathbf{r},\mathbf{v},t)\right]\mathrm{d}t\,, \end{aligned} \tag{17}$$

where notation similar to $\frac{\partial}{\partial\mathbf{u}} = \nabla_u$ (for $\mathbf{u} = \mathbf{r},\mathbf{v}$) has been used for the del operators $\nabla_r = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$ and $\nabla_v = \hat{\mathbf{x}}\frac{\partial}{\partial v_x} + \hat{\mathbf{y}}\frac{\partial}{\partial v_y} + \hat{\mathbf{z}}\frac{\partial}{\partial v_z}$. Substituting the right-hand side of (16) into the left-hand side of (17) gives the so-called *Liouville's formula*

$$\frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \mathbf{v}\cdot\nabla_r\rho_1(\mathbf{r},\mathbf{v},t) + \mathbf{a}\cdot\nabla_v\rho_1(\mathbf{r},\mathbf{v},t) = 0\,. \tag{18}$$

Liouville's formula can also be written in terms of the total derivative $\mathrm{d}\rho_1/\mathrm{d}t$ for the evolution of a volume element as it moves in μ-space (different from $\partial\rho_1/\partial t$, which refers to the change in a volume element at a specific μ-space location):

$$\frac{\mathrm{d}\rho_1(\mathbf{r},\mathbf{v},t)}{\mathrm{d}t} \equiv \frac{\partial\rho_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \mathbf{v}\cdot\nabla_r\rho_1(\mathbf{r},\mathbf{v},t) + \mathbf{a}\cdot\nabla_v\rho_1(\mathbf{r},\mathbf{v},t) = 0\,. \tag{19}$$

This means that $\rho_1(\mathbf{r},\mathbf{v},t)$ is constant along a system phase space trajectory; Eq. (19) is also called the *collisionless Boltzmann equation*.

2.2 The Boltzmann collision integral

The collisionless Boltzmann equation is an equation of motion for a one-particle probability density function $\rho_1(\mathbf{r}, \mathbf{v}, t)$, which is especially suitable for describing dilute gas. In the absence of interactions, the particles are mutually independent and ρ_1 obeys the one-particle Liouville's equation. So far we have introduced the normalized probability density $\rho(\mathbf{r}, \mathbf{v}, t)$ in Γ-space, Eq. (1), and the reduced one-particle probability density $\rho_1(\mathbf{r}, \mathbf{v}, t)$ in μ-space, Eq. (4). Another useful interpretation of the one-particle probability density is obtained when one multiplies ρ_1 by the number of particles N inside a phase space domain in μ-space; this is the one-particle *density distribution* in phase space, denoted by f_1 (see [9]):

$$f_1(\mathbf{r}, \mathbf{v}, t) = \frac{\mathrm{d}^6 N(\mathbf{r}, \mathbf{v}, t)}{\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}}, \qquad \int \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\, f_1(\mathbf{r}, \mathbf{v}, t) = N\,. \tag{20}$$

Here $\mathrm{d}^6 N(\mathbf{r}, \mathbf{v}, t)$ is the number of particles contained within the phase volume $\mathrm{d}\mu(t) = \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ around $(\mathbf{r}, \mathbf{v})$ at time t. Liouville's equation says that if during the time interval $\mathrm{d}t$, we move along with a representative particle in the phase volume element $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ (enclosing N particles) and observe the net number $\delta^6 N$ of particles that enter this volume element, we will find that $\delta^6 N \equiv 0$ (note the difference in meaning of 'd' and 'δ'). Thus, we get

$$\delta^6 N = \left[\frac{\partial f_1(\mathbf{r}, \mathbf{v}, t)}{\partial t} + \mathbf{v}\cdot\nabla_r f_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{a}\cdot\nabla_v f_1(\mathbf{r}, \mathbf{v}, t)\right]\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\,\mathrm{d}t \equiv 0 \tag{21}$$

or, in analogy to (19),

$$\frac{\mathrm{d}f_1(\mathbf{r}, \mathbf{v}, t)}{\mathrm{d}t} \equiv \frac{\partial f_1(\mathbf{r}, \mathbf{v}, t)}{\partial t} + \mathbf{v}\cdot\nabla_r f_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{a}\cdot\nabla_v f_1(\mathbf{r}, \mathbf{v}, t) = 0$$

(where, as before, $\mathbf{a} = \mathbf{F}/m$ is the particle acceleration, with m being the mass of the particle and $\mathbf{F}$ an externally supported force).

The collisionless Boltzmann equation has to be adapted to handle the effects arising from interactions between particles. The Boltzmann collision term discussed below considers only binary elastic collisions. For short-range interactions, two-particle collisions are defined in terms of the pair correlation function $f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t)$. The two colliding particles become dependent, and their density functions $f_1(\mathbf{r}, \mathbf{v}, t)$ and $f_1(\mathbf{r}_1, \mathbf{v}_1, t)$ before collision must be replaced by the two-particle density distribution f_2, which is no longer constant along the phase space trajectories; Eq. (19) needs to be modified by

$$\frac{\mathrm{d}f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t)}{\mathrm{d}t} = \left[\frac{\delta f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t)}{\delta t}\right]_{\text{coll}}. \tag{22}$$

The right-hand side of (22), $[\delta f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t)/\delta t]_{\text{coll}}$, is called the 'collision integral', and designates symbolically the rate of change of the distribution due to the two-particle collisions, which is still to be worked out. A heuristic validation of the Boltzmann equation including collisions in gases and plasmas will be carried out. In the end, this heuristic approach will give the same result as more fundamental derivations. The binary collisions occur in charged and neutral plasmas and involve atoms or molecules in a dilute gas. Multiple coulomb interactions in a plasma, although they may be as important as binary collisions, are ignored here. For binary collisions in the time interval $\mathrm{d}t$, the interaction result is characterized by the net rate at which collisions either decrease or increase the number of particles in a μ-space volume element $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ (cf. [9, 10]).

Figure 3 shows the numbers of particles $\mathrm{d}^6 N_1$ and $\mathrm{d}^6 N_2$ that, at two instants t and $t+\mathrm{d}t$, are within a phase space volume element $\mathrm{d}\mu = \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$, possibly distorted by the particle motion. It illustrates particles entering and leaving the phase volume by virtue of collisions during the interval $\mathrm{d}t$. Some of the particles that were at first in $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ may be removed from it, and particles originally outside this volume element may end up inside it.

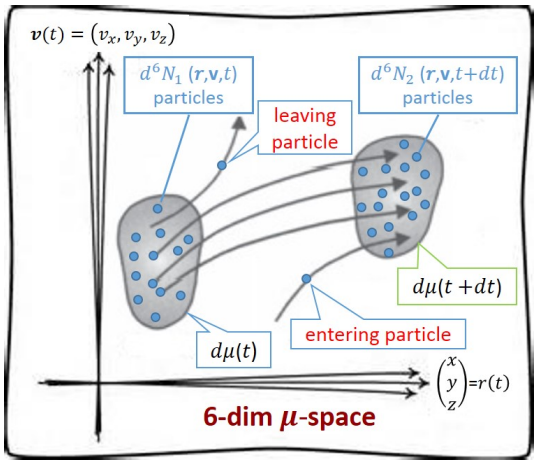


Fig. 3: Evolution of the number of particles in a volume element $\mathrm{d}^3\mathbf{r}(t)\,\mathrm{d}^3\mathbf{v}(t)$ over a time interval $\mathrm{d}t$ in μ-space

The net loss or gain of particles resulting from collisions in the time interval $\mathrm{d}t$ is, by (19) and (22),

$$\delta^6 N = \left[\frac{\delta f_2(\mathbf{r},\mathbf{v},\mathbf{r}_1,\mathbf{v}_1,t)}{\delta t}\right]_{\text{coll}} \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\,\mathrm{d}t, \qquad \delta^6 N = \delta^6 N^+ - \delta^6 N^- \,. \tag{23}$$

Here, $\delta^6 N^-$ is the loss part, due to collisions for which a particle within $\mathrm{d}^3\mathbf{r}$ of $\mathbf{r}$ has a velocity *before* collision that is within $\mathrm{d}^3\mathbf{v}$ of $\mathbf{v}$; similarly, $\delta^6 N^+$ is the gain part, caused by collisions for which a particle within $\mathrm{d}^3\mathbf{r}$ of $\mathbf{r}$ has a velocity *after* collision that is within $\mathrm{d}^3\mathbf{v}$ of $\mathbf{v}$.

For $\delta^6 N^-$, the velocities of the particles may be split into two groups: one contains velocities in the slice $\mathrm{d}^3\mathbf{v}$ about $\mathbf{v}$, and the other includes all other velocities, referred to as $\mathbf{v}_1$. The number of particles removed from the phase volume element $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ in time $\mathrm{d}t$ is the total number of collisions that the particles $\mathbf{v}$ have with all the other particles $\mathbf{v}_1$ during the time interval $\mathrm{d}t$. To compute $\delta^6 N^-$, each collision between a pair of particles must satisfy the following: one particle of the first group (called the $\mathbf{v}$-particle) in the phase volume $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ about $(\mathbf{r},\mathbf{v})$ is scattered out of the velocity slice $\mathrm{d}^3\mathbf{v}$ in the time $\mathrm{d}t$ as a result of a collision with a particle of the second group (a $\mathbf{v}_1$-particle), which has a velocity in $\mathrm{d}^3\mathbf{v}_1$ about $\mathbf{v}_1$ and a location in $\mathrm{d}^3\mathbf{r}_1$ about $\mathbf{r}_1$ (a priori not necessarily the same as $\mathrm{d}^3\mathbf{v}$). Then

$$\delta^6 N^- = \left(\int_{(\mathbf{r}_1,\mathbf{v}_1)} f_2(\mathbf{r},\mathbf{v},\mathbf{r}_1,\mathbf{v}_1,t)\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{v}_1\right)\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\,. \tag{24}$$

Recall that $\mathrm{d}^3\mathbf{r}_1$ must be such that during δt, the $\mathbf{v}_1$-particles in $\mathrm{d}^3\mathbf{r}_1$ experience a collision with the $\mathbf{v}$-particles inside $\mathrm{d}^3\mathbf{r}$. We set up $\mathrm{d}^3\mathbf{r}_1$ by considering a scattering event in the frame of the single $\mathbf{v}$-particle, as shown in Fig. 4. The particles inside $\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{v}_1$ about $(\mathbf{r}_1,\mathbf{v}_1)$ may be viewed as a $\mathbf{v}_1$-particle flux incident on this $\mathbf{v}$-particle. Figure 4 illustrates the scattering where the $\mathbf{v}_1$-particle flux approaches the $\mathbf{v}$-particle from the right at a velocity of $|\mathbf{v}_1-\mathbf{v}|$ with an *impact parameter* between b and $b+\mathrm{d}b$, in a collision plane lying between the angles ϕ and $\phi+\mathrm{d}\phi$. Accordingly, all $\mathbf{v}_1$-particles in the volume of the cylinder of length $|\mathbf{v}_1-\mathbf{v}|\,\mathrm{d}t$ and base area $b\,\mathrm{d}b\,\mathrm{d}\phi$ experience a collision with the $\mathbf{v}$-particles in time $\mathrm{d}t$. So

$$\mathrm{d}^3\mathbf{r}_1 = b\,\mathrm{d}b\,\mathrm{d}\phi\,|\mathbf{v}_1-\mathbf{v}|\,\mathrm{d}t\,. \tag{25}$$

Substituting (25) into (24) transforms the equation into

$$\delta^6 N^- = \left(\int_{(\mathbf{v}_1,b,\phi)} f_2(\mathbf{r},\mathbf{v},\mathbf{r}_1,\mathbf{v}_1,t)\,\mathrm{d}^3\mathbf{v}_1\,|\mathbf{v}_1-\mathbf{v}|\,b\,\mathrm{d}b\,\mathrm{d}\phi\right)\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}\,\mathrm{d}t\,. \tag{26}$$

For $\delta^6 N^+$, consider all particle-pair collisions that send one particle into the velocity slice $\mathrm{d}^3\mathbf{v}$ about $\mathbf{v}$ in the time interval $\mathrm{d}t$, which is the *inverse* of the original collision $(\mathbf{v},\mathbf{v}_1) \rightleftarrows (\mathbf{v}',\mathbf{v}_1')$. The

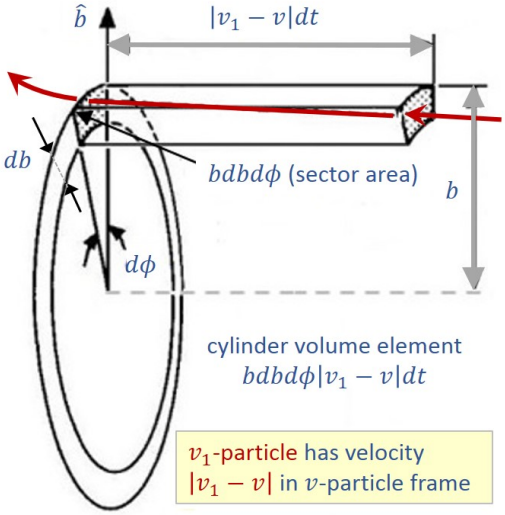


Fig. 4: Sketch of scattering for $\delta^6 N^-$ in the $\mathbf{v}$-particle frame, where a $\mathbf{v}_1$-particle moves from the right towards the $\mathbf{v}$ particle sitting in the vertical collision plane.

primed variables represent the inverse collision of the unprimed ones in (26). In analogy to (24), $\delta^6 N^+$ can be written as

$$\delta^6 N^+ = \left(\int_{(\mathbf{r}_1', \mathbf{v}_1')} f_2(\mathbf{r}', \mathbf{v}', \mathbf{r}_1', \mathbf{v}_1', t)\, \mathrm{d}^3\mathbf{r}_1'\, \mathrm{d}^3\mathbf{v}_1' \right) \mathrm{d}^3\mathbf{r}'\, \mathrm{d}^3\mathbf{v}' \,. \tag{27}$$

Equivalently, from the conservation of momentum and energy for elastic collisions between identical particles, i.e.

$$\mathbf{v} + \mathbf{v}_1 = \mathbf{v}' + \mathbf{v}_1', \quad \mathbf{v}^2 + \mathbf{v}_1^2 = \mathbf{v}'^2 + {\mathbf{v}_1'}^2 \quad \Longrightarrow \quad |\mathbf{v} - \mathbf{v}_1| = |\mathbf{v}' - \mathbf{v}_1'| \,,$$

we obtain

$$\delta^6 N^+ = \left(\int_{(\mathbf{v}_1', b, \phi)} f_2(\mathbf{r}', \mathbf{v}', \mathbf{r}_1', \mathbf{v}_1', t)\, \mathrm{d}^3\mathbf{v}_1'\, |\mathbf{v}_1' - \mathbf{v}'|\, b\, \mathrm{d}b\, \mathrm{d}\phi \right) \mathrm{d}^3\mathbf{r}'\, \mathrm{d}^3\mathbf{v}'\, \mathrm{d}t \,. \tag{28}$$

In order to combine Eqs. (26)–(28) and express $\delta^6 N$ as a single integral with variables of integration $\mathbf{v}_1$, b and ϕ, the integrands of $\delta^6 N^+$ and $\delta^6 N^-$ have to be compatible. To verify this, one can use the phase volume-invariant techniques based on change of variables involving unit Jacobian determinants (see [10] and the above proof of Liouville's theorem, for example). In particular, the phase volume $\mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{v}\, \mathrm{d}^3\mathbf{r}_1\, \mathrm{d}^3\mathbf{v}_1$ is a *collisional invariant*:

$$\mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{v}\, \mathrm{d}^3\mathbf{r}_1\, \mathrm{d}^3\mathbf{v}_1 = \mathrm{d}^3\mathbf{r}'\, \mathrm{d}^3\mathbf{v}'\, \mathrm{d}^3\mathbf{r}_1'\, \mathrm{d}^3\mathbf{v}_1' \,.$$

It follows that

$$\delta^6 N = \left(\int_{(\mathbf{v}_1, b, \phi)} \left[f_2(\mathbf{r}', \mathbf{v}', \mathbf{r}_1', \mathbf{v}_1', t) - f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t) \right] \mathrm{d}^3\mathbf{v}_1\, |\mathbf{v}_1 - \mathbf{v}|\, b\, \mathrm{d}b\, \mathrm{d}\phi \right) \mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{v}\, \mathrm{d}t \,, \tag{29}$$

where there is no extra integration over $\mathbf{v}_1'$ thanks to the above collisional invariant. Substituting (29) into (21) and cancelling the products $\mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{v}\, \mathrm{d}t$, the total time-derivative $\mathrm{d}f_1/\mathrm{d}t$ turns into

$$\begin{aligned} \frac{\mathrm{d}f_1(\mathbf{r}, \mathbf{v}, t)}{\mathrm{d}t} &\equiv \frac{\partial f_1(\mathbf{r}, \mathbf{v}, t)}{\partial t} + \mathbf{v} \cdot \nabla_r f_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{a} \cdot \nabla_v f_1(\mathbf{r}, \mathbf{v}, t) \\ &= \int_{(\mathbf{v}_1, b, \phi)} \left[f_2(\mathbf{r}', \mathbf{v}', \mathbf{r}_1', \mathbf{v}_1', t) - f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t) \right] \mathrm{d}^3\mathbf{v}_1\, |\mathbf{v}_1 - \mathbf{v}|\, b\, \mathrm{d}b\, \mathrm{d}\phi \,, \end{aligned} \tag{30}$$

where the unprimed and primed distributions refer to the states before and after collision, respectively.

Finally, to deduce the Boltzmann equation from Eq. (30) we need a physical approximation enabling us to write f_2 and f_2' in terms of f_1 and f_1', so that (30) will reduce to an expression involving a single integro-differential equation for $f_1(\mathbf{r}, \mathbf{v}, t)$. A conventional derivation of the Boltzmann equation might be considered too heuristic and does not provide a suitable basis for accurate investigation. A better formal derivation of (30) could start out from knowledge of the N-particle joint probability density $\rho_N = \rho$ in the $6N$-dimensional Γ-space (see Eq. (1)), in which the particle velocity $\mathbf{v}$ replaces the momentum $\mathbf{p}$. As the description of a system determined by a full phase density distribution is not feasible, one could instead consider a subset of particles defining a *reduced s-particle* density function f_s. A particularly ingenious method for dealing with such a reduced density function f_s (though not pursued here) is the BBGKY hierarchy. In this formalism, f_1, f_2 and f_s are shown without proof as they will not be used hereafter (notice our way of numbering the N particles $\{\mathbf{r}, \mathbf{v}, \mathbf{r}_i, \mathbf{v}_i\}_{i=1,\ldots,N-1}$, rather than using the usual order $\{\mathbf{r}_i, \mathbf{v}_i\}_{i=1,\ldots,N}$):

$$
\begin{aligned}
f_1(\mathbf{r}, \mathbf{v}, t) &= N \int \prod_{i=1}^{N-1} \mathrm{d}^3\mathbf{r}_i \, \mathrm{d}^3\mathbf{v}_i \, \rho(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, \ldots, \mathbf{r}_{N-1}, \mathbf{v}_{N-1}, t), \\
f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t) &= N(N-1) \int \prod_{i=2}^{N-1} \mathrm{d}^3\mathbf{r}_i \, \mathrm{d}^3\mathbf{v}_i \, \rho(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, \ldots, \mathbf{r}_{N-1}, \mathbf{v}_{N-1}, t), \\
f_s(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, \ldots, \mathbf{r}_{s-1}, \mathbf{v}_{s-1}, t) & \\
&= \frac{N!}{(N-s+1)!} \int \prod_{i=s}^{N-1} \mathrm{d}^3\mathbf{r}_i \, \mathrm{d}^3\mathbf{v}_i \, \rho(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, \ldots, \mathbf{r}_{N-1}, \mathbf{v}_{N-1}, t) \, . \qquad (31)
\end{aligned}
$$

Even so, this way of proceeding is not fully effective in practice. The one-particle density distribution $f_1(\mathbf{r}, \mathbf{v}, t)$ in this sequence is important as it governs the evolution of the collision Boltzmann equation, provided that manageable approximate forms are available for the two-particle density distributions $f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}, \mathbf{v}_1, t)$ and $f_2(\mathbf{r}', \mathbf{v}', \mathbf{r}', \mathbf{v}_1', t)$. Many attempts have been made to derive the Boltzmann equation from first principles without resorting to approximations. However, a number of assumptions come into play in all these derivations, which renders even the more formal analyses (e.g. the BBGKY hierarchy) somewhat ad hoc (see [8, 10–14]).

To find approximate closed-form solutions to Eq. (30) in terms of expressions relating $f_1(\mathbf{r}, \mathbf{v}, t)$ and $f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}, \mathbf{v}_1, t)$, again consider a collision between $\mathbf{v}$- and $\mathbf{v}_1$-particles, which emerge with velocities $\mathbf{v}'$ and $\mathbf{v}_1'$ from the collision. Suppose that the $\mathbf{v}$-particle is located at $(\mathbf{r}, \mathbf{v})$ in the phase volume $\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}$ and the $\mathbf{v}_1$-particle at $(\mathbf{r}, \mathbf{v}_1)$ in the phase volume $\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{v}_1$. This means that collisions are local in space and that the two particles are located at the same point; since $\mathbf{r}$ and $\mathbf{r}'$ can be any points in the respective phase volume elements, we must have $\mathrm{d}^3\mathbf{r}_1 \equiv \mathrm{d}^3\mathbf{r}$. Hence

$$
f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}_1, \mathbf{v}_1, t) \equiv f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}, \mathbf{v}_1, t) \, .
$$

The joint two-particle density distribution f_2 is therefore *homogeneous* over the interaction domain and may be characterized by a circle of radius $\mathbf{r}_{\text{int}}$, considered to be infinitesimal relative to the mean free path of the particles. Evaluating f_2 at the same point $\mathbf{r}$ in space, regardless of the velocity, supposes that the density distribution is below some small resolution in μ-space and does not vary on scales of order $\mathbf{r}_{\text{int}}$. When the particles are sufficiently far away from each other ($|\mathbf{r}| > |\mathbf{r}_{\text{int}}|$), the interaction vanishes. When the particles enter the interaction domain ($|\mathbf{r}| \leq |\mathbf{r}_{\text{int}}|$), they experience a collision. Figures 5 and 6 illustrate the geometry of a particle-pair scattering event involving two particles of velocities $\mathbf{v}$ and $\mathbf{v}_1$, viewed from a coordinate system in which the $\mathbf{v}$-particle is at rest.

As the simplest and most drastic approximation that bypasses all the transformations related to the BBGKY hierarchy, we adopt the 'molecular chaos assumption', which postulates the statistical independence of colliding particles in the derivation of the Boltzmann collision integral. This means that the

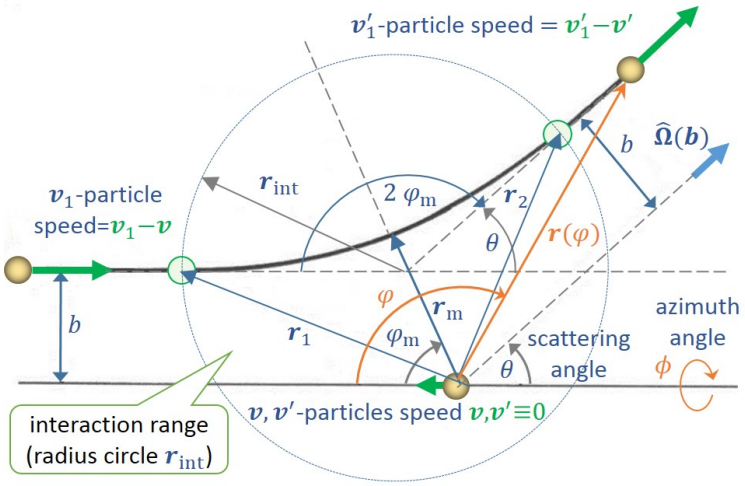


Fig. 5: Particle-pair elastic scattering: the path $\mathbf{r}(\varphi)$ of the $\mathbf{v}_1$-particle of mass m relative to the path of the $\mathbf{v}$-particle of the same mass in a frame where the origin of the $\mathbf{v}$-particle is fixed (equivalent to the centre-of-mass frame, in which the relative velocity of the $\mathbf{v}_1$-particle is $\mathbf{v}_1 - \mathbf{v}$ and that of the $\mathbf{v}$-particle is zero). The position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ mark the entry and exit points of the $\mathbf{v}_1$-particle into and out of the interaction region; $\mathbf{r}_{\rm int}$ is the interaction circle radius; θ is the scattering angle, which depends on the impact parameter b; and $\mathbf{r}_{\rm m}$ is the position of the $\mathbf{v}_1$-particle when it is at the distance of 'closest approach' from the scatterer (the $\mathbf{v}$-particle). The scattering angle θ is related to φ_m via the rule $\theta + 2\varphi_m = \pi$.

particles are assumed to be uncorrelated outside the effective range of their interaction. Therefore, their trajectories before and after a collision are rectilinear. These constraints justify the above homogeneous property, $\mathbf{r} \equiv \mathbf{r}_1$, of f_2 over the interaction range. Moreover, taking into account short-range interactions as well, the pair density distribution $f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}, \mathbf{v}_1, t)$ can be approximated by the product of two single-particle density distribution functions:

$$f_2(\mathbf{r}, \mathbf{v}, \mathbf{r}, \mathbf{v}_1, t) = f_1(\mathbf{r}, \mathbf{v}, t) f_1(\mathbf{r}, \mathbf{v}_1, t) \,, \tag{32}$$

and similarly for $f_2(\mathbf{r}, \mathbf{v}', \mathbf{r}, \mathbf{v}_1', t)$, where f_2 is evaluated at the same point assuming that $\mathbf{r}' \approx \mathbf{r}$ don't vary on scales of order $\mathbf{r}_{\rm int}$.

For example, in an inter-atomic potential the collision duration $\tau_{\rm c}$ is the time over which two particles are within the effective range $\mathbf{r}_{\rm int}$ of their interaction. For particles of a sufficiently dilute gas (speed $v \approx 10^2\,{\rm m\,s^{-1}}$) with short-range interactions $\mathbf{r}_{\rm int} \approx 10^{-10}\,{\rm m}$ (typically an atomic size under standard conditions), the collision time is $\tau_{\rm c} = \mathbf{r}_{\rm int}/v \approx 10^{-12}\,{\rm s}$. From Figs. 5 and 6 we can see that the initial (relative) velocity $\mathbf{v}_1 - \mathbf{v}$ is transformed to the final velocity $\mathbf{v}_1' - \mathbf{v}'$ through the relationship between the impact parameter b and the scattering and azimuth angles (θ, ϕ). The expression for $\mathbf{v}_1'(\mathbf{b}) = \mathbf{v}_1'(\theta, \phi)$ is obtained by integration of the equations of motion. In elastic collisions, the relative velocity $|\mathbf{v}_1 - \mathbf{v}|$ (or $|\mathbf{v}_1|$) just rotates without changing its magnitude to a final direction $\mathbf{v}_1' - \mathbf{v}'$ (or $\mathbf{v}_1'$) indicated by the angles $(\theta, \phi) \equiv \hat{\mathbf{\Omega}}(b)$ (a unit vector). Equivalently, $|\mathbf{v}_1'| = |\mathbf{v}_1|$ and $\mathbf{v}_1' = |\mathbf{v}_1|\,\hat{\mathbf{\Omega}}(b)$,

Substituting (32) into (30) results in the closed-form Boltzmann equation for f_1, involving integrals and partial derivatives of the distribution function:

$$\frac{{\rm d}f_1(\mathbf{r}, \mathbf{v}, t)}{{\rm d}t} = \int_{(\mathbf{v}_1, b, \phi)} \left[f_1(\mathbf{r}, \mathbf{v}', t) f_1(\mathbf{r}, \mathbf{v}_1', t) - f_1(\mathbf{r}, \mathbf{v}, t) f_1(\mathbf{r}, \mathbf{v}_1, t) \right] {\rm d}^3\mathbf{v}_1 \, |\mathbf{v}_1 - \mathbf{v}| \, b \, {\rm d}b \, {\rm d}\phi \,.$$

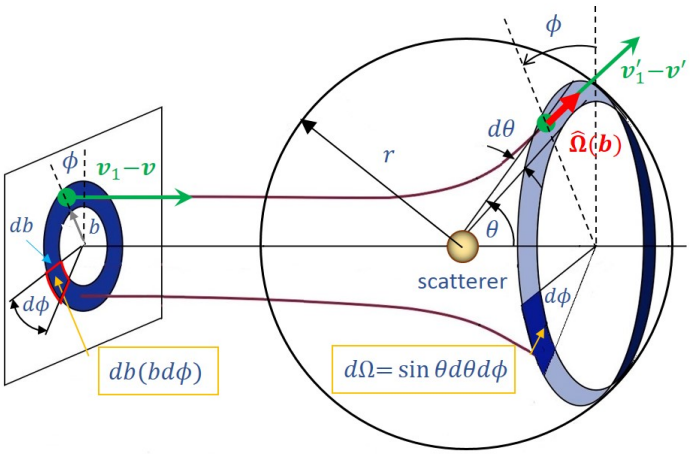


Fig. 6: Particle-pair elastic scattering (different view): particles that hit the small ring (left part of diagram) between impact parameter values b and $b+\mathrm{d}b$ are scattered by an angle between θ and $\theta+\mathrm{d}\theta$ into a larger ring on a sphere (right part of diagram), with a scatterer (the $\mathbf{v}$-particle) at rest in the centre of the sphere (i.e. $\mathbf{v}=\mathbf{v}'=0$). Note that the impact parameter $\mathbf{r}$ is equal to zero for a head-on collision. The incident $\mathbf{v}_1$-particle moves from the left in the direction of the scatterer. As in Fig. 5, the angle θ represents the scattering angle between the $\mathbf{v}_1'$-particle velocity $\mathbf{v}_1'$ and the $\mathbf{v}_1$-particle velocity $\mathbf{v}_1$; ϕ is the azimuth (rotational) angle of $\mathbf{v}_1'-\mathbf{v}'$ about the vertical axis; and $\mathrm{d}\Omega$ is the differential solid angle element of the small area in the large ring of the sphere.

Alternatively, upon expanding the total time-derivative $\mathrm{d}f_1/\mathrm{d}t$,

$$\begin{aligned}&\frac{\partial f_1(\mathbf{r},\mathbf{v},t)}{\partial t}+\mathbf{v}\cdot\nabla_r f_1(\mathbf{r},\mathbf{v},t)+\mathbf{a}\cdot\nabla_v f_1(\mathbf{r},\mathbf{v},t)\\&\quad=\int_{(\mathbf{v}_1,b,\phi)}\left[f_1(\mathbf{r},\mathbf{v}',t)f_1(\mathbf{r},\mathbf{v}_1',t)-f_1(\mathbf{r},\mathbf{v},t)f_1(\mathbf{r},\mathbf{v}_1,t)\right]\mathrm{d}^3\mathbf{v}_1\,|\mathbf{v}_1-\mathbf{v}|\,b\,\mathrm{d}b\,\mathrm{d}\phi\,.\end{aligned}\tag{33}$$

Consider a particle beam of flux (intensity) I (in units of $[\text{particle}/(\text{s m}^2)]$) incident on a scatterer located at the origin. The *differential scattering cross-section* $\sigma(\theta,\phi)$ $[\text{m}^2]$ is defined as the number of particles scattered per second per unit incident flux, at a solid angle oriented in the direction of the outgoing flux after the collision, labelled by the solid angle vector $\hat{\mathbf{\Omega}}$. Geometrically, $\sigma(\theta,\phi)$ can be understood as saying that the number of particles scattered into the solid angle element $\mathrm{d}\Omega$ per unit time is equal to the number of particles crossing an area equal to $\sigma(\theta,\phi)\,\mathrm{d}\Omega$ in the incident beam, as illustrated in Fig. 6, where the differential solid angle element of the small area in the large ring of the sphere is expressed as

$$\begin{aligned}I\sigma(\theta,\phi)\,\mathrm{d}\Omega=\;&\text{number of particles scattered per second into the solid angle element }\mathrm{d}\Omega\\&\text{oriented at }\hat{\mathbf{\Omega}}\text{ with differential solid angle}\\&\mathrm{d}\Omega=r\sin\theta\,\mathrm{d}\phi(r\,\mathrm{d}\theta)/r^2=\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\;\;[\text{rad}^2]\,.\end{aligned}$$

This number of particles is the number traversing the annulus element $b\,\mathrm{d}b\,\mathrm{d}\phi$, so

$$I\sigma(\theta,\phi)\,\mathrm{d}\Omega=Ib\,\mathrm{d}b\,\mathrm{d}\phi\implies\sigma(\theta,\phi)=\frac{b}{\sin\theta}\left|\frac{\mathrm{d}b}{\mathrm{d}\theta}\right|,\tag{34}$$

where the modulus sign is needed because b and θ may change in opposite directions but the cross-section $\sigma(\theta,\phi)$ is always positive. The detailed form of $\sigma(\theta,\phi)$ depends on the inter-particle potential.

Comment: $\sigma(\theta,\phi)$ can depend on the azimuth angle ϕ, but nearly all potentials are spherically symmetric so that $\sigma=\sigma(\theta)$ depends only on θ (or on the impact parameter b). Note that the alternative notation $|\mathrm{d}\sigma/\mathrm{d}\Omega|$ is often used instead of σ for the differential cross-section.

To illustrate the meaning of the differential cross-section with a simple example, consider an elastic collision involving a light point-like particle hitting a heavy hard sphere of radius r. From Fig. 7 it can be seen that the scattering angle θ is linked to the impact parameter through $b = r\sin\varphi_m \equiv r\sin[(\pi-\theta)/2]$, with $b \le r$. The differential cross-section is then calculated from (34) as

$$\begin{aligned}\sigma(\theta,\phi) &= \frac{r}{\sin\theta}\sin\left[\frac{\pi}{2}-\frac{\theta}{2}\right]\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\left(r\sin\left[\frac{\pi}{2}-\frac{\theta}{2}\right]\right)\right| \\ &= \frac{r^2}{2\sin\theta}\sin\left[\frac{\pi}{2}-\frac{\theta}{2}\right]\cos\left[\frac{\pi}{2}-\frac{\theta}{2}\right] \\ &= \frac{r^2}{2\sin\theta}\left(\frac{1}{2}\sin\left[2\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right]\right) = \frac{r^2}{4\sin\theta}\sin\theta \\ \implies\quad \sigma(\theta,\phi) &= \frac{r^2}{4}, \quad \theta = \pi - 2\arcsin\left(\frac{b}{r}\right) . \end{aligned} \tag{35}$$

In this example, the cross-section $\sigma = r^2/4$ does not depend on the scattering angle θ (nor on the azimuth angle ϕ). In particular, $\theta = \pi$ for $b = 0$ (head-on collision) and $\theta = 0$ for $b \ge r$.

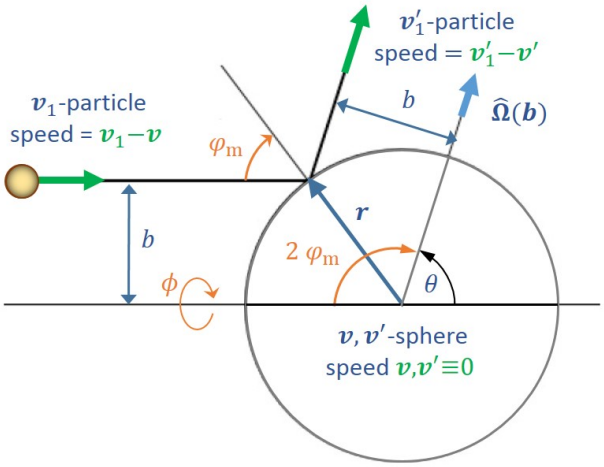


Fig. 7: Tiny elastic particle scattering from a rigid sphere. Irrespective of the impact parameter value $b \le r$, the sphere radius r is equal to the distance of 'closest approach' of the scattering sphere, and $\theta + 2\varphi_m = \pi$.

As another example, illustrated by Figs. 5 and 6, the differential scattering cross-section $\sigma(\theta,\phi)$ for the Coulomb potential $U(r)$ can be cast into the form (see e.g. [9])

$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r}, \quad \sigma(\theta) = \frac{b_0^2}{4\sin^4(\theta/2)}, \quad b_0 = \frac{e^2}{2\pi\epsilon_0 m\,|\mathbf{v}_1 - \mathbf{v}|^2} . \tag{36}$$

The equation for the scattering angle in a Coulomb potential can be written as

$$\theta = 2\arctan\left(\frac{b_0}{b}\right) \iff \tan\left(\frac{\theta}{2}\right) = \frac{b_0}{b} . \tag{37}$$

Notice that b_0 is the value of the impact parameter b for a $\pi/2$ scattering angle (not true for relativistic elastic particle collisions). Also, for $b = 0$ we have $\theta = \pi$.

Replacing $b\,\mathrm{d}b\,\mathrm{d}\phi$ by $\sigma(\theta,\phi)\,\mathrm{d}\Omega$ in Eq. (33), we obtain the following expression for the Boltzmann collision integral:

$$\begin{aligned}&\frac{\partial f_1(\mathbf{r},\mathbf{v},t)}{\partial t} + \mathbf{v}\cdot\nabla_r f_1(\mathbf{r},\mathbf{v},t) + \mathbf{a}\cdot\nabla_v f_1(\mathbf{r},\mathbf{v},t) \\ &= \int_{(\mathbf{v}_1,\theta,\phi)} \left[f_1(\mathbf{r},\mathbf{v}',t) f_1(\mathbf{r},\mathbf{v}_1',t) - f_1(\mathbf{r},\mathbf{v},t) f_1(\mathbf{r},\mathbf{v}_1,t)\right] \mathrm{d}^3\mathbf{v}_1\,|\mathbf{v}_1 - \mathbf{v}|\,\sigma(\theta,\phi)\,\mathrm{d}\Omega\,. \end{aligned} \tag{38}$$

The external force $\mathbf{F} = m\mathbf{a}$ may also include the Lorentz force $\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ due to externally applied fields. If the force is derived from a potential, i.e. $\mathbf{F} = -\nabla_r U(\mathbf{r})$, then (38) can be written in the following form, using $\partial/\partial\mathbf{u} \equiv \nabla_u$ ($\mathbf{u} = \mathbf{r}, \mathbf{v}$):

$$\begin{aligned}&\left[\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}} - \frac{1}{m}\frac{\partial U}{\partial \mathbf{r}}\cdot\frac{\partial}{\partial \mathbf{v}}\right] f_1(\mathbf{r},\mathbf{v},t) \\ &\quad = \int_{(\mathbf{v}_1,\,\theta,\,\phi)} \left[f_1(\mathbf{r},\mathbf{v}',t) f_1(\mathbf{r},\mathbf{v}_1',t) - f_1(\mathbf{r},\mathbf{v},t) f_1(\mathbf{r},\mathbf{v}_1,t)\right] \mathrm{d}^3\mathbf{v}_1\, |\mathbf{v}_1 - \mathbf{v}|\, \sigma(\theta,\phi)\, \mathrm{d}\Omega\,.\end{aligned} \tag{39}$$

The Boltzmann equation is a nonlinear integro-differential equation, which is not easy to solve. The terms on the left-hand side of (39) describe the motion of a single particle in an external potential, while the right-hand side consists of the collision terms. The equation can be physically interpreted as meaning that 'the probability of finding a particle of velocity $\mathbf{v}$ at position $\mathbf{r}$ suddenly changes if that particle experiences a collision with another particle of velocity $\mathbf{v}_1$'. The probability of a collision depends on the differential cross-section σ of the incident particle flux, which is proportional to $|\mathbf{v}_1 - \mathbf{v}|$, and on the joint probability of finding the two particles at $\mathbf{r}$ given their velocities, approximated by $f_1(\mathbf{r},\mathbf{v}) f_1(\mathbf{r},\mathbf{v}_1)$.

To summarize, the 'heuristic derivation' of the one-particle Boltzmann equation has required several strong assumptions, namely:

- the two-particle density distribution is homogeneous over the range of interaction;
- molecular chaos, i.e. the velocities of two colliding particles are uncorrelated;
- the range of particle interaction over the mean free path is much smaller than unity;
- the particle trajectories are rectilinear before and after collision.

2.2.1 The Maxwell–Boltzmann distribution

For gas molecules in a closed box, the modelling approach consists of establishing the equation(s) determining the evolution of the particle density function at equilibrium [8, 9, 13]. An *equilibrium* density distribution function is defined as a solution $f_1(\mathbf{r},\mathbf{v})$ of the Boltzmann equation that has no explicit time dependence and so satisfies $\partial f_1(\mathbf{r},\mathbf{v})/\partial t = 0$. Moreover, one assumes the absence of external forces, i.e. $\mathbf{a} = \mathbf{F}/m \equiv 0$, and a uniform particle distribution in space, i.e. $f_1(\mathbf{r},\mathbf{v})$ is homogeneous so that the density distribution is independent of $\mathbf{r}$ and thus $\nabla_r f_1(\mathbf{v}) = 0$. By Eq. (39), this equilibrium distribution function, denoted by $f_1^{\mathrm{eq}}(\mathbf{v})$, satisfies

$$0 = \int_{(\mathbf{v}_1,\,\theta,\,\phi)} \left[f_1^{\mathrm{eq}}(\mathbf{v}') f_1^{\mathrm{eq}}(\mathbf{v}_1') - f_1^{\mathrm{eq}}(\mathbf{v}) f_1^{\mathrm{eq}}(\mathbf{v}_1)\right] \mathrm{d}^3\mathbf{v}_1\, |\mathbf{v}_1 - \mathbf{v}|\, \sigma(\Omega)\, \mathrm{d}\Omega\,, \tag{40}$$

where $\mathbf{v}_1$ is an arbitrary velocity. It follows that $f_1^{\mathrm{eq}}(\mathbf{v})$, known as the *Maxwell–Boltzmann distribution function*, satisfies the condition

$$f_1^{\mathrm{eq}}(\mathbf{v}') f_1^{\mathrm{eq}}(\mathbf{v}_1') - f_1^{\mathrm{eq}}(\mathbf{v}) f_1^{\mathrm{eq}}(\mathbf{v}_1) = 0\,. \tag{41}$$

Taking the logarithm of (41) yields

$$\ln f_1^{\mathrm{eq}}(\mathbf{v}') + \ln f_1^{\mathrm{eq}}(\mathbf{v}_1') = \ln f_1^{\mathrm{eq}}(\mathbf{v}) + \ln f_1^{\mathrm{eq}}(\mathbf{v}_1)\,. \tag{42}$$

This equation is a summation invariant because the doublets $\{\mathbf{v},\mathbf{v}_1\}$ and $\{\mathbf{v}',\mathbf{v}_1'\}$ are the potential 'initial' and 'final' velocities of a particle-pair collision process $\{\mathbf{v},\mathbf{v}_1\} \to \{\mathbf{v}',\mathbf{v}_1'\}$. Thus, $\ln f_1^{\mathrm{eq}}(\mathbf{v})$ can be formulated by linearly mixing the invariants 'mass' m, 'momentum' $m\mathbf{v}$, and 'kinetic energy' $m|\mathbf{v}|^2/2 \equiv m\mathbf{v}^2/2$ with the constants a_0, a_2 and $\mathbf{a}_1 = a_{1,x}\hat{\mathbf{x}} + a_{1,y}\hat{\mathbf{y}} + a_{1,z}\hat{\mathbf{z}}$ in the form

$$\ln f_1^{\mathrm{eq}}(\mathbf{v}) = m\left(a_0 + \mathbf{a}_1\cdot\mathbf{v} - \frac{a_2\mathbf{v}^2}{2}\right), \tag{43}$$

where the minus sign is introduced for later convenience. This equation can be expressed in a compact form by completing the square on its right-hand side using the formula

$$ax^2 + bx + c \equiv a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right) . \tag{44}$$

Letting $x \to \mathbf{v}$, $x^2 \to \mathbf{v}^2$, $a \to -a_0/2$, $b \to \mathbf{a}_1$, $c \to a_0$, the quadratic equation (43) can be written as

$$\begin{aligned}\ln f_1^{\mathrm{eq}}(\mathbf{v}) &= m\left(-\frac{a_2\mathbf{v}^2}{2} + \mathbf{a}_1\cdot\mathbf{v} + a_0\right)\\ &= -\frac{ma_2}{2}\left(\mathbf{v} - \frac{\mathbf{a}_1}{a_2}\right)^2 + m\left(a_0 + \frac{\mathbf{a}_1^2}{2a_2}\right) = -\frac{ma_2}{2}(\mathbf{v}-\mathbf{v}_0)^2 + \ln C\,,\end{aligned} \tag{45}$$

where the extra constants $\ln C = m(a_0 + \mathbf{a}_1^2/(2a_2))$ and $\mathbf{v}_0 = \mathbf{a}_1/a_2$ have been introduced to simplify the expression. Taking the exponential of $\ln f_1^{\mathrm{eq}}$ gives

$$f_1^{\mathrm{eq}}(\mathbf{v}) = C\exp\left[-\frac{1}{2}ma_2(\mathbf{v}-\mathbf{v}_0)^2\right] , \tag{46}$$

which is the Maxwell–Boltzmann equilibrium distribution function. Equation (46) contains five constant coefficients to be determined, namely C, a_2 and the three components of $\mathbf{v}_0 = v_{0,x}\hat{\mathbf{x}} + v_{0,y}\hat{\mathbf{y}} + v_{0,z}\hat{\mathbf{z}}$. These constants can be deduced from observable physical properties of the system, such as the particle density n, the average velocity $\langle\mathbf{v}\rangle$, the temperature T and the average kinetic energy $\langle m\mathbf{v}^2/2\rangle$.

For a uniformly distributed system of N particles enclosed in a box of volume V (i.e. f_1^{eq} is independent of $\mathbf{r}$), the particle density n is constant and is expressed according to the following normalization conditions:

$$N = \int_{\mathbf{v}} f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{v}, \qquad n = \frac{N}{V} = \int_{\mathbf{v}} f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{v}. \tag{47}$$

Performing the change of variables $\mathbf{w} = \mathbf{v} - \mathbf{v}_0$ with $\mathbf{w} = w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}$, we find that

$$\begin{aligned} n &\equiv C\int_{\mathbf{w}} \exp\left[-\frac{1}{2}ma_2\mathbf{w}^2\right]\mathrm{d}^3\mathbf{w}\\ &= C\int_{-\infty}^{+\infty}\!\!\!\iiint \exp\left[-\frac{1}{2}ma_2(w_x^2+w_y^2+w_z^2)\right]\mathrm{d}w_x\,\mathrm{d}w_y\,\mathrm{d}w_z = C\left(\frac{2\pi}{ma_2}\right)^{3/2} ,\end{aligned} \tag{48}$$

so that C is determined as a function of the constant a_2,

$$C = n\left(\frac{ma_2}{2\pi}\right)^{3/2} . \tag{49}$$

Using (49) and $\mathbf{v} = \mathbf{w} + \mathbf{v}_0$, the average velocity $\langle\mathbf{v}\rangle$ is similarly evaluated as

$$\begin{aligned}\langle\mathbf{v}\rangle &= \frac{\int_{\mathbf{v}}\mathbf{v}f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{v}}{\int_{\mathbf{v}} f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{v}} = \left(\frac{ma_2}{2\pi}\right)^{3/2}\int_{\mathbf{v}}\mathbf{v}\exp\left[-\frac{1}{2}ma_2(\mathbf{v}-\mathbf{v}_0)^2\right]\mathrm{d}^3\mathbf{v}\\ &= \left(\frac{ma_2}{2\pi}\right)^{3/2}\int_{\mathbf{w}}\mathbf{w}\exp\left[-\frac{1}{2}ma_2\mathbf{w}^2\right]\mathrm{d}^3\mathbf{w} + \frac{C}{n}\mathbf{v}_0\int_{\mathbf{w}}\exp\left[-\frac{1}{2}ma_2\mathbf{w}^2\right]\mathrm{d}^3\mathbf{w}\\ &= \left(\frac{ma_2}{2\pi}\right)^{3/2}\int_{-\infty}^{+\infty}\!\!\!\iiint (w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}})\exp\left[-\frac{1}{2}ma_2\left(w_x^2+w_y^2+w_z^2\right)\right]\mathrm{d}w_x\,\mathrm{d}w_y\,\mathrm{d}w_z\end{aligned}$$

$$+\left(\frac{ma_2}{2\pi}\right)^{3/2}\mathbf{v}_0\iiint_{-\infty}^{+\infty}\exp\left[-\frac{1}{2}ma_2\left(w_x^2+w_y^2+w_z^2\right)\right]\mathrm{d}w_x\,\mathrm{d}w_y\,\mathrm{d}w_z \tag{50}$$

$$=\left(\frac{ma_2}{2\pi}\right)^{3/2}\mathbf{v}_0\left(\frac{2\pi}{ma_2}\right)^{3/2}=\mathbf{v}_0\ ,$$

where we have used the fact that the integrand $\mathbf{w}\exp[-(ma_2/2)\mathbf{w}^2]$ in (50) is an odd function of $\mathbf{w}$ so that this integral is equal to zero. Therefore

$$\mathbf{v}_0=\langle\mathbf{v}\rangle.$$

This shows that the constant $\mathbf{v}_0$ represents the mean particle velocity $\langle\mathbf{v}\rangle$. Indeed, for a stationary box containing N gas particles moving at random, there is evidently no particle translational motion as a whole, so $\mathbf{v}_0\equiv 0$.

Finally, let us calculate the average kinetic energy $\langle m\mathbf{v}^2/2\rangle$ where the 'overall particle drift' $\mathbf{v}_0$ is set to zero for convenience (i.e. there is no global particle translational motion). Symmetry considerations suggest that

$$\langle v_x^2\rangle=\langle v_y^2\rangle=\langle v_x^2\rangle\implies\langle\mathbf{v}^2\rangle\equiv\langle v_x^2+v_y^2+v_z^2\rangle=3\langle v_x^2\rangle,$$

so it suffices to compute $\langle v_x^2\rangle$ to obtain $\langle\mathbf{v}^2\rangle$. Using (49),

$$\left\langle\frac{1}{2}m\mathbf{v}^2\right\rangle=\frac{\int_{\mathbf{v}}\frac{m}{2}\mathbf{v}^2f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{v}}{\int_{\mathbf{v}}f_1^{\mathrm{eq}}(\mathbf{v})\,\mathrm{d}^3\mathbf{v}}=\left(\frac{ma_2}{2\pi}\right)^{3/2}\int_{\mathbf{v}}\frac{m\mathbf{v}^2}{2}\exp\left[-\frac{1}{2}ma_2\mathbf{v}^2\right]\mathrm{d}^3\mathbf{v}$$

$$=\frac{1}{2}\left(\frac{ma_2}{2\pi}\right)^{3/2}m\iiint_{-\infty}^{+\infty}\left(v_x^2+v_y^2+v_z^2\right)\exp\left[-\frac{ma_2}{2}\left(v_x^2+v_y^2+v_z^2\right)\right]\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z$$

$$=\frac{1}{2}\left(\frac{ma_2}{2\pi}\right)^{3/2}m\iiint_{-\infty}^{+\infty}3v_x^2\exp\left[-\frac{ma_2}{2}\left(v_x^2+v_y^2+v_z^2\right)\right]\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z$$

$$=\frac{3}{2}\left(\frac{ma_2}{2\pi}\right)^{3/2}m\int_{-\infty}^{+\infty}v_x^2\exp\left[-\frac{ma_2}{2}v_x^2\right]\mathrm{d}v_x$$

$$\times\int_{-\infty}^{+\infty}\exp\left[-\frac{ma_2}{2}v_y^2\right]\mathrm{d}v_y\int_{-\infty}^{+\infty}\exp\left[-\frac{ma_2}{2}v_z^2\right]\mathrm{d}v_z$$

$$=\frac{3}{2}\left(\frac{ma_2}{2\pi}\right)^{1/2}m\int_{-\infty}^{+\infty}v_x^2\exp\left[-\frac{ma_2}{2}v_x^2\right]\mathrm{d}v_x=\frac{3}{2a_2}\ . \tag{51}$$

To quantify the constant a_2, we need to introduce some 'physics'. We use the thermodynamic definition of temperature T, 'experimentally' related to the particle kinetic energy via

$$\frac{3}{2}kT=\left\langle\frac{1}{2}m\mathbf{v}^2\right\rangle\equiv\frac{3}{2a_2}\implies a_2=\frac{1}{kT}\implies C=n\left(\frac{m}{2\pi kT}\right)^{3/2}\ , \tag{52}$$

where k is the Boltzmann constant. Substituting the constants C and a_2 into Eq. (46) and reinserting the overall particle drift constant $\mathbf{v}_0$ for completeness, the equilibrium distribution function $f_1^{\mathrm{eq}}(\mathbf{v})$ in the absence of external forcing becomes

$$f_1^{\mathrm{eq}}(\mathbf{v})=\frac{n}{(2\pi kT/m)^{3/2}}\exp\left[-\frac{m}{2kT}(\mathbf{v}-\mathbf{v}_0)^2\right]\ . \tag{53}$$

3 Intrabeam scattering

3.1 Particle-pair collisions

Intrabeam scattering (IBS), the scattering of particles within a beam, belongs to the category of processes not governed by Liouville's theorem. Other such processes include scattering on residual gas in particle accelerators and storage rings, which leads to a continuous rise in normalized emittance. A related phenomenon is the cooling of one beam by another mixed with it and travelling at the same velocity (e.g. the electron cooling of antiproton beams [15]). A characteristic feature of IBS is the rise time or damping time of the beam dimensions. In some situations IBS leads to the redistribution of partial beam emittances, which can cause undesirable beam dilution in phase space or could heat the beam as a whole (i.e. it may increase the partial beam emittances simultaneously).

Here, we follow the ingenious approach of Ya. S. Derbenev [16] and A. H. Sørensen [17] to studying the self-blowing and damping of a relativistic stored bunch caused by particle-pair collisions. In this process, the scattering between particles induces an energy exchange between transverse and longitudinal motions. Small transverse momenta are transformed into amplified longitudinal fluctuations due to the relativistic Lorentz factor in the transformation.

Comment: If the longitudinal momenta acquired during a single particle-pair collision exceed the momentum acceptance of the RF bucket that keeps the beam bunched, or if the particles hit the aperture when displaced by dispersion, the particles will get lost. This process, referred to as the *Touschek effect*, results in a finite lifetime for a bunched beam.

In greater detail, let us consider a 'simple' Coulomb collision model involving two particles (labelled 1 and 2) with initially *equal* and *opposite* momenta, $\mathbf{p}_{x_{1,2}} = \pm|\mathbf{p}_x|\hat{\mathbf{x}}$. As a result of the $\pi/2$ scattering angle, the two momenta completely transfer into longitudinal momenta (along the s-axis) parallel to the circulating beam. Equilibrium beam conditions at energies below and above the ring transition energy are assessed. In the 'laboratory frame', the transverse components of the two momenta remain unchanged from their 'beam frame' values.

The two opposite post-collision longitudinal momentum components, written as $\Delta\mathbf{p}'_{\|_{1,2}} = \pm\Delta\mathbf{p}'_{\|}$, represent the departures from the beam average momentum $\mathbf{p}_0 = |\mathbf{p}_0|\hat{\mathbf{s}} = \gamma m\mathbf{v}_0$ (in the beam frame $\bar{\mathbf{p}}_0 = |\bar{\mathbf{p}}_0|\hat{\mathbf{s}} = 0$). Viewed in the laboratory frame, $\Delta\mathbf{p}'_{\|_1}$ and $\Delta\mathbf{p}'_{\|_2}$ are both larger by a Lorentz factor $\gamma = (1 - \mathbf{v}_0^2/c^2)^{-1/2}$ than their transverse momentum components before the collision. The factor γ is related to the average beam relativistic energy $E = \gamma mc^2$ and momentum $\mathbf{p}_0 = \gamma m\mathbf{v}_0$, as shown in Fig. 8 (see also [16] and [18]). The following equation expresses the preservation of the sum of the squared momenta in the beam frame after the collision:

$$|\bar{\mathbf{p}}_{\perp_1}|^2 + |\bar{\mathbf{p}}_{\perp_2}|^2 = |\bar{\mathbf{p}}'_{\|_1}|^2 + |\bar{\mathbf{p}}'_{\|_2}|^2 \,.$$

Observe that the momentum deviations in (8) are defined in the two frames (with $|\mathbf{p}_x| = |\bar{\mathbf{p}}_x| = p_x$) as

$$\begin{aligned}
&\text{Beam frame } (\bar{\mathbf{p}}_0 = |\bar{\mathbf{p}}_0|\hat{\mathbf{s}} \stackrel{\text{def}}{=} 0): \quad && \bar{\mathbf{p}}_{\perp_{1,2}} = \bar{\mathbf{p}}_{x_{1,2}} = \mp|\bar{\mathbf{p}}_x|\hat{\mathbf{x}}\,, \\
& && \bar{\mathbf{p}}'_{\|_{1,2}} = \bar{\mathbf{p}}'_{s_{1,2}} = \pm|\bar{\mathbf{p}}_x|\hat{\mathbf{s}}\,; \\
&\text{Laboratory frame } (\mathbf{p}_0 = \gamma m\mathbf{v}_0): \quad && \Delta\mathbf{p}_{\perp_{1,2}} \stackrel{\text{def}}{=} \mathbf{p}_{\perp_{1,2}} = \mathbf{p}_{s_{1,2}} = \mp|\mathbf{p}_x|\hat{\mathbf{x}}\,, \\
& && \Delta\mathbf{p}'_{\|_{1,2}} \stackrel{\text{def}}{=} \mathbf{p}'_{\|_{1,2}} = \mathbf{p}'_{s_{1,2}} = \pm\gamma|\mathbf{p}_x|\hat{\mathbf{s}}\,.
\end{aligned} \tag{54}$$

The horizontal single-particle emittance relates to the Courant–Snyder invariant (see Eq. (6)), where α_x, β_x and γ_x are the Twiss parameters of the lattice, D_x is the momentum dispersion function and $x' = |\mathbf{p}_x|/|\mathbf{p}_0| = p_x/p_0$:

$$\varepsilon_x = \gamma_x x_\beta^2 + 2\alpha_x x_\beta x'_\beta + \beta_x x'^2_\beta = \frac{1+\alpha_x^2}{\beta_x} x_\beta^2 - \beta'_x x_\beta x'_\beta + \beta_x x'^2_\beta \,, \qquad x_\beta = x - D_x\frac{\Delta p}{p_0} \,. \tag{55}$$

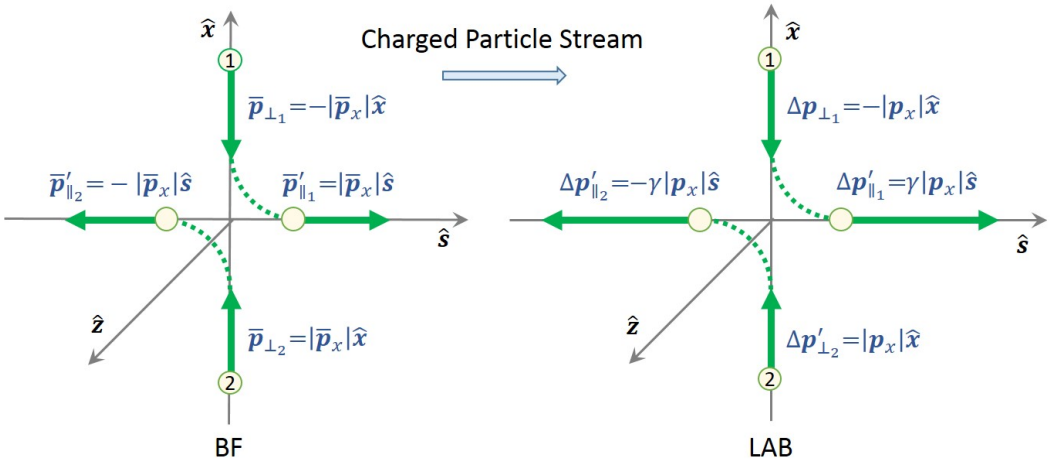


Fig. 8: Particle-pair elastic Coulomb scattering between two particles with momenta $\mathbf{p}_{x,1,2} \equiv \mathbf{p}_{\perp,1,2}$ before collision when viewed in the laboratory frame (LAB), or $\bar{\mathbf{p}}_{\perp,1,2}$ if viewed in the beam frame (BF). After collision the particles' initial momenta become longitudinal, $\pm\mathbf{p}'_{\parallel,1,2}$. A bar on top of a momentum indicates the beam frame, and a prime denotes momentum after collision. The post-collision momenta $\Delta\mathbf{p}'_{\parallel_{1,2}}$ are just the momentum deviations from the average velocity $\mathbf{p}_0$ of the particle stream.

Now, neglecting the derivatives of the lattice parameters β_x and D_x yields $\beta'_x = D'_x = 0$, and the emittance reduces to

$$\varepsilon_x = \frac{1}{\beta_x}\left(x_\beta^2 + \beta_x^2 x_\beta'^2\right) . \tag{56}$$

Assume that the particle interaction point $x \equiv x_{1,2}$ stays constant during the collision time and that the scattering angles $x'_{1,2} = p_{x_{1,2}}/|\mathbf{p}_0|$ vary instantaneously with the momentum change. Then, the kinematic picture of the particle collision can be cast in terms of the betatron amplitude x_β and its derivative x'_β as follows:

$$\begin{aligned} &\text{Before collision:} \quad x_{\beta_{1,2}} \equiv x, && x'_{\beta_{1,2}} = \frac{p_{x_{1,2}}}{p_0} = \pm\frac{p_x}{p_0}. \\ &\text{After collision:} \quad x_{\beta_{1,2}} \equiv x_{1,2} - D_x\frac{\Delta p}{p_0} = x \mp D_x\gamma\frac{p_x}{p_0}, && x'_{1,2} = x'_{\beta_{1,2}} = 0 . \end{aligned} \tag{57}$$

In the above we have used the relations

$$\Delta\mathbf{p}'_{s_{1,2}} = \Delta\mathbf{p}'_{\parallel_{1,2}} = \pm\gamma|\mathbf{p}_{\perp_{1,2}}|\hat{\mathbf{x}} = \pm\gamma|\mathbf{p}_{x_{1,2}}|\hat{\mathbf{x}} = \pm\gamma|\mathbf{p}_x|\hat{\mathbf{x}}, \tag{58}$$

$$D_x\frac{|\Delta\mathbf{p}'_{s_{1,2}}|}{|\mathbf{p}_0|} = \pm\gamma D_x\frac{|\mathbf{p}_x|}{|\mathbf{p}_0|} . \tag{59}$$

Using Eqs. (55)–(59), we can evaluate the change in the sum of emittances of the two colliding particles:

$$\begin{aligned} \beta_x\Delta(\varepsilon_{x_1} + \varepsilon_{x_2}) = \Delta\varepsilon_{x_1} + \Delta\varepsilon_{x_2} &= \left(\varepsilon_{x_1}^{\text{aftcoll}} - \varepsilon_{x_1}^{\text{befcoll}}\right) + \left(\varepsilon_{x_2}^{\text{aftcoll}} - \varepsilon_{x_2}^{\text{befcoll}}\right) \\ &= \left\{\left[\left(x_1 - D_x\gamma\frac{p_{x_1}}{p_0}\right)^2 + 0\right] - \left[x_1^2 + \beta_x^2\left(\frac{p_{x_1}}{p_0}\right)^2\right]\right\} \\ &\quad + \left\{\left[\left(x_2 - D_x\gamma\frac{p_{x_2}}{p_0}\right)^2 + 0\right] - \left[x_2^2 + \beta_x^2\left(\frac{p_{x_2}}{p_0}\right)^2\right]\right\} \end{aligned}$$

$$
\begin{aligned}
&= \left\{ \left[\left(x - D_x \gamma \frac{p_x}{p_0} \right)^2 \right] - \left[x^2 + \beta_x^2 \left(\frac{p_x}{p_0} \right)^2 \right] \right\} \\
&\quad + \left\{ \left[\left(x + D_x \gamma \frac{p_x}{p_0} \right)^2 \right] - \left[x_2^2 + \beta_x^2 \left(-\frac{p_x}{p_0} \right)^2 \right] \right\} \\
&= \left\{ -2x D_x \gamma \frac{p_x}{p_0} + \left(\frac{p_x}{p_0} \right)^2 \left(D_x^2 \gamma^2 - \beta_x^2 \right) \right\} \\
&\quad + \left\{ 2x D_x \gamma \frac{p_x}{p_0} + \left(\frac{p_x}{p_0} \right)^2 \left(D_x^2 \gamma^2 - \beta_x^2 \right) \right\} \\
&= 2 \left(\frac{p_x}{p_0} \right)^2 \left(D_x^2 \gamma^2 - \beta_x^2 \right) , \qquad (60)
\end{aligned}
$$

where $x_1 = x_2 \equiv x$ and $\Delta(\varepsilon_{x_1} + \varepsilon_{x_2})$ symbolically represents the sum of the two particles' emittance variations before and after the collision. Hence

$$
\Delta(\varepsilon_{x_1} + \varepsilon_{x_2}) = \frac{2}{\beta_x} \left(\frac{p_x}{p_0} \right)^2 \left(D_x^2 \gamma^2 - \beta_x^2 \right) = 2 \beta_x \gamma^2 \left(\frac{p_x}{p_0} \right)^2 \left(\frac{D_x^2}{\beta_x^2} - \frac{1}{\gamma^2} \right) . \qquad (61)
$$

The following approximations for the betatron and momentum dispersion functions, referred to as the *smooth focusing approximation*, can be written in the form

$$
\langle \beta_x \rangle \approx \frac{R}{Q_x}, \quad \langle D_x \rangle \approx \frac{\langle \beta_x \rangle}{Q_x} \quad \Longrightarrow \quad \frac{D_x^2}{\beta_x^2} \approx \frac{\langle D_x \rangle^2}{\langle \beta_x \rangle^2} \approx \frac{1}{Q_x^2} . \qquad (62)
$$

Here, R is the mean ring radius and Q_x is the horizontal tune. Introducing the *momentum compaction factor* α_p with associated *transition energy* $\gamma_t mc^2$ and *slip factor* η_t, plus ring curvature radius ρ, we get (see e.g. [18, 19])

$$
\begin{gathered}
\gamma_t \overset{\text{def}}{=} \frac{1}{\sqrt{\alpha_p}}, \qquad \eta_t \overset{\text{def}}{=} \frac{1}{\gamma_t^2} - \frac{1}{\gamma^2}, \\
\alpha_p = \frac{1}{2\pi R} \oint \frac{D_x(s)}{\rho(s)} \, \mathrm{d}s = \left\langle \frac{D_x(s)}{\rho(s)} \right\rangle = \frac{1}{\gamma_t^2} \quad \Longrightarrow \quad \left\langle \frac{D_x(s)}{\rho(s)} \right\rangle \approx \frac{\langle D_x \rangle}{R} \approx \frac{1}{Q_x^2} \approx \frac{1}{\gamma_t^2} . \qquad (63)
\end{gathered}
$$

Note that the contribution to the integral in (63) vanishes in the straight section of the lattice where $\rho(s) \to \infty$. The transition energy is therefore the energy for which the slip factor vanishes. Combining Eq. (61) with the smooth approximations (62) and (63), the change in the sum of the particle emittances becomes

$$
\begin{aligned}
\Delta(\varepsilon_{x_1} + \varepsilon_{x_2}) &= 2 \beta_x \gamma^2 \left(\frac{p_x}{p_0} \right)^2 \left(\frac{\langle D_x \rangle^2}{\langle \beta_x \rangle^2} - \frac{1}{\gamma^2} \right) = 2 \beta_x \gamma^2 \left(\frac{p_x}{p_0} \right)^2 \left(\frac{1}{\gamma_t^2} - \frac{1}{\gamma^2} \right) \\
&= 2 \beta_x \gamma^2 \left(\frac{p_x}{p_0} \right)^2 \eta_t . \qquad (64)
\end{aligned}
$$

In summary:

- Above the transition ($\gamma > \gamma_t$, $\eta_t > 0$): the collisions lead to increased oscillation amplitudes, giving rise to horizontal emittance growth, so that the beam cannot reach an equilibrium.
- Below the transition ($\gamma < \gamma_t$, $\eta_t < 0$): the collisions lead to decreased oscillation amplitudes, giving rise to horizontal emittance reduction, and so a beam equilibrium can exist.

3.2 The original Piwinski IBS model

3.2.1 Introduction

Intrabeam scattering in *weak-focusing* or *smooth ring* lattices can be related to scattering of gas molecules in a closed box, where the walls behave like quadrupole focusing forces and the RF voltage keeps the particles together. The scattering of the molecules leads to the Maxwell–Boltzmann distribution (53) of the three velocity components (v_x, v_z, v_s), where m is the molecule mass, T the temperature, k the Boltzmann's constant, n the volume density of the gas and f_1 the density distribution

$$f_1(v_x, v_z, v_s) = \frac{1}{(2\pi kT/m)^{3/2}} \exp\left[-m(v_x^2 + v_z^2 + v_s^2)/(2kT)\right] \ . \tag{65}$$

The original coordinates $\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}}$ of Eq. (53) have been transformed to the curvilinear coordinate system $v_x\hat{\mathbf{x}} + v_z\hat{\mathbf{z}} + v_s\hat{\mathbf{s}}$, often used to describe particle motion in synchrotrons, where s is the arc length along the reference orbit. The difference between IBS and scattering of gas molecules enclosed in a box is due to the curvature of the ring orbit.

Orbit curvature

- The curvature of the reference orbit produces a *dispersion*, so that a sudden change of energy will change the betatron amplitudes and initiate a *synchro-betatron* oscillation coupling.
- The curvature also gives rise to the *negative mass instability*; that is, when a particle accelerates above transition it becomes slower and behaves like a particle with negative mass, and so an equilibrium of particles above the transition energy cannot exist. Additional comment: In a particle accelerator the *transition energy* $\gamma_t^2 mc^2$ is attained once $\gamma^2 = \gamma_t^2 \equiv \alpha_p^{-1} = (\mathrm{d}p/p)/(\mathrm{d}R/R)$, where the last term of this formula is the ratio of the relative momentum change to the relative orbit radius change.

Above transition

- The IBS effect is to increase the three bunch dimensions; that is, there is a continuous emittance increase in both the transverse and the longitudinal directions.
- For instance, in the LHC at 7 TeV, although $\gamma = 7461 \gg \gamma_t \approx 53.8$ ($\eta_t \approx 3.4{\times}10^{-4}$), the undesirable growth of the bunch emittances caused by IBS is counterbalanced by the *synchrotron radiation damping* effect.

Below transition

- An equilibrium particle distribution between the partial transverse and longitudinal emittances can exist, provided the conditions for the *smooth focusing approximation* hold, i.e. for weak-focusing accelerators and storage rings or for sufficiently smooth lattices. If these conditions do not hold, an equilibrium particle distribution may not exist.
- For example, in the strong-focusing compact ring ELENA, to decelerate at 100 keV and cool the antiprotons sent by the Antiproton Decelerator to give dense beams, a redistribution of partial emittances due to IBS is anticipated, even though $\gamma \approx 1.0 \ll \gamma_t \approx 1.9$ ($\eta_t \approx -0.72$).

3.2.2 Core intrabeam scattering model

The mathematics of IBS is rather complicated. Calculation of the growth rates in each degree of freedom involves integration and averaging procedures that cannot be undertaken entirely analytically and may need to be finished by computer. At present, various IBS computer codes are available which implement the different models developed so far. The inputs are the lattice parameters of the accelerator or storage ring (Twiss parameters, momentum dispersion function and its derivative, etc.) and the beam characteristics (e.g. bunched or coasting beam, number of circulating particles, momentum spread and emittances).

The main outputs are the rise/damping times (or growth/damping rates) for the horizontal, vertical and longitudinal emittances and the momentum spread.

Here we follow the approach of A. Piwinski [1,20] to work out in some detail the growth or decay rates of the beam dimensions due to the IBS effect. The strategy can be outlined in six steps as follows.

Step 1: Transform the momenta of the two colliding particles from the laboratory frame to the beam frame.

Step 2: Calculate the changes in momenta due to an elastic collision.

Step 3: Transform the momenta back to the laboratory frame.

Step 4: Relate the changes in momenta to changes in transverse and longitudinal emittances.

Step 5: Average over the scattering angle distribution using the classical Rutherford cross-section.

Step 6: Average over the distributions of the particle momenta and positions within a bunch.

Steps* 1–3: *Momentum kinematics

In line with Piwinski's approach, the relative momentum changes $\delta\mathbf{p}_{1,2}/|\mathbf{p}|$ from the average particle momentum $\mathbf{p}$ after a collision between two particles (labelled 1 and 2) can be obtained from the first three steps listed above (which is a fairly lengthy task). For brevity, we will omit some details of the calculations.

First of all, the partial longitudinal, horizontal and vertical particle momenta $(p_{s_{1,2}}, p_{x_{1,2}}, p_{z_{1,2}})$ before the collision can be represented in the (s, x, z) coordinate system of the laboratory (or rest) frame (LAB) attached to the reference orbit of the storage ring (supposing that $p_{s_{1,2}} \approx p_{1,2}$):

$$\mathbf{p}_{1,2} = p_{s_{1,2}}\hat{\mathbf{s}} + p_{x_{1,2}}\hat{\mathbf{x}} + p_{z_{1,2}}\hat{\mathbf{z}} = p_{s_{1,2}}\big(\hat{\mathbf{s}} + x'_{1,2}\hat{\mathbf{x}} + z'_{1,2}\hat{\mathbf{z}}\big) \approx p_{1,2}\big(\hat{\mathbf{s}} + x'_{1,2}\hat{\mathbf{x}} + z'_{1,2}\hat{\mathbf{z}}\big)\ , \tag{66}$$

where $x'_{1,2} = p_{x_{1,2}}/p_{s_{1,2}}$ and $z'_{1,2} = p_{z_{1,2}}/p_{s_{1,2}}$ are the betatron angles and $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ are unit vectors parallel to the s, x and z coordinate axes. Then, an extra coordinate system with axes (u, v, w) oriented along the unit vectors $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{z}})$ is defined in the LAB frame for an ensuing Lorentz transformation parallel to $\mathbf{p}_1 + \mathbf{p}_2$ longitudinally, to $\mathbf{p}_1 \times \mathbf{p}_2$ horizontally and to $(\mathbf{p}_1 + \mathbf{p}_2) \times (\mathbf{p}_1 \times \mathbf{p}_2)$ vertically:

$$\hat{\mathbf{u}} = \frac{\mathbf{p}_1 + \mathbf{p}_2}{|\mathbf{p}_1 + \mathbf{p}_2|}, \quad \hat{\mathbf{v}} = \frac{\mathbf{p}_1 \times \mathbf{p}_2}{|\mathbf{p}_1 \times \mathbf{p}_2|}, \quad \hat{\mathbf{w}} = \hat{\mathbf{u}} \times \hat{\mathbf{v}}\ . \tag{67}$$

The particle momenta can then be represented in this coordinate system as

$$\mathbf{p}_{1,2} = p_{s_{1,2}}\big(\cos\alpha_{1,2}\hat{\mathbf{u}} + 0\,\hat{\mathbf{v}} \pm \sin\alpha_{1,2}\hat{\mathbf{w}}\big) \approx p_{1,2}\big(\cos\alpha_{1,2}\hat{\mathbf{u}} + 0\,\hat{\mathbf{v}} \pm \sin\alpha_{1,2}\hat{\mathbf{w}}\big)\ , \tag{68}$$

where $\alpha_{1,2}$ are the angles between the vectors $\mathbf{p}_1 + \mathbf{p}_2$ and $\mathbf{p}_{1,2}$ (see Fig. 9).

Comment: ***Momentum-Energy Lorentz Transformation.*** Let a particle moving at velocity $V_u\hat{\mathbf{u}} \stackrel{\text{def}}{=} \beta_u c\,\hat{\mathbf{u}}$ along the u-axis in the LAB frame (u, v) and at $\bar{V}_u\hat{\bar{\mathbf{u}}} \stackrel{\text{def}}{=} \bar{\beta}_u c\,\hat{\bar{\mathbf{u}}}$ in the CM frame $(\bar{u}, \bar{v})$ (overbars refer to the CM system). The relative velocity parallel to the u-axis between the two inertial frames is denoted $V_r\hat{\mathbf{u}} \stackrel{\text{def}}{=} \beta_r c\,\hat{\mathbf{u}}$. The subscript on the β's and γ's mean $\beta(V_u), \gamma(V_u)$ or $\beta(V_r), \gamma(V_r)$ (the relative frame 'velocity' $V_r\hat{\mathbf{u}}$). The total energy and momentum in the LAB frame are $E = m\gamma_u c^2$ and $p_u = m\gamma_u\beta_u c \equiv E\beta_u/c$, with $\gamma_u = \big(1-\beta_u^2\big)^{-1/2}$. The factor $\bar{\gamma}_u$ stated in term of γ_u and γ_r is needful to compute energy $\bar{E} = m\gamma_{\bar{u}}c^2$ and momentum $\bar{p}_u = m\gamma_{\bar{u}}\beta_{\bar{u}}c$. Using the formula for the additions of velocities $\bar{\beta}_u = (\beta_u - \beta_r)/(1-\beta_u\beta_r)$, we compute $\gamma_{\bar{u}} = \gamma_u\gamma_r(1-\beta_u\beta_r)$. So, the Lorentz transformations from the LAB to the CM frame (and vice versa) on the u- and v-directions $V_v\hat{\mathbf{v}} \stackrel{\text{def}}{=} \beta_v c\,\hat{\mathbf{v}}$ are (cf. [21]):

$$\{\bar{p}_u = \gamma_{\rm r}(p_u - \beta_{\rm r}E/c), \bar{E}/c = \gamma_{\rm r}(E/c - \beta_{\rm r}\,p_u)\}, \quad \{p_u = (\gamma_{\rm r}\,\bar{p}_u + \beta_{\rm r}\bar{E}/c), E/c = \gamma_{\rm r}(\bar{E}/c - \beta_{\rm r}\,\bar{p}_u)\}$$
$$\{\bar{p}_v = p_v, \bar{E}/c = E/c\}, \quad \{p_v = \bar{p}_v, E/c = \bar{E}/c\}$$

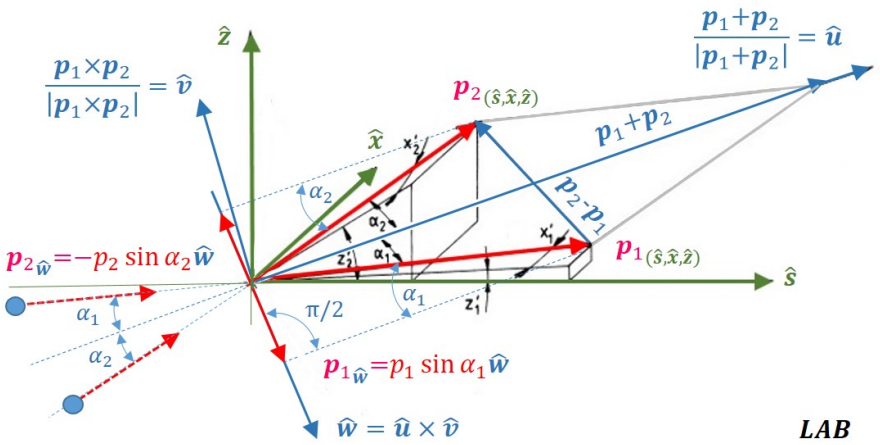


Fig. 9: Relationship between the rest coordinate system $(\hat{\mathbf{s}},\hat{\mathbf{x}},\hat{\mathbf{z}})$ (i.e. LAB frame) and the overlaid $(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{w}})$ coordinate system aligned on the centre-of-mass motion. Geometrically we have $\mathbf{p}_{1_{\hat{\mathbf{w}}},2_{\hat{\mathbf{w}}}}=\pm p_{1,2}\sin\alpha_{1,2}\hat{\mathbf{w}}$, $\mathbf{p}_{1_{\hat{\mathbf{u}}},2_{\hat{\mathbf{u}}}}=p_{1,2}\cos\alpha_{1,2}\hat{\mathbf{u}}$ and $\mathbf{p}_{1_{\hat{\mathbf{v}}},2_{\hat{\mathbf{v}}}}=0$, since $\hat{\mathbf{v}}$ is perpendicular to the plane containing $\mathbf{p}_{1,2}$. Also, $\mathbf{p}_1\cdot\mathbf{p}_2=p_1p_2\cos[\alpha_1+\alpha_2]$ and $p_1\sin\alpha_1=p_2\sin\alpha_2$ (cf. Piwinski [1]).

We now perform the Lorentz transformations along the u-axis parallel to the sum of the two momenta in Fig. 9, $\mathbf{p}_u\overset{\text{def}}{=}\mathbf{p}_1+\mathbf{p}_2$, and the sum of the two energies, $E_u\overset{\text{def}}{=}E_1+E_2$. The velocity of the CM frame $(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{w}})$ relative to the LAB frame $(\hat{\mathbf{s}},\hat{\mathbf{x}},\hat{\mathbf{z}})$ is β_r, with Lorentz factor γ_r, and $\beta_{1,2}c$ and $\gamma_{1,2}$ are the velocities and Lorentz factors of $\mathbf{p}_{1,2}$ in the LAB frame. Hence, the momentum-energy Lorentz transformation parallel to the u-axis gives the momentum in the CM frame, with the help of Eq. (68):

$$\begin{aligned}\bar{\mathbf{p}}_{1,2_{\hat{\mathbf{u}}}}&=\gamma_\mathrm{r}\left(|\mathbf{p}_{1,2_{\hat{\mathbf{u}}}}|-\beta_\mathrm{r}\frac{E_{1,2}}{c}\right)\hat{\mathbf{u}}=\gamma_\mathrm{r}\left(|\mathbf{p}_{1,2_{\hat{\mathbf{u}}}}|-\beta_\mathrm{r}\frac{|\mathbf{p}_{1,2}|}{\beta_{1,2}}\right)\hat{\mathbf{u}}\\&=\gamma_\mathrm{r}\left(|\mathbf{p}_{1,2}|\cos\alpha_{1,2}-\beta_\mathrm{r}\frac{|\mathbf{p}_{1,2}|}{\beta_{1,2}}\right)\hat{\mathbf{u}}=p_{1,2}\gamma_\mathrm{r}\left(\cos\alpha_{1,2}-\frac{\beta_\mathrm{r}}{\beta_{1,2}}\right)\hat{\mathbf{u}}\,,\end{aligned}\tag{69}$$

where the subscript $\hat{\mathbf{u}}$ on the left-hand side refers to the component $\bar{\mathbf{p}}_{1,2}$ along the $\bar{u}$-axis. The relative velocity β_r of the CM frame is fixed by the necessity that $\bar{\mathbf{p}}_1+\bar{\mathbf{p}}_2=0$ (cf. Piwinski [22] Appendix A1). Using Eq. (69) and with $|\mathbf{p}_{1,2}|=m\gamma_{1,2}\beta_{1,2}c$ for the two LAB frame particles one get:

$$\begin{aligned}&|\bar{\mathbf{p}}_{1_{\hat{\mathbf{u}}}}+\bar{\mathbf{p}}_{2_{\hat{\mathbf{u}}}}|=m\gamma_r c(-\beta_r(\gamma_1+\gamma_2)+\beta_1\gamma_1\cos\alpha_1+\beta_2\gamma_2\cos\alpha_2)\equiv 0,\quad\text{solving for }\beta_\mathrm{r}\text{ gives:}\\&\beta_\mathrm{r}=\frac{\beta_1\gamma_1\cos\alpha_1+\beta_2\gamma_2\cos\alpha_2}{\gamma_1+\gamma_2}\equiv\frac{|\mathbf{p}_1+\mathbf{p}_2|c}{E_1+E_2},\quad\gamma_\mathrm{r}=\left(1-\beta_\mathrm{r}^2\right)^{-1/2}=\left(1+\frac{|\mathbf{p}_1+\mathbf{p}_2|^2c^2}{(E_1+E_2)^2}\right)^{-1/2},\\&\gamma_\mathrm{r}^2=\left(1-\beta_\mathrm{r}^2\right)^{-1}=\frac{1}{2}\frac{(\gamma_1+\gamma_2)^2}{1+\gamma_1\gamma_2-\beta_1\gamma_1\beta_2\gamma_2\cos[\alpha_1+\alpha_2]}\approx\frac{\gamma^2}{1+\gamma^2-\beta^2\gamma^2(1-2\alpha^2)}=\frac{\gamma^2}{1+\beta^2\gamma^2\alpha^2}\,.\end{aligned}\tag{70}$$

which in turn yields the approximate relative velocity of the CM frame $\beta_\mathrm{r}=\sqrt{\gamma_\mathrm{r}^2-1}/\gamma_\mathrm{r}\approx\beta(1-\alpha^2/2)$. Also $E_{1,2}=mc^2\gamma_{1,2}$, $p_{1,2}c=mc^2\beta_{1,2}\gamma_{1,2}$, $\cos[\alpha_1+\alpha_2]=\cos 2\alpha$, $\mathbf{p}_1\cdot\mathbf{p}_2=p_1p_2\cos 2\alpha$, $\beta^2\gamma^2=\gamma^2-1$, assuming $p_{s_{1,2}}\approx p_{1,2}$, $x'_{1,2}\approx p_{x_{1,2}}/p\ll 1$, $z'_{1,2}\approx p_{z_{1,2}}/p\ll 1$, and $\alpha_1\approx\alpha_2\approx\alpha$, see Eq. (75), and since:

$$\begin{aligned}&\mathbf{p}_1+\mathbf{p}_2=\left(p_1\cos\alpha_1+p_2\cos\alpha_2\right)\hat{\mathbf{u}}+\left(p_1\sin\alpha_1-p_2\sin\alpha_2\right)\hat{\mathbf{w}}=\left(p_1\cos\alpha_1+p_2\cos\alpha_2\right)\hat{\mathbf{u}}\text{ (Fig. 9)},\\&|\mathbf{p}_1|\sin\alpha_1=|\mathbf{p}_2|\sin\alpha_2,\quad|\mathbf{p}_1+\mathbf{p}_2|=mc\left(\beta_1\gamma_1\cos\alpha_1+\beta_2\gamma_2\cos\alpha_2\right),\quad E_1+E_2=mc^2(\gamma_1+\gamma_2).\end{aligned}$$

Completing Eq. (69) with the Lorentz transformation on the w-axis using Eq. (68), the component on the $\bar{v}$-axis being null, and adding the Lorentz transformation for the energy, we get the three momenta and the energy expressed in the CM frame $(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{w}})$ (Eq. (71) will be set in a more operable form below):

$$\bar{\mathbf{p}}_{1,2}=p_{1,2}\left[\gamma_\mathrm{r}\left(\cos\alpha_{1,2}-\frac{\beta_\mathrm{r}}{\beta_{1,2}}\right)\hat{\mathbf{u}}+0\,\hat{\mathbf{v}}\pm\sin\alpha_{1,2}\hat{\mathbf{w}}\right],\quad\bar{E}_{1,2}=\gamma_\mathrm{r}\left(1-\beta_\mathrm{r}\beta_{1,2}\cos\alpha_{1,2}\right)=\frac{E_1+E_2}{2\gamma_\mathrm{r}}.\tag{71}$$

By symmetry, the changes in momenta of the two colliding particles have the same absolute values but opposite signs in the CM frame, their directions are described by the two angles $\bar{\psi}$ and $\bar{\phi}$ (see Fig. 10).

$$\bar{\mathbf{p}}'_{1,2} = \pm\Big[\big(\bar{p}_{\bar{w}}\sin\bar{\psi}\cos\bar{\phi} + \bar{p}_{\bar{u}}\cos\bar{\psi}\big)\hat{\mathbf{u}} + \bar{p}\sin\bar{\psi}\sin\bar{\phi}\,\hat{\mathbf{v}} + \big(\bar{p}_{\bar{w}}\cos\bar{\psi} - \bar{p}_{\bar{u}}\sin\bar{\psi}\cos\bar{\phi}\big)\hat{\mathbf{w}}\Big] \ . \quad (72)$$

Let us summarize this 'gymnastic' back and forth between the rest and centre-of-mass frames:

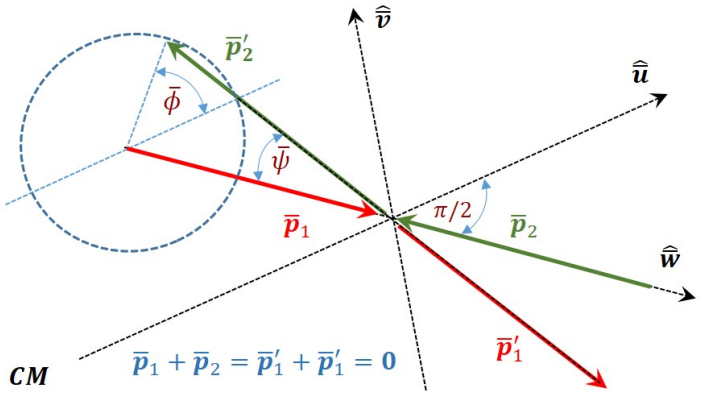


Fig. 10: Changes of momenta in a particle-pair collision in the centre-of-mass coordinate system (Ref. [1])

Before collision (Fig. 9)

(a) LAB frame coordinate system $(\hat{\mathbf{s}}, \hat{\mathbf{x}}, \hat{\mathbf{z}})$: the three momenta of the two particles are given in (66),

$$\mathbf{p}_{1,2} = p_{s_{1,2}}\hat{\mathbf{s}} + p_{x_{1,2}}\hat{\mathbf{x}} + p_{z_{1,2}}\hat{\mathbf{z}} \approx p_{1,2}\big(\hat{\mathbf{s}} + x'_{1,2}\hat{\mathbf{x}} + z'_{1,2}\hat{\mathbf{z}}\big) \ .$$

(b) LAB frame $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$: the three momenta of the two particles reduce to two momenta, which can be written using the angles $\alpha_{1,2}$ between the u-axis and the particle momenta as (68),

$$\mathbf{p}_{1,2} = p_{1,2}\big(\cos\alpha_{1,2}\hat{\mathbf{u}} + 0\,\hat{\mathbf{v}} \pm \sin\alpha_{1,2}\hat{\mathbf{w}}\big) \ .$$

(c) CM frame $(\hat{\bar{\mathbf{u}}}, \hat{\bar{\mathbf{v}}}, \hat{\bar{\mathbf{w}}})$: the two particle momenta in the CM frame calculated by way of a Lorentz transformation parallel to the u-axis can be cast into the form (71) ,

$$\bar{\mathbf{p}}_{1,2} = p_{1,2}\left[\gamma_{\mathrm{r}}\left(\cos\alpha_{1,2} - \frac{\beta_{\mathrm{r}}}{\beta_{1,2}}\right)\hat{\bar{\mathbf{u}}} + 0\,\hat{\bar{\mathbf{v}}} \pm \sin\alpha_{1,2}\hat{\bar{\mathbf{w}}}\right] \ .$$

After collision (Fig. 10)

(a) CM frame $(\hat{\bar{\mathbf{u}}}, \hat{\bar{\mathbf{v}}}, \hat{\bar{\mathbf{w}}})$: the two particle momenta are rotated by the axial and azimuthal scattering angles $(\bar{\psi}, \bar{\phi})$; they are stated in (72),

$$\bar{\mathbf{p}}'_{1,2} = \pm\Big[\big(\bar{p}_{\bar{w}}\sin\bar{\psi}\cos\bar{\phi} + \bar{p}_{\bar{u}}\cos\bar{\psi}\big)\hat{\bar{\mathbf{u}}} + \bar{p}\sin\bar{\psi}\sin\bar{\phi}\,\hat{\bar{\mathbf{v}}} + \big(\bar{p}_{\bar{w}}\cos\bar{\psi} - \bar{p}_{\bar{u}}\sin\bar{\psi}\cos\bar{\phi}\big)\hat{\bar{\mathbf{w}}}\Big] \ .$$

(b) LAB frame $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$: the two momenta are transformed back to the laboratory coordinate system after the collision; they can be expressed as

$$\begin{aligned}\mathbf{p}'_{1,2} = p_{1,2}\Bigg\{&\gamma_{\mathrm{r}}\left[\pm\sin\bar{\psi}\cos\bar{\phi}\sin\alpha_{1,2} + \cos\bar{\psi}\gamma_{\mathrm{r}}\left(\cos\alpha_{1,2} - \frac{\beta_{\mathrm{r}}}{\beta_{1,2}}\right)\right]\hat{\mathbf{u}} \\ &+ \left[\pm\sin\bar{\psi}\sin\bar{\phi}\frac{p}{p_{1,2}} - \sin\bar{\psi}\cos\bar{\phi}\left(\cos\alpha_{1,2} - \frac{\beta_{\mathrm{r}}}{\beta_{1,2}}\right)\right]\hat{\mathbf{v}} + \big[\cos\bar{\phi}\sin\alpha_{1,2}\big]\,\hat{\mathbf{w}}\Bigg\} \ . \quad (73)\end{aligned}$$

Therefore, for the two particles colliding with each other, the change in the particle momenta $\delta\mathbf{p}_{1,2}$ in the laboratory frame $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$ is obtained immediately by subtracting (68) from (73). The result is

$$\begin{aligned}\delta\mathbf{p}_{1,2} &= \mathbf{p}'_{1,2} - \mathbf{p}_{1,2} \\ &= p_{1,2}\Bigg\{\gamma_{\mathrm{r}}\left[\pm\sin\bar{\psi}\cos\bar{\phi}\sin\alpha_{1,2} + (\cos\bar{\psi}-1)\gamma_{\mathrm{r}}\left(\cos\alpha_{1,2} - \frac{\beta_{\mathrm{r}}}{\beta_{1,2}}\right)\right]\hat{\mathbf{u}} \\ &\quad \pm\left[\sin\bar{\psi}\sin\bar{\phi}\frac{p}{p_{1,2}}\right]\hat{\mathbf{v}} - \left[\sin\bar{\psi}\cos\bar{\phi}\left(\cos\alpha_{1,2} - \frac{\beta_{\mathrm{r}}}{\beta_{1,2}}\right) \pm (\cos\bar{\phi}-1)\sin\alpha_{1,2}\right]\hat{\mathbf{w}}\Bigg\}\,. \quad (74)\end{aligned}$$

More approximations are needed to progress further. Let us introduce the three angles ξ, θ and ζ, used as integration variables for the averaging process over the particle distribution inside the bunch. It is assumed that, in the LAB frame, the angle between the two colliding particle momenta is small, and that the derivatives $x'_{1,2}, z'_{1,2}$ of $x_{1,2}, z_{1,2}$ with respect to the longitudinal axis are equal to the small angles the particles make with this s-axis, that is $p_{s_{1,2}} \approx p_{1,2}$, $x'_{1,2} \ll 1$ and $z'_{1,2} \ll 1$, which implies $\alpha_1 \approx \alpha_2 \approx \alpha$ as seen above (cf. also Fig. 9).

$$\begin{gathered}\gamma\xi = \frac{p_1 - p_2}{p}, \quad \theta = \frac{p_{x_1} - p_{x_2}}{p} \equiv x'_1 - x'_2, \quad \zeta = \frac{p_{z_1} - p_{z_2}}{p} \equiv z'_1 - z'_2\,, \\ 2\alpha = \alpha_1 + \alpha_2 = \sqrt{(x'_1 - x'_2)^2 + (z'_1 - z'_2)^2} = \sqrt{\theta^2 + \zeta^2}\,.\end{gathered} \quad (75)$$

Comment: The first expression in (75) can also be written as $\gamma\xi = [(p_1 - p) - (p_2 - p)]/p = \Delta p_1/p - \Delta p_2/p$.

Using Eqs. (70) and (75) with the above approximations and expanding the cosine and sine functions in power series to second-order, Eq. (71) simplifies (see [6, 22] for more details) supposing that $\gamma^2\alpha^2 \ll 1$, where $p = |\mathbf{p}|$ is the average momentum value of all particles in the bunch. One obtain

$$\begin{aligned}\bar{\mathbf{p}}_{1,2} &= \pm\frac{p}{2}\left[\xi\sqrt{1+\beta^2\gamma^2\alpha^2}\hat{\bar{\mathbf{u}}} + 0\hat{\bar{\mathbf{v}}} + 2\alpha\hat{\bar{\mathbf{w}}}\right] \\ &\approx \pm\frac{p}{2}\left[\xi\hat{\bar{\mathbf{u}}} + 0\hat{\bar{\mathbf{v}}} + 2\alpha\hat{\bar{\mathbf{w}}}\right] \approx \pm\frac{p}{2}\left[\frac{p_1 - p_2}{\gamma p}\hat{\bar{\mathbf{u}}} + 0\hat{\bar{\mathbf{v}}} + 2\alpha\hat{\bar{\mathbf{w}}}\right]. \quad (76)\end{aligned}$$

Finally, it remains to formulate the change of momenta $\delta\mathbf{p}_{1,2}$ for the two colliding particles (74) in the laboratory frame $(\hat{\mathbf{s}}, \hat{\mathbf{x}}, \hat{\mathbf{z}})$ of the storage ring coordinate system. After some manipulations the result can be written in the form (see Piwinski [1])

$$\begin{aligned}\frac{\delta\mathbf{p}_{1,2}}{|\mathbf{p}|} = \frac{\mathbf{p}'_{1,2} - \mathbf{p}_{1,2}}{|\mathbf{p}|} &= \frac{\delta p_{s_{1,2}}}{p}\hat{\mathbf{s}} + \frac{\delta p_{x_{1,2}}}{p}\hat{\mathbf{x}} + \frac{\delta p_{z_{1,2}}}{p}\hat{\mathbf{z}} = \frac{1}{2}\left[2\gamma\alpha\cos\bar{\phi}\sin\bar{\psi} + \gamma\xi(\cos\bar{\psi}-1)\right]\hat{\mathbf{s}} \\ &\quad + \frac{1}{2}\left[\left(\zeta\sqrt{1+\frac{\xi^2}{4\alpha^2}}\sin\bar{\phi} - \frac{\xi\theta}{2\alpha}\cos\bar{\phi}\right)\sin\bar{\psi} + \theta(\cos\bar{\psi}-1)\right]\hat{\mathbf{x}} \\ &\quad + \frac{1}{2}\left[\left(\theta\sqrt{1+\frac{\xi^2}{4\alpha^2}}\sin\bar{\phi} - \frac{\xi\zeta}{2\alpha}\cos\bar{\phi}\right)\sin\bar{\psi} + \zeta(\cos\bar{\psi}-1)\right]\hat{\mathbf{z}}\,. \quad (77)\end{aligned}$$

Equation (77) captures the essence of the intrabeam collision process.

Step* 4: *Emittance change induced by momentum change

The change in the particle momenta after collision leads to a parallel change in the particle invariants, namely the longitudinal and transverse emittances. These changes can be calculated by supposing that the transverse particle positions are not altered during the interaction time (assumed to be short enough). From now on we focus the analysis of *bunched beams*. The radial particle movement from the reference closed orbit is the sum of the betatron motion x_β and the momentum deviation contribution from the

product of the momentum dispersion function $D_{x,z}$ and the relative momentum deviation $\Delta p/p$. To simplify the notation, we assume that the reference orbit lies in the plane containing the s- and x-axes; that is, no orbit curvature takes place in the vertical direction z, so that $D_z(s) \equiv 0$. Then

$$x = x_\beta + D_x \frac{\Delta p}{p}, \quad z = z_\beta \quad \text{and} \quad x' \equiv \frac{p_x}{p} = x'_\beta - D'_x \frac{\Delta p}{p}, \quad z' \equiv \frac{p_z}{p} = z'_\beta \,. \tag{78}$$

The invariants of the motion are the single-particle *transverse emittances* $\varepsilon_{x,z}$ (identified with the *Courant–Snyder invariant*) and the *longitudinal invariant* H for a bunched beam:

$$\varepsilon_x = \gamma_x x_\beta^2 + 2\alpha_x x_\beta x'_\beta + \beta_x x'^2_\beta \quad \text{and} \quad H = \left(\frac{\Delta p}{p}\right)^2 + \frac{1}{\Omega^2}\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\Delta p}{p}\right)\right]^2 , \tag{79}$$

where α_x, β_x and γ_x are the *Twiss parameters*, with $\beta_x\gamma_x - \alpha_x^2 = 1$ and $2\alpha_x = -\beta'_x$, and Ω is the *synchrotron frequency*.

The change $\delta\varepsilon_x$ in the invariant ε_x after collision is given by

$$\delta\varepsilon_x = \gamma_x\big(2x_\beta\delta x_\beta + \delta x_\beta^2\big) + 2\alpha_x\big(x'_\beta\delta x_\beta + x_\beta\delta x'_\beta + \delta x_\beta\delta x'_\beta\big) + \beta_x\big(2x'_\beta\delta x'_\beta + \delta x'^2_\beta\big) \,. \tag{80}$$

Replace x with z to get $\delta\varepsilon_z$.

It has already been assumed that the ring lattice is vertically free of dispersion, i.e. $D_z = D'_z = 0$. Suppose also that $x_{1,2}$ and $z_{1,2}$ stay constant during the short collision time, so that only $x'_{1,2}$ and $z'_{1,2}$ vary with the change in momentum. Now $\delta(\Delta p/p) = \delta p/p$, as the average (or reference) momentum $p_s = |\mathbf{p}_s|$ remains constant if the beam is not accelerated.

Comment: Indeed, $\delta(\Delta p/p) = \delta[(p - p_s)/p_s] = \delta p/p_s - \delta p_s/p_s = \delta p/p_s \approx \delta p/p$ as p_s is constant. Defining $\eta = \Delta p/p$ gives $\delta\eta \approx \delta p/p$.

So the variations δx_β, $\delta x'_\beta$ and $\delta z'_\beta$ can be formulated in terms of the betatron amplitudes as

$$\delta x_\beta = -D_x\frac{\delta p}{p}, \qquad \delta x'_\beta = \frac{\delta p_x}{p} - D'_x\frac{\delta p}{p}, \qquad \delta z'_\beta = \frac{\delta p_z}{p} \,, \tag{81}$$

where, for instance, we have used the 'trick'

$$\delta x = \delta x_\beta + D_x\frac{\Delta p}{p} = \delta x_\beta + D_x\delta\left[\frac{\Delta p}{p}\right] = \delta x_\beta + D_x\frac{\delta p}{p} \equiv 0 \implies \delta x_\beta = -D_x\frac{\delta p}{p} \,.$$

the changes $\delta\varepsilon_{x,z}$ and δH of $\varepsilon_{x,z}$ and H can be written in the form

$$\begin{aligned}\frac{\delta\varepsilon_x}{\beta_x} &= -\frac{2}{\beta_x}\Big[x_\beta(\gamma_x D_x + \alpha_x D'_x) + x'_\beta\tilde{D}_x\Big]\frac{\delta p}{p} \\ &\quad + \frac{D_x^2 + \tilde{D}_x^2}{\beta_x^2}\left(\frac{\delta p}{p}\right)^2 + 2\left(x'_\beta + \frac{\alpha_x}{\beta_x}x_\beta\right)\frac{\delta p_x}{p} + \left(\frac{\delta p_x}{p}\right)^2 - \frac{2\tilde{D}_x}{\beta_x}\frac{\delta p}{p}\frac{\delta p_x}{p} \,, \end{aligned} \tag{82}$$

$$\frac{\delta\varepsilon_z}{\beta_z} = 2\left(z'_\beta + \frac{\alpha_z}{\beta_z}z_\beta\right)\frac{\delta p_z}{p} + \left(\frac{\delta p_z}{p}\right)^2, \qquad \delta H = 2\frac{\Delta p}{p}\frac{\delta p}{p} + \left(\frac{\delta p}{p}\right)^2 \,, \tag{83}$$

where $\tilde{D}_x = \alpha_x D_x + \beta_x D'_x$ and we have neglected possible time variation of the synchrotron frequency during the collision.

Comment: Likewise, assuming that $\mathrm{d}\eta/\mathrm{d}t = 0$ at the collision time, we have $H \equiv \eta^2$ and $\delta H = (\eta + \delta\eta)^2 - \eta^2 = 2\eta\delta\eta + (\delta\eta)^2 = 2\eta\delta p/p + (\delta p/p)^2$ as $\delta\eta \equiv \delta(\Delta p/p) \approx \delta p/p$.

In the presence of radial dispersion, the momentum change δp of the particle during the collision leads to a change in the horizontal emittance given by Eq. (82). In what follows, the shorthand $\eta_{1,2} \stackrel{\text{def}}{=} \Delta p_{1,2}/p_{1,2}$ or $\eta = \Delta p/p$ will be used.

Step* 5: *Averaging over the scattering angles

The variation of the beam phase space volume can be calculated by averaging the change of the particle invariants over all the collisions in accordance with the particle distribution function. Piwinski defined the time-derivative $\mathrm{d}\langle\varepsilon_x\rangle/\mathrm{d}\bar{t}$ of the average radial emittance $\langle\varepsilon_x\rangle$ in the CM frame for all particles by means of a stepwise integration process.

Stated in detail (see [1, 3]): "To calculate the mean value of the emittance and momentum deviation change for one particle we have to average with respect to the second particle betatron angles and momentum deviations. Thus, to derive the overall mean value of the emittance and momentum deviation change for all particles we have to average further with respect to all betatron angles, momentum deviations and positions of the first particle. This means that we have to integrate over all phase space betatron coordinates, momentum spread values and azimuthal locations of two interacting particles, by means of a probability density function P (in the rest frame) for the betatron amplitudes and angles, the momentum deviations and the azimuthal positions of the interacting particles. This is done by integrating $\bar{P}$ (in the CM frame) over the particle bunch phase space volume $\bar{V}$ with respect to the differential phase space volume element $\mathrm{d}\bar{V}$:

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}\bar{t}} \frac{\langle\varepsilon_{x_1}\rangle}{\beta_x} \right\rangle = \int_{\bar{V}} 2c\bar{\beta}\bar{P}\,\mathrm{d}\bar{V} \int_{\bar{\psi}_{\min}}^{\pi} \mathrm{d}\bar{\psi} \int_0^{2\pi} \mathrm{d}\bar{\phi}\,\bar{\sigma}(\bar{\psi}) \frac{\delta\varepsilon_{x_1}}{\beta_x} \sin\bar{\psi}, \tag{84}$$

where the outer brackets $\langle\,\cdot\,\rangle$ denote the average value over the lattice parameters of the ring. Here $\bar{\sigma}(\psi)$ is the differential Coulomb scattering cross-section for the scattering into a solid angle element $\mathrm{d}\bar{\Omega}(\bar{\psi},\bar{\phi})$ in the CM frame, $\mathrm{d}\bar{t}$ and $\mathrm{d}t$ are the *proper time* intervals in the CM and rest LAB frames, such that $\mathrm{d}t = \gamma\,\mathrm{d}\bar{t}$, c is the speed of light, and $2c\bar{\beta}$ is the relative velocity of two interacting particles (labelled 1 and 2) with velocities $\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 = 0$ in the CM frame. The probability density function P is defined by a product of 12 variables and can be formulated in the LAB frame as

$$P_{12\text{var}} = P_{\eta s}(\eta_1, s_1) P_{\eta s}(\eta_2, s_2) P_{x_\beta x'_\beta}(x_{\beta_1}, x'_{\beta_1}) P_{x_\beta x'_\beta}(x_{\beta_2}, x'_{\beta_2}) P_{zz'}(z_1, z'_1) P_{zz'}(z_2, z'_2)\,,$$

and the two-particle infinitesimal phase space volume element is

$$\mathrm{d}V_{12\text{var}} = \mathrm{d}\eta_1\,\mathrm{d}s_1\,\mathrm{d}\eta_2\,\mathrm{d}s_2\,\mathrm{d}x_{\beta_1}\,\mathrm{d}x'_{\beta_1}\,\mathrm{d}x_{\beta_2}\,\mathrm{d}x'_{\beta_2}\,\mathrm{d}z_1\,\mathrm{d}z'_1\,\mathrm{d}z_2\,\mathrm{d}z'_2\,. \tag{85}$$

Of these 12 partial joint probability density functions, three are dependent, since during the time of the interaction the positions of the particles are supposed to remain unchanged on account of the short collision duration (equivalently, this follows from the assumption that the two colliding particles have the same position). So, the variables $s_{1,2}$, $x_{\beta_{1,2}}$ and $z_{1,2}$ satisfy the following conditions if $D_z = 0$:

$$s_1 = s_2, \qquad x_{\beta_1} + D_x\eta_1 \equiv x_{\beta_2} + D_x\eta_2, \qquad z_1 = z_{\beta_1} \equiv z_2 = z_{\beta_2}\,. \tag{86}$$

Then $P_{12\text{var}}$ and $\mathrm{d}V_{12\text{var}}$ can be reduced to 9 variables

$$P = P_\eta(\eta_1) P_\eta(\eta_2) P_s(s_1) P_{x_\beta}(x_{\beta_1}) P_{x'_\beta}(x'_{\beta_1}) P_{x'_\beta}(x'_{\beta_2}) P_z(z_1) P_{z'}(z'_1) P_{z'}(z'_2)\,, \tag{87}$$

$$\mathrm{d}V = \mathrm{d}\eta_1\,\mathrm{d}\eta_2\,\mathrm{d}s_1\,\mathrm{d}x_{\beta_1}\,\mathrm{d}x'_{\beta_1}\,\mathrm{d}x'_{\beta_2}\,\mathrm{d}z_1\,\mathrm{d}z'_1\,\mathrm{d}z'_2\,. \tag{88}$$

We now investigate the distribution of the scattering angle $\bar{\psi}$ resulting from the Coulomb interaction of two non-relativistic ions (with $\bar{\beta} \ll 1$) of charge Z and atomic mass A in the CM frame, for which we can use the 'classical' *Rutherford differential cross-section* formula:

$$\bar{\sigma}(\bar{\psi}) = \left(\frac{AmZ^2e^2}{4\pi\epsilon_0|\bar{\mathbf{p}}_2 - \bar{\mathbf{p}}_1|^2}\right)^2 \frac{1}{\sin^4[\bar{\psi}/2]} = \left(\frac{Z^2 r_0 mc^2}{2\bar{T}}\right)^2 \frac{1}{\sin^4[\bar{\psi}/2]} = \left(\frac{Z^2}{A}\frac{r_0}{4\bar{\beta}^2}\right)^2 \frac{1}{\sin^4[\bar{\psi}/2]}\,, \tag{89}$$

where $\bar{T}=|\bar{\mathbf{p}}_2-\bar{\mathbf{p}}_1|^2/2Am=2Am\bar{\beta}^2c^2$ is the kinetic energy of the ion, $2Am\bar{\beta}c$ is the relative momentum of the two colliding ions, with $\bar{\mathbf{p}}_1+\bar{\mathbf{p}}_2=0$ in the CM frame as the collision is elastic, and

$$r_0=\frac{e^2}{4\pi\epsilon_0 mc^2}\quad\text{and}\quad r_{\rm i}=\frac{r_0Z^2}{A} \tag{90}$$

are the *classical proton radius* and *classical ion radius*, respectively.

With the assumptions $p_{s_{1,2}}\approx p_{1,2}$, $(p_1-p_2)/p\ll 1$, $p_{x_{1,2}}/p=x'_{1,2}\ll 1$ and $p_{z_{1,2}}/p=z'_{1,2}\ll 1$ (see Eq. (75) in which $2\alpha\equiv\sqrt{\theta^2+\zeta^2}$), Eq. (76) can be rewritten in the following way to show the connection between the particle velocities in the CM and LAB systems.

$$\bar{\mathbf{p}}_{1,2}=m\bar{\gamma}_{1,2}\bar{\beta}_{1,2}c\approx\pm\frac{p}{2}\left[\xi\hat{\bar{\mathbf{u}}}+0\hat{\bar{\mathbf{v}}}+2\alpha\hat{\bar{\mathbf{w}}}\right]=\pm\frac{m\gamma\beta c}{2}\left[\xi\hat{\bar{\mathbf{u}}}+0\hat{\bar{\mathbf{v}}}+2\alpha\hat{\bar{\mathbf{w}}}\right]=\pm\frac{m\gamma\beta c}{2}\left[\xi\hat{\bar{\mathbf{u}}}+0\hat{\bar{\mathbf{v}}}+\sqrt{\theta^2+\zeta^2}\hat{\bar{\mathbf{w}}}\right]$$
$$\Longrightarrow\quad|\bar{\mathbf{p}}_{1,2}|^2\approx(m\bar{\gamma}_{1,2}\bar{\beta}_{1,2}c)^2\approx\frac{m^2\gamma^2\beta^2c}{4}(\alpha^2+\theta^2+\zeta^2)\,. \tag{91}$$

Moreover, assuming non-relativistic particle velocities in the CM frame, i.e. $\bar{\beta}\ll 1$, $\bar{\gamma}\approx 1$, then with (91)

$$\bar{\beta}\approx\frac{\beta\gamma}{2}\sqrt{\xi^2+\alpha^2}=\frac{\beta\gamma}{2}\sqrt{\xi^2+\theta^2+\zeta^2}=\frac{\beta\gamma}{2}\sqrt{\left(\frac{p_1-p_2}{\gamma p}\right)^2+\left(x'_1-x'_2\right)^2+\left(z'_1-z'_2\right)^2}\,. \tag{92}$$

The integrations required to calculate Eq. (84) can be done as follows, where the next integral I_{x_1}, needed to evaluate part of the mean time-derivative of $\langle\varepsilon_{x_1}\rangle/\beta_x$ is computed by replacing $\delta p/p$ and $\delta p_x/p$ with their expressions in terms of the parameters $\alpha,\xi,\theta,\bar{\phi}$ and $\bar{\psi}$ (77). Integrating $\delta\varepsilon_{x_1}/\beta_x$ over the azimuthal and scattering angle $\bar{\psi}$ and $\bar{\phi}$ using the *Mathematica* program gives, upon expanding the scattering integrals into first order series development in $\bar{\psi}_{\rm min}$, using the Eqs. (94)–(95) below to approximate $\bar{\psi}_{\rm min}/2\approx r_{\rm i}/(2\bar{\beta}^2\bar{b}_{\rm max})$, we find that

$$\begin{aligned}
I_{x_1}&\equiv\int_{\bar{\psi}_{\rm min}}^{\pi}\mathrm{d}\bar{\psi}\int_0^{2\pi}\mathrm{d}\bar{\phi}\,\bar{\sigma}(\bar{\psi})\frac{\delta\varepsilon_{x_1}}{\beta_x}\sin\bar{\psi}\\
&=-\frac{\pi r_{\rm i}^2}{8\bar{\beta}^4}\left\{\xi^2+\zeta^2-2\theta^2+\frac{D_x^2+\tilde{D}_x^2}{\beta_x^2}\gamma^2\left(\zeta^2+\theta^2-2\xi^2\right)+\frac{6\tilde{D}_x}{\beta_x}\gamma\theta\xi\right\}\\
&+\frac{\pi r_{\rm i}^2}{4\bar{\beta}^4}\left\{\frac{4x_{\beta_1}}{\beta_x}\left(\gamma_xD_x\gamma\xi+\alpha_x(D'_x\gamma\xi-\theta)\right)+4x'_{\beta_1}\left(\frac{\tilde{D}_x\gamma\xi}{\beta_x}-\theta\right)+\xi^2+\zeta^2\right.\\
&\qquad\left.+\frac{D_x^2+\tilde{D}_x^2}{\beta_x^2}\gamma^2\left(\zeta^2+\theta^2\right)+\frac{2\tilde{D}_x}{\beta_x}\gamma\xi\theta\right\}\ln\left[\frac{2}{\bar{\psi}_{\rm min}}\right].
\end{aligned} \tag{93}$$

Notice that the smallest scattering angle $\bar{\psi}_{\rm min}$ is defined by the maximum *impact parameter* $\bar{b}_{\rm max}$, as shown in Eqs. (36) and (37) for a classical Coulomb scattering process,

$$\tan\left[\frac{\bar{\psi}_{\rm min}}{2}\right]\approx\frac{r_{\rm i}}{2\bar{\beta}^2\bar{b}_{\rm max}}\,. \tag{94}$$

Also, the maximum impact parameter $\bar{b}_{\rm max}$ gives a cut-off angle for the scattering angle ψ, it is often defined as the half the beam diameter or beam height $2\sigma_z$ since $D_z=0$.

Comment: From Eq. (36) we write, for $A=Z=1$ and with Eq. (90), $b_0=e^2/2\pi\epsilon_0 m\,|\bar{\mathbf{v}}_1-\bar{\mathbf{v}}|^2=e^2/2\pi\epsilon_0 m|2\bar{\mathbf{v}}|^2=e^2/4\pi\epsilon_0 2m\bar{\beta}^2c^2=r_0/2\bar{\beta}^2$, supposing that $\bar{\mathbf{v}}_1=-\bar{\mathbf{v}}$. Hence, by means of Eq. (37), we get $\tan[\bar{\psi}_{\rm min}/2]=r_0/2\bar{\beta}^2\bar{b}_{\rm max}$.

To obtain manageable results, we assume that $\bar{\psi}_{\text{min}} \ll 1$ (i.e. $\tan[\bar{\psi}_{\text{min}}/2] \ll 1$), so that

$$2\bar{\beta}^2\bar{b}_{\text{max}}/r_i \gg 1\,. \tag{95}$$

The two brackets in (93) have comparable small values as the angles ξ, θ and $\zeta \ll 1$; however the first bracket is negligible compared to the second one because it is multiplied by the *Coulomb logarithm* (97) (with usual values between 10 and 20).

$$\begin{aligned} I_{x_1} &\equiv \int_{\bar{\psi}_{\text{min}}}^{\pi} \mathrm{d}\bar{\psi} \int_0^{2\pi} \mathrm{d}\bar{\phi}\, \bar{\sigma}(\bar{\psi}) \frac{\delta\varepsilon_{x_1}}{\beta_x} \sin\bar{\psi} \\ &= \frac{\pi r_{\text{i}}^2}{4\bar{\beta}^4} \Bigg\{ \frac{4x_{\beta_1}}{\beta_x}\big(\gamma_x D_x \gamma\xi + \alpha_x(D'_x\gamma\xi - \theta)\big) + 4x'_{\beta_1}\left(\frac{\tilde{D}_x\gamma\xi}{\beta_x} - \theta\right) + \xi^2 + \zeta^2 \\ &\qquad + \frac{D_x^2 + \tilde{D}_x^2}{\beta_x^2}\gamma^2\big(\zeta^2 + \theta^2\big) + \frac{2\tilde{D}_x}{\beta_x}\gamma\xi\theta \Bigg\} \ln\left[\frac{2\bar{\beta}^2\bar{b}_{\text{max}}}{r_{\text{i}}}\right]. \end{aligned} \tag{96}$$

The logarithm factor in (96) is the so-called Coulomb logarithm $\bar{C}_{\log}$ defined in the CM system (see e.g. [2, 23]) as

$$\bar{C}_{\log} \equiv \ln\left[\frac{2\bar{\beta}^2\bar{b}_{\text{max}}}{r_{\text{i}}}\right] = \ln\left[\frac{2}{\bar{\psi}_{\text{min}}}\right]\,. \tag{97}$$

Comment: Observe that $C_{\log} \neq \bar{C}_{\log}$ because β in the rest frame is not equal to $\bar{\beta}$ in the CM frame as shown in Eq. (92).

Alternative definitions of the Coulomb logarithm are proposed, for example in [18]. However, its logarithmic dependence means that it changes slowly over a large range of the elements involved in its definition. In summary:

The other integrals I_{z_1} and I_{s_1} for the vertical and longitudinal momenta can be worked out in a similar way, assuming no vertical dispersion ($D_z = D'_z = 0$, with $\alpha_z \neq 0$). Combined, they yield the transverse and longitudinal scattering integrals (in which δH in Eq. 83 is now rewritten as $\delta H \approx 2\eta\delta p_s/p + (\delta p_s/p)^2$ since $\delta p \approx \delta p_s$, where $\eta = \Delta p/p$)

$$\begin{pmatrix} I_{s_1} \\ I_{x_1} \\ I_{z_1} \end{pmatrix} \equiv \int_{\bar{\psi}_{\text{min}}}^{\pi} \mathrm{d}\bar{\psi} \int_0^{2\pi} \mathrm{d}\bar{\phi}\, \sin\bar{\psi}\, \bar{\sigma}(\bar{\psi}) \begin{Bmatrix} \delta H_1/\gamma^2 \\ \delta\varepsilon_{x_1}/\beta_x \\ \delta\varepsilon_{z_1}/\beta_z \end{Bmatrix} = \frac{\pi r_{\text{i}}^2}{4\bar{\beta}^4} \ln\left[\frac{2\bar{\beta}^2\bar{b}_{\text{max}}}{r_{\text{i}}}\right] \tag{98}$$

$$\times \begin{Bmatrix} -\dfrac{4\eta_1}{\gamma}\xi + \theta^2 + \zeta^2 \\ \dfrac{4x_{\beta_1}}{\beta_x}\big(\gamma_x D_x\gamma\xi + \alpha_x(D'_x\gamma\xi - \theta)\big) + 4x'_{\beta_1}\Big(\frac{\tilde{D}_x\gamma\xi}{\beta_x} - \theta\Big) + \xi^2 + \zeta^2 + \frac{D_x^2+\tilde{D}_x^2}{\beta_x^2}\gamma^2\big(\zeta^2+\theta^2\big) + \dfrac{2\tilde{D}_x}{\beta_x}\gamma\xi\theta \\ -\dfrac{4\alpha_z z_1}{\beta_z}\zeta - 4z'_1\zeta + \xi^2 + \theta^2 \end{Bmatrix}.$$

In the centre of mass system, the derivatives $\mathrm{d}/\mathrm{d}s$ are reduced by the factor γ due to the Lorentz contraction along the longitudinal direction s, the transverse beam sizes and the relative momentum spread remain unchanged, the impact parameter $\bar{b}_{\text{max}} = b_{\text{max}}$ since it is perpendicular to the s-axis, the bunch length becomes $\bar{\sigma}_s = \gamma\sigma_s$

$$\begin{aligned} \bar{P} &= P/\gamma\,, \quad \mathrm{d}\bar{\mathrm{t}} = \mathrm{d}\mathrm{t}/\gamma \quad \text{and} \quad \bar{\sigma}_{x'_\beta} = \sigma_{x'_\beta}/\gamma \quad \bar{\sigma}'_z = \sigma'_z/\gamma\,, \\ \bar{\sigma}_{x_\beta} &= \sigma_{x_\beta}\,, \quad \bar{\sigma}_z = \sigma_z \quad \text{and} \quad \bar{\sigma}_{x'_\beta} = \sigma_{x'_\beta}/\gamma \quad \bar{\sigma}'_z = \sigma'_z/\gamma\,, \end{aligned} \tag{99}$$

$$\bar{\sigma}_\eta \;= \sigma_\eta \;, \qquad \bar{\sigma}_s \;= \gamma\sigma_s.$$

In accordance with Piwinski [6] - [20] the relative velocity between two colliding particles in the centre of mass system is $2\bar{\beta}c$. Thus, using the Rutherford differential cross-section $\bar{\sigma}(\bar{\psi})$ the likelihood of a collision per unit time and solid angle element $\mathrm{d}\bar{\Omega}(\bar{\psi},\bar{\phi})$, denoted $\bar{P}_{\text{scat}}$ is determined by the particle density distribution in phase space $\bar{P}$. We can write

$$\bar{P}_{\text{scat}} = 2\bar{\beta}c\,\bar{P}\,\bar{\sigma}(\bar{\psi}) \quad \text{with} \quad \bar{\sigma}(\bar{\psi})\mathrm{d}\bar{\Omega} = \left(\frac{r_{\text{i}}}{4\bar{\beta}^2\sin^2[\bar{\psi}/2]}\right)^2 \sin\bar{\psi}\mathrm{d}\bar{\psi}\mathrm{d}\bar{\phi}\;. \tag{100}$$

The phase space density distribution $\bar{P}$ is given by P/γ, P being defined in LAB frame; also, the transformation of the time step $\mathrm{d}\bar{\mathrm{t}} = \mathrm{d}t/\gamma$ in LAB frame produces another factor γ. It follows that Eq. (84) for the change per unit time of the mean values $\langle H\rangle/\gamma^2$ and $\langle\varepsilon_{x,z}\rangle/\beta_{x,z}$, averaged over all particles, integrating P over the phase space volume element dV introduced in Eqs. (87)–(88) can be cast into the form, with Eq. (98):

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\langle H_1\rangle/\gamma^2\\ \langle\varepsilon_{x_1}\rangle/\beta_x\\ \langle\varepsilon_{z_1}\rangle/\beta_z\end{bmatrix}\right\rangle = \int_V \frac{2\bar{\beta}cP}{\gamma^2}\begin{pmatrix}I_{s_1}\\ I_{x_1}\\ I_{z_1}\end{pmatrix}\mathrm{d}V\;. \tag{101}$$

Equation (101) is stated in the LAB frame except for β. It will be fully converted back to the laboratory system, with $\bar{\beta}$ replaced by its approximation $\beta\gamma\sqrt{\xi^2+\theta^2+\zeta^2}/2$ (Eq. (92)) after a suitable change of variables in P.

Step* 6: *Averaging over the particle momenta and positions

Computation of the mean change of the invariants $\varepsilon_{x,z}$ and H of all particles due to the multiple particle collisions requires averaging the above three integrals for the colliding particles over the joint density distribution P, where the 12 variables are reduced to nine $(\eta, s, \xi, x_\beta, x'_\beta, \theta, z, z', \zeta)$, as three of them are dependent; see Eq. (86). The mapping to the following new variables transformation P into $\mathcal{P}$ is

$$P(\eta_1,\eta_2,s_1,x_{\beta_1},x'_{\beta_1},x'_{\beta_2},z_1,z'_1,z'_2) \;\longmapsto\; \mathcal{P}(\eta,\xi,s,x_\beta,x'_\beta,\theta,z,z',\zeta)\;, \tag{102}$$

where the three angles ξ, θ and ζ have been introduced. Let us make the variable substitution (in conformity with Eq. (75)), taking into account that $D'_x \neq 0$:

$$\begin{aligned} x_{\beta_{1,2}} &= x_\beta \mp \frac{D_x\gamma\xi}{2}\;, & \eta_{1,2} &= \eta \pm \frac{\gamma\xi}{2}\;,\\ x'_{\beta_{1,2}} &= x'_\beta \pm \frac{\theta - D'_x\gamma\xi}{2}\;, & z'_{1,2} &= z' \pm \frac{\zeta}{2}\;,\\ x_{1,2} &= x\;, & z_{1,2} &= z\;, \qquad s_{1,2}= s. \end{aligned} \tag{103}$$

Hence, the phase space volume element $\mathrm{d}V$ (88) can be expressed in terms of these new variables via the 9×9 Jacobian matrix J of the transformation (103).

$$\begin{aligned} J &= \begin{pmatrix} \partial\eta_1/\partial\eta & \partial\eta_2/\partial\eta & \cdots & x'_{\beta_1}/\partial\eta & x'_{\beta_2}/\partial\eta & \cdots & z'_{\beta_1}/\partial\eta & z'_{\beta_2}/\partial\eta\\ \partial\eta_1/\partial\xi & \partial\eta_2/\partial\xi & \cdots & x'_{\beta_1}/\partial\xi & x'_{\beta_2}/\partial\xi & \cdots & z'_{\beta_1}/\partial\xi & z'_{\beta_2}/\partial\xi\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & & \\ \partial\eta_1/\partial\zeta & \partial\eta_2/\partial\zeta & \cdots & x'_{\beta_1}/\partial\zeta & x'_{\beta_2}/\partial\zeta & \cdots & z'_{\beta_1}/\partial\zeta & z'_{\beta_2}/\partial\zeta \end{pmatrix}\\ &= \begin{pmatrix} 1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0\\ \gamma/2 & -\gamma/2 & \cdots & -D'_x\gamma/2 & D'_x\gamma/2 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1/2 & -1/2 \end{pmatrix}, \end{aligned} \tag{104}$$

The absolute value of the determinant of J is easily found to be $|\det \mathrm{J}| = \gamma$, the density distribution and the phase space volume element in the new variables being labelled $\mathcal{P}$ (102) and $\mathrm{d}\mathcal{V}$. The relation between the new and the initial phase volume elements is related to the transformation of multiple integrals by

$$\int_V P \,\mathrm{d}V = \int_{\mathcal{V}} |\det \mathrm{J}\,| \mathcal{P} \,\mathrm{d}\mathcal{V} \tag{105}$$

with $\mathrm{d}\mathcal{V} = \mathrm{d}\eta \,\mathrm{d}\xi \,\mathrm{d}s \,\mathrm{d}x_\beta \,\mathrm{d}x'_\beta \,\mathrm{d}\theta \,\mathrm{d}z \,\mathrm{d}z' \,\mathrm{d}\zeta$ and $\mathrm{d}V = \mathrm{d}\eta_1 \,\mathrm{d}\eta_2 \,\mathrm{d}s_1 \,\mathrm{d}x_{\beta_1} \,\mathrm{d}x'_{\beta_1} \,\mathrm{d}x'_{\beta_2} \,\mathrm{d}z_1 \,\mathrm{d}z'_1 \,\mathrm{d}z'_2$.

Hence, the change per unit time of the three averaged invariants (101) can be rewritten using (105), taking γ for $|\det \mathrm{J}|$

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \langle H\rangle/\gamma^2 \\ \langle \varepsilon_x\rangle/\beta_x \\ \langle \varepsilon_z\rangle/\beta_z \end{bmatrix} \right\rangle = \int_{\mathcal{V}} \frac{2\bar{\beta} c \mathcal{P}}{\gamma} \begin{pmatrix} I_{s_1} \\ I_{x_1} \\ I_{z_1} \end{pmatrix} \mathrm{d}\mathcal{V} \,. \tag{106}$$

Let us consider I_{x_1} in Eq. (98) and replace the variables x_{β_1} and x'_{β_1} in the bracket by the relevant new variables (103) into Eq. (106). We obtain

$$\begin{aligned} \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \frac{\langle \varepsilon_x\rangle}{\beta_x} \right\rangle &= \frac{\pi c r_\mathrm{i}^2}{2} \int_{\mathcal{V}} \frac{\mathrm{d}\mathcal{V}}{\bar{\beta}^3 \gamma} \mathcal{P}\big(\eta, s, \xi, x_\beta, x'_\beta, \theta, z, z', \zeta\big) \ln\left[\frac{2\bar{\beta}^2 \bar{b}_{\max}}{r_\mathrm{i}}\right] \\ &\times \left\{ \xi^2 + \zeta^2 - 2\theta^2 + \frac{D_x^2 + \tilde{D}_x^2}{\beta_x^2} \gamma^2 \big(\zeta^2 + \theta^2\big) - \frac{2\gamma_x D_x^2}{\beta_x} \gamma^2 \xi^2 - \frac{2D'_x}{\beta_x} \big(\alpha_x D_x + \tilde{D}_x\big) \gamma^2 \xi^2 + \frac{4\tilde{D}_x}{\beta_x} \gamma (x'_\beta + \theta) \xi \right. \\ &\left. - \frac{4\gamma}{\beta_x} (\alpha_x x_\beta + \beta_x x'_\beta) \theta + \frac{2D_x}{\beta_x} \gamma (2\gamma_x x_\beta + \alpha_x \theta) \xi + \frac{4D'_x}{\beta_x} \gamma \alpha_x x_\beta \xi + 2D'_x \gamma \theta \xi \right\} . \end{aligned} \tag{107}$$

By construction the probability density law $\mathcal{P}$ is clearly symmetrical with respect to the angles ξ, θ and ζ, so the integrals vanish for the linear terms in ξ, θ and ζ of the integrands. Consequently, keeping only the factors ξ^2, θ^2 and ζ^2, Eq. (107) reduces to the expression to

$$\begin{aligned} \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \frac{\langle \varepsilon_x\rangle}{\beta_x} \right\rangle &= \frac{\pi c r_\mathrm{i}^2}{2} \int_{\mathcal{V}} \frac{\mathrm{d}\mathcal{V}}{\bar{\beta}^3 \gamma} \mathcal{P}\big(\eta, s, \xi, x_\beta, x'_\beta, \theta, z, z', \zeta\big) \ln\left[\frac{2\bar{\beta}^2 \bar{b}_{\max}}{r_\mathrm{i}}\right] \\ &\times \left\{ \xi^2 + \zeta^2 - 2\theta^2 + \frac{D_x^2 + \tilde{D}_x^2}{\beta_x^2} \gamma^2 \big(\zeta^2 + \theta^2\big) - \frac{2\gamma_x D_x^2}{\beta_x} \gamma^2 \xi^2 - \frac{2D'_x}{\beta_x} (\alpha_x D_x + \tilde{D}_x) \gamma^2 \xi^2 \right\} . \end{aligned} \tag{108}$$

Comment: Using Eq. (102), the symmetry of $\mathcal{P}$ means that $\mathcal{P}(\eta, -\xi, s, x_\beta, x'_\beta, -\theta, z, z', -\zeta) \longmapsto P(\eta_2, \eta_1, s_1, x_{\beta_1}, x'_{\beta_2}, x'_{\beta_1}, z_1, z'_2, z'_1)$ where the order of the variables $\{\eta_1, \eta_2\}$, $\{x'_{\beta_1}, x'_{\beta_2}\}$ and $\{z'_1, z'_2\}$ in P permutes without changing the probability law.

Proceeding similarly with I_{s_1} and I_{z_1} in (98) and by inserting the upgrade terms into Eq. (107) yields:

$$\begin{aligned} \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \langle H\rangle/\gamma^2 \\ \langle \varepsilon_x\rangle/\beta_x \\ \langle \varepsilon_z\rangle/\beta_z \end{bmatrix} \right\rangle &= \frac{\pi c r_\mathrm{i}^2}{2} \int_{\mathcal{V}} \frac{\mathrm{d}\mathcal{V}}{\bar{\beta}^3 \gamma} \mathcal{P}\big(\eta, s, \xi, x_\beta, x'_\beta, \theta, z, z', \zeta\big) \ln\left[\frac{2\bar{\beta}^2 \bar{b}_{\max}}{r_\mathrm{i}}\right] \\ &\times \left\{ \begin{array}{c} \theta^2 + \zeta^2 - 2\xi^2 \\ \xi^2 + \zeta^2 - 2\theta^2 + \dfrac{D_x^2 + \tilde{D}_x^2}{\beta_x^2} \gamma^2 \big(\zeta^2 + \theta^2\big) - \dfrac{2\gamma_x D_x^2}{\beta_x} \gamma^2 \xi^2 - \dfrac{2D'_x}{\beta_x} (\alpha_x D_x + \tilde{D}_x) \gamma^2 \xi^2 \\ \xi^2 + \theta^2 - 2\zeta^2 \end{array} \right\} . \end{aligned} \tag{109}$$

This formula for the mean change of the invariants $\varepsilon_{x,z}$ and H makes no a priori assumption about the density distribution P ($\mathcal{P}$) of the particles within the bunch. So in principle the integral can be calculated for arbitrary distribution functions. However, since 'Gaussian integration' is quite easily

performed, many analytical IBS models are based on the assumption that all the betatron amplitudes and angles, as well as the momentum deviations and the coordinates of the synchrotron motion (for bunched beams), follow Gaussian distributions.

Consequently, at this stage we assume that all the variables follow Gaussian probability laws. We then introduce bi-Gaussian formulations of the betatron amplitude and angle distributions, $P_{x_\beta x'_\beta}(x_{\beta_{1,2}}, x'_{\beta_{1,2}})$ and $P_{zz'}(z_{1,2}, z'_{1,2})$, along with the momentum and bunch-length deviation distributions $P_{\eta s}(\eta_{1,2}, s_{1,2})$. The density distributions, in which $u = x, z$ stands for both horizontal and vertical betatron motions, can be expressed as (see [24])

$$P_{u_\beta u'_\beta}(u_\beta, u'_\beta) = \frac{\sqrt{1+\alpha_u^2}}{2\sigma_{u_\beta}\sigma'_{u_\beta}} \exp\left[-Q\left(u_\beta, u'_\beta\right)\right] , \tag{110}$$

$$Q\left(u_\beta, u'_\beta\right) = \frac{1+\alpha_u^2}{2}\left[\frac{u_\beta^2}{\sigma_{u_\beta}^2} + \frac{2u_\beta u'_\beta \alpha_u}{\sigma_{u_\beta}\sigma_{u'_\beta}\sqrt{1+\alpha_u^2}} + \frac{u'^2_\beta}{\sigma_{u'_\beta}^2}\right] ,$$

$$P_{\eta s}(\eta, s) = P_\eta(\eta) P_s(s) = \frac{1}{2\pi\sigma_\eta\sigma_s} \exp\left[-\frac{\eta^2}{2\sigma_\eta^2} - \frac{(s-s_0)^2}{2\sigma_s^2}\right] . \tag{111}$$

In standard Gaussian notation, σ_{x_β} and $\sigma_{x'_\beta}$ are the r.m.s. beam size $\sqrt{\langle x^2\rangle}$ and angular spread $\sqrt{\langle x'^2\rangle}$, respectively, and similarly for the vertical betatron motion z, for which $z = z_\beta$ and $z' = z'_\beta$ by Eq. (78). Likewise σ_η is the momentum spread $\sqrt{\langle\eta^2\rangle}$, and σ_s is the r.m.s. bunch length $\sqrt{\langle s^2\rangle}$, with $\Delta s = s - s_0$ being the synchrotron coordinates, i.e. the position relative to the synchronous particle. The quadratic form $Q =$ constant is a tilted ellipse with correlation coefficient $\rho_x = \alpha_x/\sqrt{1+\alpha_x^2}$. The above probability distributions P must be well-matched to the Courant–Snyder invariant $\varepsilon_x = \gamma_x x_\beta^2 + 2\alpha_x x_\beta x'_\beta + \beta_x x'^2_\beta$. Using the related betatron amplitude and angle r.m.s. values σ_{x_β} and $\sigma_{x'_\beta} = \sigma_{x_\beta}\sqrt{\gamma_x/\beta_x}$, the probability $P_{x_\beta x'_\beta}$ can be rewritten as

$$P_{x_\beta x'_\beta}(x_\beta, x'_\beta) = \frac{\beta_x}{2\sigma_{x_\beta}^2} \exp\left[-\frac{\beta_x}{2\sigma_{x_\beta}^2}\left(\gamma_x x_\beta^2 + 2\alpha_x x_\beta x'_\beta + \beta_x x'^2_\beta\right)\right] . \tag{112}$$

or using the emittance ε_x in place of the variables position and angle x_β, x'_β

$$P_{\varepsilon_x}(\varepsilon_x) = \frac{\beta_x}{2\sigma_{x_\beta}^2} \exp\left[-\frac{\beta_x\varepsilon_x}{2\sigma_{x_\beta}^2}\right] \iff P_{\varepsilon_x}(\varepsilon_x) = \frac{1}{\langle\varepsilon_x\rangle} \exp\left[-\frac{\varepsilon_x}{\langle\varepsilon_x\rangle}\right] . \tag{113}$$

Here, the emittance describes the phase space area used by the beam; that is, for a phase space area covering a fraction F_{ε_x} of a Gaussian beam with r.m.s. value σ_{x_β}, the emittance at F_{ε_x} % of particles is

$$\varepsilon_x = -\frac{2\sigma_{x_\beta}^2}{\beta_x}\ln(1-F_{\varepsilon_x}) \quad \text{or} \quad F_{\varepsilon_x}(\varepsilon_x) = 1 - \exp\left[-\frac{\beta_x\varepsilon_x}{2\sigma_{x_\beta}^2}\right] \quad \text{with} \quad 0 < F_{\varepsilon_x} \le 1 . \tag{114}$$

$F_{\varepsilon_x}(\varepsilon_x)$ is the cumulative probability function and its derivative is the probability $P_{\varepsilon_x}(\varepsilon_x) = \mathrm{d}F/\mathrm{d}t$ (113). The second moment of $P_\eta(\eta)$ and the mean value of $P_{\varepsilon_x}(\varepsilon_x)$ are:

$$\langle\eta^2\rangle = \int_{-\infty}^{\infty} \eta^2 P_\eta(\eta)\mathrm{d}\eta = \sigma_\eta^2 \ \text{ and } \ \langle\varepsilon_x\rangle = \int_0^\infty \varepsilon_x P_{\varepsilon_x}(\varepsilon_x)\mathrm{d}\varepsilon_x = \frac{2\sigma_{x_\beta}^2}{\beta_x} \implies$$

$$\langle x_\beta^2\rangle \equiv \sigma_{x_\beta}^2 = \frac{\beta_x\langle\varepsilon_x\rangle}{2} \ \text{ and } \ \langle x'^2_\beta\rangle \equiv \sigma_{x'_\beta}^2 = \frac{\gamma_x\langle\varepsilon_x\rangle}{2} \implies \frac{\sigma_{x'_\beta}}{\sigma_{x_\beta}} = \sqrt{\frac{\gamma_x}{\beta_x}} = \frac{1}{\beta_x} \ \text{ if } \ \alpha_x = 0. \tag{115}$$

The last two formulae in (115) are obtained looking at Fig. 11, which shows beam boundaries at $\pm\sqrt{6}\sigma_{x_\beta}$ and $\pm\sqrt{6}\sigma_{x'_\beta}$. In analogy, the boundaries for $\langle\varepsilon_x\rangle$ would be at $\pm\sigma_{x_\beta} = \pm\sqrt{\beta_x\langle\varepsilon_x\rangle/2}$ and $\pm\sigma_{x'_\beta} = \pm\sqrt{\gamma_x\langle\varepsilon_x\rangle/2}$, yielding $\sigma_{x'_\beta}^2 = \gamma_x\langle\varepsilon_x\rangle/2$.

Figure 11 depicts the area occupied by the distribution in phase space, which parameterizes an elliptical contour surrounding a given fraction of the beam. Thus, a larger ellipse containing, say, 95% of the particles in the bunch has envelope boundaries at $\pm\sqrt{6}\sigma_{x_\beta}=\pm\sqrt{\beta_x\varepsilon_x}$ and $\pm\sqrt{6}\sigma_{x'_\beta}=\pm\sqrt{\gamma_x\varepsilon_x}$. For instance, the emittances at $F_{\varepsilon_x}=(39,\,87,\,95)\%$ of particles in phase space are $\varepsilon_x=(1,\,4,\,6)\,\sigma^2_{x_\beta}/\beta_x$. The emittance at 39% is the r.m.s. emittance, labelled $\varepsilon_{x_{\rm rms}}$.

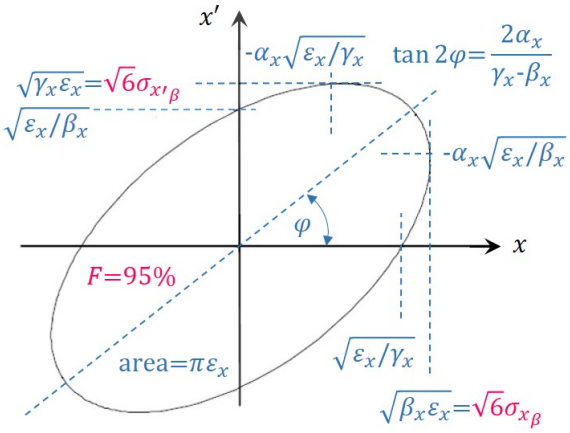


Fig. 11: Extrema, intercepts and slope of the phase space ellipse enclosing $F_{x_\beta x'_\beta}$ % of particles characterized by the Twiss parameters. The area bounded by the ellipse is equal to $\pi\varepsilon_x$.

Comment: Besides ε_x, which bounds a fraction F_{ε_x} of particles in phase space, an alternative well known emittance concept is defined as $\varepsilon_x^{\rm proj}$ ('projected emittance'), whose beam width contains a fraction $F_{\varepsilon_x}^{\rm proj}$ of particles projected onto a betatron amplitude axis. Unlike F_{ε_x} for ε_x, the beam fraction $F_{\varepsilon_x}^{\rm proj}$, for instance $\varepsilon_x^{\rm proj}=(1,\,4,\,6)\,\sigma^2_{x_\beta}/\beta_x$ is equal to $F_{\varepsilon_x}^{\rm proj}=(68,\,95,\,99)\%$. Actually, emittance measurements using beam profile monitors, like wire-scanners and gas ionization monitors, measure projected transverse beam distributions through 'integration' over the angles x', from whence σ_x and $\varepsilon_x^{\rm proj}$ are derived. On the other end, beam destructive type devices, such as scrapers, ongoing decrease of beam intensity by moving blades into the beam. This process delivers the transverse beam size versus the fraction of remaining intensity and, by differentiation, the betatron beam amplitude distribution. Therefore, emittance ε_x are by nature of the process expressed in term of the enclosed fraction of particles.

Summing up for the r.m.s. emittances derived from phase space and projected beam densities one gets (see also (115))

$$F_{\varepsilon_{x_{\rm rms}}}=F_{\varepsilon_{x_{\rm rms}}}\left(\varepsilon_{x_{\rm rms}}=\sigma^2_{x_\beta}/\beta_x\right)=1-\exp\left[-\frac{1}{2}\right]=0.39\quad\text{with}\quad\langle\varepsilon_{x_{\rm rms}}\rangle=\frac{2\sigma^2_{x_\beta}}{\beta_x}\neq\varepsilon_{x_{\rm rms}}=\frac{\sigma^2_{x_\beta}}{\beta_x}\,,$$

$$F^{\rm proj}_{\varepsilon_{x_{\rm rms}}}=F^{\rm proj}_{\varepsilon_{x_{\rm rms}}}\left(\varepsilon^{\rm proj}_{x_{\rm rms}}=\frac{\sigma^2_{x_\beta}}{\beta_x}\right)=\int_{-\sigma_{x_\beta}}^{\sigma_{x_\beta}}P^{\rm proj}_{x_\beta}(x_\beta)\mathrm{d}x_\beta=0.68\quad\text{where}\quad P^{\rm proj}_{x_\beta}(x_\beta)=\frac{1}{\sqrt{2\pi}\,\sigma_{x_\beta}}\exp\left[-\frac{x^2_\beta}{2\sigma^2_{x_\beta}}\right],$$

$$\langle x^2_\beta\rangle=\int_{-\infty}^{\infty}x^2_\beta P^{\rm proj}_{x_\beta}(x_\beta)\mathrm{d}x_\beta=\sigma^2_{x_\beta}\quad\Longrightarrow\quad\langle\varepsilon^{\rm proj}_{x_{\rm rms}}\rangle=\frac{\langle x^2_\beta\rangle}{\beta_x}=\frac{\sigma^2_{x_\beta}}{\beta_x}\equiv\varepsilon^{\rm proj}_{x_{\rm rms}}\neq\langle\varepsilon_{x_{\rm rms}}\rangle\,.\qquad(116)$$

Figure 12 shows the reference orbit of a synchrotron accelerator or storage ring system, referred to as the rest frame. The coordinate system attached to the rotating vector $\mathbf{r}_0(s)$ is referred to as the centre-of-mass frame, even though it is not an inertial system moving at constant velocity along a straight line.

To simplify and clarify the formalism, from now on we neglect the derivatives of the dispersion and transverse betatron functions (the vertical dispersion was earlier assumed to be null) that is:

$$D'_{x,z}=0,\quad\beta'_{x,z}=-2\alpha_{x,z}=0\quad\Longrightarrow\quad\tilde{D}_{x,z}=\alpha_{x,z}D_{x,z}+\beta_{x,z}D'_{x,z}=0,\quad\gamma_{x,z}=1/\beta_{x,z}\,.\qquad(117)$$

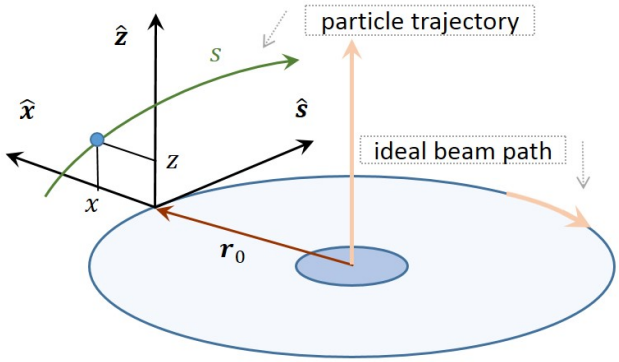


Fig. 12: Curvilinear Frenet–Serret coordinate system for particle motion in accelerator and storage rings: $\mathbf{r}(s) = \mathbf{r}_0(s) + x(s)\hat{\mathbf{x}} + z(s)\hat{\mathbf{z}}$, with $\hat{\mathbf{s}} = \frac{\mathrm{d}\mathbf{r}_0}{\mathrm{d}s} \big/ \left|\frac{\mathrm{d}\mathbf{r}_0}{\mathrm{d}s}\right|$ and $\hat{\mathbf{z}} = \hat{\mathbf{s}} \times \hat{\mathbf{x}}$.

Using the change of variables (103), in which $x'_{\beta_{1,2}}$ reduces to $x'_\beta \pm \theta/2$ since $D'_x = 0$, the Gaussian density distributions (110) and (111) can be rewritten in terms of the nine variables $\eta, s, \xi, x_\beta, x'_\beta, \theta, z, z', \zeta$ as (swapping from P to $\mathcal{P}$)

$$\mathcal{P}_{x_\beta x'_\beta}\left(x_\beta \mp \frac{D_x\gamma\xi}{2},\, x'_\beta \pm \frac{\theta}{2}\right), \quad \mathcal{P}_{zz'}\left(z,\, z' \pm \frac{\zeta}{2}\right), \quad \mathcal{P}_\eta\left(\eta \pm \frac{\gamma\xi}{2}\right), \quad \mathcal{P}_s^2(s)\,. \tag{118}$$

Consequently the transformation mapping P into $\mathcal{P}$ (Eq. 102) of the density distributions $P_{x_\beta x'_\beta}(x_\beta, x'_\beta)$, $P_{zz'}(z, z')$ and $P_{\eta s}(\eta, s)$, Eqs. (110) and (111), which are now separable since $\alpha_{x,z} = 0$, can thus be rephrased (remembering that two particles are involved and that (78) apply) as

$$P_\eta(\eta_1, \eta_2) = P_\eta(\eta_1)P_\eta(\eta_2) \mapsto \mathcal{P}_\eta\left(\eta + \frac{\gamma\xi}{2}\right)\mathcal{P}_\eta\left(\eta - \frac{\gamma\xi}{2}\right), \tag{119}$$

$$P_s(s_1, s_2) = P_s(s_1)P_s(s_2) \mapsto \mathcal{P}_s(s)\mathcal{P}_s(s) = \mathcal{P}_s^2(s), \tag{120}$$

$$P_{x_\beta}(x_{\beta_1}, x_{\beta_2}) = P_{x_\beta}(x_{\beta_1})P_{x_\beta}(x_{\beta_2}) \mapsto \mathcal{P}_{x_\beta}\left(x_\beta - \frac{D_x\gamma\xi}{2}\right)\mathcal{P}_{x_\beta}\left(x_\beta + \frac{D_x\gamma\xi}{2}\right), \tag{121}$$

$$P_{x'_\beta}(x'_{\beta_1}, x'_{\beta_2}) = P_{x'_\beta}(x'_{\beta_1})P_{x'_\beta}(x'_{\beta_2}) \mapsto \mathcal{P}_{x'_\beta}\left(x'_\beta + \frac{\theta}{2}\right)\mathcal{P}_{x'_\beta}\left(z'_\beta - \frac{\theta}{2}\right), \tag{122}$$

$$P_z(z_1, z_2) = P_z(z_1)P_z(z_2) \mapsto \mathcal{P}_z^2(z), \tag{123}$$

$$P_{z'}(z'_1, z'_2) = P_{z'}(z'_1)P_{z'}(z'_2) \mapsto \mathcal{P}_{z'}\left(z' + \frac{\zeta}{2}\right)\mathcal{P}_{z'}\left(z'_\beta - \frac{\zeta}{2}\right)\,. \tag{124}$$

The Gaussian integrals over the six variables $\eta, s, x_\beta, x'_\beta, z, z'$ can be made by means of *Mathematica*. Some calculation details of the density distribution integrals are shown for (119) and (120), which are representative of the other four integrals. Substituting the η-exp part of (111) into (119) and integrating over the momentum deviation range $\{-\infty, \infty\}$ gives,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \mathrm{d}\eta\, P_\eta\Big(\eta + \frac{\gamma\xi}{2}\Big)\mathcal{P}_\eta\Big(\eta - \frac{\gamma\xi}{2}\Big) \\
&= \frac{1}{2\pi\sigma_\eta^2}\int_{-\infty}^{\infty} \mathrm{d}\eta\, \exp\left[-\frac{1}{2\sigma_\eta^2}\Big(\eta + \frac{\gamma\xi}{2}\Big)^2 - \frac{1}{2\sigma_\eta^2}\Big(\eta - \frac{\gamma\xi}{2}\Big)^2\right] \\
&= \frac{1}{2\pi\sigma_\eta^2}\int_{-\infty}^{\infty} \mathrm{d}\eta\, \exp\left[-\frac{1}{2\sigma_\eta^2}\left(2\eta^2 + \frac{\gamma^2\xi^2}{2}\right)\right] \\
&= \frac{1}{2\pi\sigma_\eta^2}\, \exp\left[-\frac{\gamma^2\xi^2}{4\sigma_\eta^2}\right]\int_{-\infty}^{\infty} \mathrm{d}\eta\, \exp\left[-\frac{\eta^2}{\sigma_\eta^2}\right]
\end{aligned}$$

$$= \frac{1}{2\pi\sigma_\eta^2}\exp\left[-\frac{\gamma^2\xi^2}{4\sigma_\eta^2}\right]\sqrt{\pi}\sigma_\eta = \frac{1}{2\sqrt{\pi}\sigma_\eta}\exp\left[-\frac{\gamma^2\xi^2}{4\sigma_\eta^2}\right] . \tag{125}$$

Likewise, substituting the s-exp part of (111) into (120) and integrating gives

$$\int_{-\infty}^{\infty}\mathrm{d}s\,\mathcal{P}_s^2(s) = \frac{1}{2\pi\sigma_s^2}\int_{-\infty}^{\infty}\mathrm{d}s\exp\left[-\frac{(s-s_0)^2}{2\sigma_s^2}\right]\exp\left[-\frac{(s-s_0)^2}{2\sigma_s^2}\right] = \frac{1}{2\sqrt{\pi}\sigma_s} . \tag{126}$$

Hence, the particle density distribution $\mathcal{P}$ is the product of the six terms obtained after integrating the six density distributions (right part of Eqs. (119) to (124)) with respect to the variables $\eta, s, x_\beta, x'_\beta, z, z'$, times N_b, the number of particles per bunch is

$$\begin{aligned}\mathcal{P}(\xi,\theta,\zeta) &= N_b\prod_{u=\eta,s,x_\beta}^{x'_\beta,z,z'}\int_{-\infty}^{\infty}\mathrm{d}u\,\mathcal{P}_u\Big(u\pm\frac{\gamma\lambda_u}{2}\Big)\mathcal{P}_u\Big(u\mp\frac{\gamma\lambda_u}{2}\Big)\\ &= N_b\,\frac{\exp\Big[-\frac{\gamma^2\xi^2}{4}\Big(\frac{1}{\sigma_\eta^2}+\frac{D_x^2}{\sigma_{x_\beta}^2}\Big)-\frac{\theta^2}{4\sigma_{x'_\beta}^2}-\frac{\zeta^2}{4\sigma_{z'}^2}\Big]}{64\pi^3\sigma_{x_\beta}\sigma_{x'_\beta}\sigma_z\sigma_{z'}\sigma_\eta\sigma_s} .\end{aligned} \tag{127}$$

where λ_u stands for either ξ, 0, $D_x\xi$, θ, or ζ. At present $\mathcal{P}$ is reduced to the variables ξ, θ, ζ.

To complete the transformation back to the laboratory system, $\bar\beta$ must be replaced by its LAB frame approximation $\beta\gamma\sqrt{\xi^2+\theta^2+\zeta^2}/2)$. Let us define the Piwinski IBS constant $\mathcal{A}_\mathrm{P}$ as

$$\mathcal{A}_\mathrm{P} = \frac{cr_\mathrm{i}^2N_b}{64\pi^2\beta^3\gamma^4\sigma_{x_\beta}\sigma_{x'_\beta}\sigma_z\sigma_{z'}\sigma_\eta\sigma_s} = \frac{cr_\mathrm{i}^2N_b}{64\pi^2\beta^3\gamma^4\varepsilon_x\varepsilon_z\varepsilon_s} , \quad \text{wherein} \tag{128}$$

$$\sigma_x^2=\sigma_{x_\beta}^2+D_x^2\sigma_\eta^2 , \;\; \sigma_z^2=\sigma_{z_\beta}^2 , \;\; \sigma_{x_\beta,z_\beta}^2=\beta_{x,z}\varepsilon_{x,z} , \;\; \sigma_{x'_\beta,z'_\beta}^2=\frac{\varepsilon_{x,z}}{\beta_{x,z}} , \;\; \sigma_{x_\beta,z_\beta}\sigma_{x'_\beta,z'_\beta}=\varepsilon_{x,z} , \;\; \sigma_\eta\sigma_s=\varepsilon_s .$$

Now, using (92) and (127) the full set of Eq. (109), in which the factors $D'_{x,z}$ and $\alpha_{x,z}$ have to be removed because or the additional assumption $D'_{x,z}=\beta'_{x,z}=0$, can be easily rearranged as

$$\begin{aligned}\left\langle\begin{pmatrix}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle H\rangle}{\gamma^2}\\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle\varepsilon_x\rangle}{\beta_x}\\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle\varepsilon_z\rangle}{\beta_z}\end{pmatrix}\right\rangle &= 4\mathcal{A}_\mathrm{P}\iiint_{-\infty}^{\infty}\frac{\mathrm{d}\xi\,\mathrm{d}\theta\,\mathrm{d}\zeta}{(\xi^2+\theta^2+\zeta^2)^{3/2}}\exp\left[-\frac{\gamma^2\xi^2}{4}\left(\frac{1}{\sigma_\eta^2}+\frac{D_x^2}{\sigma_{x_\beta}^2}\right)-\frac{\theta^2}{4\sigma_{x'_\beta}^2}-\frac{\zeta^2}{4\sigma_{z'_\beta}^2}\right]\\ &\times\begin{Bmatrix}\theta^2+\zeta^2-2\xi^2\\ \xi^2+\zeta^2-2\theta^2+\frac{D_x^2}{\beta_x^2}\gamma^2\big(\zeta^2+\theta^2-2\xi^2\big)\\ \xi^2+\theta^2-2\zeta^2\end{Bmatrix}\ln\left[\frac{q^2}{4}\big(\xi^2+\theta^2+\zeta^2\big)\right] .\end{aligned} \tag{129}$$

in which the parameter q is introduced for convenience

$$q=\beta\gamma\sqrt{\frac{2b_\mathrm{max}}{r_\mathrm{i}}}\approx 2\beta\gamma\sqrt{\frac{\sigma_z}{r_\mathrm{i}}} , \quad \text{with} \quad b_\mathrm{max}\overset{\mathrm{def}}{=}2\sigma_z , \quad \text{(half beam or bunch height)} . \tag{130}$$

Comment: The Coulomb logarithm C_log in the rest frame is 'hidden' inside Eq. (129) since $\ln\big[\frac{q^2}{4}\big(\xi^2+\theta^2+\zeta^2\big)\big]=C_\mathrm{log}+\ln\big[\frac{\gamma^2}{4}\big(\xi^2+\theta^2+\zeta^2\big)\big]$, with $C_\mathrm{log}=\ln\big[\frac{2\beta^2b_\mathrm{max}}{r_\mathrm{i}}\big]$; see (97).

Discussion on an invariant

In the simplest circumstances, intrabeam scattering of particles in a bunched beam may cause dilution of the beam in phase space, leading to continuous growth of the momentum spread and/or growth of one or both transverse emittances. The importance of the transition energy is briefly revisited here in the light of the invariants.

The behaviour of the beam can be described via a global invariant which can be arranged in a form close to the sum of the mean value of the change in emittance $\langle \varepsilon_{x,z} \rangle$ and the change in momentum deviation $\langle H \rangle$ over the collisions of all particles, as discussed above. Then, multiplying the top momentum deviation change term $\frac{\mathrm{d}}{\mathrm{d}t}\left[\langle H \rangle / \gamma^2\right]$ in the left part of Eq. (129) by the factor $(1-\gamma^2 D_x^2/\beta_x^2)$ and adding it to the emittance change middle and bottom terms $\frac{\mathrm{d}}{\mathrm{d}t}\left[\langle \varepsilon_{x,z} \rangle \beta_{x,z}\right]$ gives a result equal to zero:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\langle H \rangle \left(\frac{1}{\gamma^2} - \frac{D_x^2}{\beta_x^2}\right) + \frac{\langle \varepsilon_x \rangle}{\beta_x} + \frac{\langle \varepsilon_z \rangle}{\beta_z}\right] = 0 \,. \tag{131}$$

Unfortunately, the quantity acted on by the derivative operator $\mathrm{d}/\mathrm{d}t$ is not an invariant, because the ratio D_x/β_x varies along the ring lattice (i.e. it depends on the longitudinal or azimuthal coordinate s along the lattice). However, the smooth focusing approximation holds for a weak-focusing or smooth lattice. So, from the betatron and dispersion functions, the momentum compaction factor and the transition energy factor, we have

$$\langle \beta_{x,z} \rangle \equiv \frac{R}{Q_{x,z}}, \quad \langle D_x \rangle \equiv \frac{R}{Q_{x,z}^2}, \quad \alpha_p = \frac{\langle D_{x,z}(s) \rangle}{\rho(s)} = \frac{1}{\gamma_t^2} \equiv \frac{\langle D_{x,z} \rangle}{R} = Q_{x,z}^{-2} \implies \left\langle \frac{D_{x,z}}{\beta_{x,z}} \right\rangle \equiv \frac{1}{\gamma_t^2} \,.$$

(where $\langle \cdot \rangle$ denotes an average over the lattice). After straightforward integration with respect to t (with the integrand now being a constant), we get the invariant

$$\langle H \rangle \left(\frac{1}{\gamma^2} - \frac{1}{\gamma_t^2}\right) + \frac{\langle \varepsilon_x \rangle}{\beta_x} + \frac{\langle \varepsilon_z \rangle}{\beta_z} = \text{constant} \,, \tag{132}$$

where $\eta_t = \gamma_t^{-2} - \gamma^{-2}$ is the *slip factor*.

Above transition energy ($\eta_t \geq 0$)

- The coefficient of $\langle H \rangle$ in (132) is negative and the total oscillation energy can increase as long as it does not exceed other limitations. Therefore no equilibrium distribution can exist.

Below transition energy ($\eta_t < 0$)

- The sum of the three positive invariants, and hence of the three oscillation energies, is bounded (i.e. the IBS emittance growth is constrained). So the emittances are redistributed in all three phase planes, holding the whole phase space invariant; the particle density distribution P is stable, and an equilibrium exists (like the situation of gas molecules in a closed box where the focusing forces are produced by the wall of the box).

Growth rates calculation (neglecting $\alpha_{x,z}$ and $D'_{x,z}$)

In his original model of 1974 [1], Piwinski derived expressions for the variations in the transverse oscillation amplitudes (similarly the square root of the transverse emittances) and momentum spread per unit time caused by a scattering event, using the *smooth focusing approximation* in which only the mean values of the lattice functions are considered; see Eq. (62). In addition, as seen earlier, the variations of the betatron and dispersion functions are neglected and a zero vertical dispersion function along the ring lattice is assumed, namely

$$\langle \beta_x \rangle = \frac{R}{Q_x} \,, \quad \langle D_x \rangle = \frac{R}{Q_x^2} \,, \quad \beta'_{x,z} = -2\alpha_{x,z} = 0 \quad D'_{x,z} = 0 \,, \quad \tilde{D}_x = \alpha_x D_x + \beta_x D'_x = 0 \,, \quad D_z = 0 \,, \tag{133}$$

where R is the mean radius of the ring and $Q_{x,z}$ are the transverse betatron tunes.

The longitudinal, horizontal and vertical growth times $\tau_{\eta,x,z}^{-1}$ are given by

$$\frac{1}{\tau_\eta}=\frac{1}{\sigma_\eta}\frac{d\sigma_\eta}{dt}=\frac{1}{2\sigma_\eta^2}\frac{d\sigma_\eta^2}{dt}=\frac{1}{2\langle\eta^2\rangle}\frac{d\langle\eta^2\rangle}{dt}=\frac{1}{2\langle H\rangle}\frac{d\langle H\rangle}{dt}\quad\overset{\text{sol.}}{\Longrightarrow}\quad\langle H\rangle=\langle H_0\rangle\exp\Big[\frac{2t}{\tau_\eta}\Big]\,,\tag{134}$$
$$\frac{1}{\tau_x}=\frac{1}{\sigma_{x_\beta}}\frac{d\sigma_{x_\beta}}{dt}=\frac{1}{2\sigma_{x_\beta}^2}\frac{d\sigma_{x_\beta}^2}{dt}=\frac{1}{2\langle x_\beta^2\rangle}\frac{d\langle x_\beta^2\rangle}{dt}=\frac{1}{2\langle\varepsilon_x\rangle}\frac{d\langle\varepsilon_x\rangle}{dt}\quad\overset{\text{sol.}}{\Longrightarrow}\quad\langle\varepsilon_x\rangle=\langle\varepsilon_{x_0}\rangle\exp\Big[\frac{2\,t}{\tau_\eta}\Big]\,.$$

where σ_η is the r.m.s. relative momentum spread, H_0 and ε_{x_0} are the natural relative momentum spread squared and equilibrium transverse emittance in the absence of IBS.

At this place, the projected r.m.s. emittance onto the x-axis $\langle\varepsilon_{x_{\rm rms}}^{\rm proj}\rangle=\sigma_{x_\beta}^2/\beta_x$ $\big($equal to $\varepsilon_{x_{\rm rms}}^{\rm proj}\big)$ is chosen rather than the emittance defined by (116), that is $\langle\varepsilon_{x_{\rm rms}}\rangle=2\sigma_{x_\beta}^2/\beta_x$ $\big($not equal to $\varepsilon_{x_{\rm rms}}\big)$, idem for $\langle\varepsilon_{z_{\rm rms}}\rangle$, and using $\sigma_\eta^2=\langle\eta^2\rangle$ for the relative momentum deviation squared. From now on we drop the flags 'proj' and 'rms' from the emittance notation.

Comment: The IBS growth times depend of the instantaneous beam emittance and momentum deviation values. So, it is necessary to use an iterative procedure to compute the growth times and derive the evolution of the emittances and momentum deviation. The IBS growth rates are defined by (134), where τ_η and $\tau_{x,z}$ are the longitudinal and transverse IBS growth times of the relative momentum spread squared H and transverse emittances ε_{x_0,z_0}. Adding 'synchrotron radiation damping'(SRD) effect with damping times $\tau_{\eta_{\rm srd}}$, $\tau_{x_{\rm srd}}$ and $\tau_{z_{\rm srd}}$ to IBS effect, the quantities $H=\eta^2$ and $\varepsilon_{x,z}$ will evolve in accordance with [30] as:

$$\frac{{\rm d}H}{dt}=-\frac{2}{\tau_{\eta_{\rm srd}}}\big(H-H_0\big)+\frac{2}{\tau_\eta}H\qquad\frac{{\rm d}\varepsilon_{x,z}}{dt}=-\frac{2}{\tau_{x_{\rm srd},z_{\rm srd}}}\big(\varepsilon_{x,z}-\varepsilon_{x_0,z_0}\big)+\frac{2}{\tau_{x,z}}\varepsilon_{x,z}\,.\tag{135}$$

Equilibrium emittances $\varepsilon_{x_{\rm eq},z_{\rm eq}}$ and $H_{\eta_{\rm eq}}$ are reached once ${\rm d}\varepsilon_{x,z}/{\rm d}t=0$ and ${\rm d}H/{\rm d}t=0$. One find

$$H_{\eta_{\rm eq}}=\frac{\tau_\eta}{\tau_\eta-\tau_{\eta_{\rm srd}}}H_0\qquad\varepsilon_{x_{\rm eq},z_{\rm eq}}=\frac{\tau_{x,z}}{\tau_{x,z}-\tau_{x_{\rm srd},z_{\rm srd}}}\varepsilon_{x_0,z_0}\,.\tag{136}$$

As said above, iterative computations are required simultaneously in longitudinal and transverse planes to get the equilibrium.

Warning! The form and the units of the first column of Eq. (129), (e.g. $\beta_x^{-1}{\rm d}\langle\varepsilon_x\rangle/{\rm d}t$ in [rad^2]) do not fit those of Eq. (134) where $1/\tau_{\eta,x,z}$ is in [s^{-1}], as shown below. Therefore, the following quantities, introduced for suitability, (the last two are in [rad^{-2}]) will be added to the last column of (138) below for compatibility

$$\frac{(1-d^2)\,q^2}{c^2}=\frac{\gamma^2}{\sigma_\eta^2}\qquad\frac{a^2q^2}{c^2}=\frac{1}{\sigma_{x'_\beta}^2}\qquad\frac{b^2q^2}{c^2}=\frac{1}{\sigma_{z'_\beta}^2}\,,\tag{137}$$

For example, considering τ_x^{-1} and ε_x, we compute $(2\langle\varepsilon_x\rangle)^{-1}{\rm d}\langle\varepsilon_x\rangle/{\rm d}t=(2\sigma_{x_\beta}^2)^{-1}{\rm d}\langle\varepsilon_x\rangle/{\rm d}t=(2\sigma_{x'_\beta}^2\beta_x)^{-1}{\rm d}\langle\varepsilon_x\rangle/{\rm d}t=(a^2q^2/2\,c^2)^{-1}{\rm d}\langle\varepsilon_x/\beta_x\rangle/{\rm d}t$ (with $\alpha_x=0$) using (115)–(137); similarly for $\tau_{z,\eta}^{-1}$, we find

$$\begin{pmatrix}\frac{1}{\tau_\eta}\\ \frac{1}{\tau_x}\\ \frac{1}{\tau_z}\end{pmatrix}=\begin{pmatrix}\frac{1}{2\sigma_\eta^2}\frac{{\rm d}\sigma_\eta^2}{{\rm d}t}\\ \frac{1}{2\sigma_{x_\beta}^2}\frac{{\rm d}\sigma_{x_\beta}^2}{{\rm d}t}\\ \frac{1}{2\sigma_{z_\beta}^2}\frac{{\rm d}\sigma_{z_\beta}^2}{{\rm d}t}\end{pmatrix}=\begin{pmatrix}\frac{1}{2\langle H\rangle}\frac{d\langle H\rangle}{dt}\\ \frac{1}{2\langle\varepsilon_x\rangle}\frac{d\langle\varepsilon_x\rangle}{dt}\\ \frac{1}{2\langle\varepsilon_z\rangle}\frac{d\langle\varepsilon_z\rangle}{dt}\end{pmatrix}=\frac{q^2}{2\,c^2}\begin{pmatrix}(1-d^2)\frac{\rm d}{{\rm d}t}\frac{\langle H\rangle}{\gamma^2}\\ a^2\quad\frac{\rm d}{{\rm d}t}\frac{\langle\varepsilon_x\rangle}{\beta_x}\\ b^2\quad\frac{\rm d}{{\rm d}t}\frac{\langle\varepsilon_z\rangle}{\beta_z}\end{pmatrix}\neq\begin{pmatrix}\frac{\rm d}{{\rm d}t}\frac{\langle H\rangle}{\gamma^2}\\ \frac{\rm d}{{\rm d}t}\frac{\langle\varepsilon_x\rangle}{\beta_x}\\ \frac{\rm d}{{\rm d}t}\frac{\langle\varepsilon_z\rangle}{\beta_z}\end{pmatrix}\,.\tag{138}$$

where for $D_z = 0$ (see also Eq. (130)),

$$a = \frac{\sigma_h}{\gamma\sigma_{x'_\beta}}, \quad b = \frac{\sigma_h}{\gamma\sigma_{z'_\beta}}, \quad c = q\sigma_h = \Big(\beta\gamma\sqrt{\frac{2b_{\max}}{r_{\rm i}}}\Big)\sigma_h, \quad q = \gamma\exp\Big[\frac{C_{\log}}{2}\Big] \quad d^2 = 1 - \frac{\sigma_h^2}{\sigma_\eta^2}\,, \tag{139}$$

$$\frac{1}{\sigma_h^2} = \left(\frac{1}{\sigma_\eta^2} + \frac{D_x^2}{\sigma_{x_\beta}^2}\right) \iff \sigma_h = \frac{\sigma_\eta\sigma_{x_\beta}}{\sigma_x}$$

or equivalently

$$a = \frac{\beta_x\sigma_\eta}{\gamma\sigma_x}, \quad b = \frac{\beta_z\sigma_\eta}{\gamma\sigma_x}, \quad c = \frac{\sigma_\eta\sigma_{x_\beta}}{\sigma_x}\exp\Big[\frac{C_{\log}}{2}\Big] \quad d = \frac{D_x\sigma_\eta}{\sigma_x}\,, \tag{140}$$

The remaining three integrals over ξ, θ, ζ in Eq. (129) still need to be computed to get the mean change of the invariants $\langle\varepsilon_{x,z}\rangle$ and $\langle H\rangle = \eta^2$ due to the multiple collisions of the bunch particles circulating along the lattice. To this end, a first change of variables $(\xi, \theta, \zeta) \to (u, v, w)$ is made as a first step toward the integration over the 'angles' ξ, θ and ζ; that is,

$$\big(2u = q\xi,\ 2v = q\theta,\ 2w = q\zeta\big) \implies u^2 + v^2 + w^2 = \frac{q^2}{4}(\xi^2 + \theta^2 + \zeta^2)\,. \tag{141}$$

The aim of these approximations and changes of variables is to derive an 'almost closed-form' and manageable IBS formula for estimation of the rise times of the mean oscillation amplitudes, which determine the bunch dimensions caused by the effects of IBS. Hence, in the framework of the Piwinski model, the growth rates are calculated in accordance with Eq. (129) reformulated using the new set of variables (u, v, w).

This gives, converting in Eq. (129) the expressions $\gamma^2(1/\sigma_\eta^2 + D_x^2/\sigma_{x_\beta}^2)$ into q^2/c^2 and $\gamma^2 D_x^2/\beta_x^2$ into d^2/a^2 using (139) and (140)

$$\begin{pmatrix} \frac{1}{\tau_\eta} \\ \frac{1}{\tau_x} \\ \frac{1}{\tau_z} \end{pmatrix} = \left\langle \frac{q^2}{2\,c^2} \begin{pmatrix} (1-d^2)\,\frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle H\rangle}{\gamma^2} \\ a^2 \quad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle\varepsilon_x\rangle}{\beta_x} \\ b^2 \quad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle\varepsilon_z\rangle}{\beta_z} \end{pmatrix} \right\rangle = \frac{8\mathcal{A}_{\rm P}}{c^2}\iiint_{-\infty}^{\infty}\exp\Big[-\frac{1}{c^2}\big(u^2 + a^2v^2 + b^2w^2\big)\Big]$$

$$\times\,\frac{\ln[u^2+v^2+w^2]}{\big(u^2+v^2+w^2\big)^{3/2}} \left\{ \begin{matrix} (1-d^2)(-2u^2+v^2+w^2) \\ a^2\big((u^2-2v^2+w^2) + \big(\frac{d}{a}\big)^2(-2u^2+v^2+w^2)\big) \\ b^2(u^2+v^2-2w^2) \end{matrix} \right\} \mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,, \tag{142}$$

Eq. (142) is then further transformed to a triple integral in spherical-like coordinates $(u, v, w) \to (\sqrt{r}, \mu, \nu)$ with

$$\big(u = \sqrt{r}\sin\mu\cos\nu,\ v = \sqrt{r}\sin\mu\sin\nu,\ w = \sqrt{r}\cos\mu\big) \implies u^2 + v^2 + w^2 = r\,, \tag{143}$$

which readily gives, using *Mathematica*

$$\begin{pmatrix} \tau_\eta^{-1} \\ \tau_x^{-1} \\ \tau_z^{-1} \end{pmatrix} = \frac{\mathcal{A}_{\rm P}}{c^2}\int_0^\infty \mathrm{d}r \int_0^\pi \mathrm{d}\mu \int_0^{2\pi} \mathrm{d}\nu\,\sin[\mu]\exp[-rD(\mu,\nu)]\ln[r]$$

$$\times \left\{ \begin{matrix} (1-d^2)\Big[\cos^2[\mu] - \frac{1}{2}(1 + 3\cos[2\nu]\sin^2[\mu])\Big] \\ \big(a^2+d^2\big)\big(1+3\cos[2\mu]\big) + 6\big(a^2-d^2\big)\cos[2\nu]\sin^2[\mu] \\ -2\,b^2\big(1+3\cos[2\mu]\big) \end{matrix} \right\}\,, \tag{144}$$

where

$$D(\mu,\nu)=\frac{1}{c^2}\Big(b^2\cos^2[\mu]+\sin^2[\mu]\big(\cos^2[\nu]+a^2\sin^2[\nu]\big)\Big)\,. \tag{145}$$

Therefore, introducing the following three functions:

$$g_1[\mu,\nu]=1-3\sin^2[\mu]\cos^2[\nu]\quad g_2[\mu,\nu]=1-3\sin^2[\mu]\sin^2[\nu]\quad g_3[\mu,\nu]=1-3\cos^2[\mu]\,, \tag{146}$$

Eq. (144) can be cast in the compact form, in which the outer bracket $\langle\,\cdot\,\rangle$ averages over the lattice parameters,

$$\begin{aligned}\begin{pmatrix}\tau_\eta^{-1}\\ \tau_x^{-1}\\ \tau_z^{-1}\end{pmatrix}&=\left\langle\frac{q^2}{2\,c^2}\begin{bmatrix}(1-d^2)\\ a^2\\ b^2\end{bmatrix}\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\langle H\rangle/\gamma^2\\ \langle\varepsilon_x\rangle/\beta_x\\ \langle\varepsilon_z\rangle/\beta_z\end{bmatrix}\right\rangle=\frac{\mathcal{A}_{\mathrm{P}}}{c^2}\int_0^\infty\mathrm{d}r\int_0^\pi\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\nu\\ &\times\sin[\mu]\exp[-rD(\mu,\nu)]\ln[r]\begin{Bmatrix}(1-d^2)\,g_1[\mu,\nu]\\ a^2g_2[\mu,\nu]+d^2g_1[\mu,\nu]\\ b^2g_3[\mu,\nu]\big)\end{Bmatrix}\,.\end{aligned} \tag{147}$$

Now, let us define the 'scattering function' by means of the triple integrals

$$f(a,b,c)=2\int_0^\pi\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\nu\,\sin[\mu]\,(1-3\cos^2[\mu])\int_0^\infty\mathrm{d}\rho\,\log[c^2\rho]\exp\Big[-\rho\,D_0(\mu,\nu)\Big] \tag{148}$$

where a new variable $\rho=r/c^2$ is introduced, a,b,c being defined in (139) and

$$D_0(\mu,\nu)=\Big(\sin^2[\mu]\big(a^2\cos^2[\nu]+b^2\sin^2[\nu]\big)+\cos^2[\mu]\Big)\,. \tag{149}$$

Notice that $D_0(\mu,\nu)\neq D(\mu,\nu)$. The function $\mathrm{f(a,b,c)}$ cannot be evaluated in closed form over the three variables, but the single integral over ρ can be solved analytically (e.g. by *Mathematica*):

$$\int_0^\infty\mathrm{d}\rho\log[c^2\rho]\exp\big[-\rho\,D_0(\mu,\nu)\big]=\frac{2\log c-C_{\text{Euler}}-\log\big[D_0(\mu,\nu)\big]}{D_0(\mu,\nu)}\,, \tag{150}$$

where $C_{\text{Euler}}=0.5772$ is Euler's constant. Thus, (148) reduces to the double integral

$$f(a,b,c)=2\int_0^\pi\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\nu\,\sin[\mu]\,\big(1-3\cos^2[\mu]\big)\,\frac{2\log c-C_{\text{Euler}}-\log\big[D_0(\mu,\nu)\big]}{D_0(\mu,\nu)}\,, \tag{151}$$

From (139) we obtain $2\log c=C_{\log}+2\log[\gamma\sigma_h]\approx C_{\log}$ assuming that $C_{\log}\gg\log[\gamma\sigma_h]$. This approximation sounds fine as usually $10\lesssim C_{\log}\lesssim 20$ $\big($e.g. 7 Tev LHC: $\gamma=7000$, $\sigma_\eta\approx10^{-4}$, $C_{\log}\approx20$, and taking $\sigma_h\approx\sigma_\eta$ we get $\log[c^2]\approx C_{\log}\gg\log\big[\gamma^2\sigma_\eta^2\big]=-0.7\big)$.

Following Evans and Zotter approach [25], the scattering function (151) is first transformed by a change of variables ($x=\cos\mu$, $y=2\,\nu$), using the periodicity of π and symmetry about $\pi/2$ of $\cos^2[\nu]$ and $\sin^2[\nu]$ (as also μ) which allows to replace the limit π of μ by $\pi/2$ and 2π of ν by $\pi/2$ and then to multiply the integral by an additional factor 8. Therefore, since $\mathrm{d}\mu\,\mathrm{d}\nu=-(2\,\sin[\mu])^{-1}\mathrm{d}x\,\mathrm{d}y$, and with the new limits of integration ($0\leq\mu\leq\pi/2\rightarrow1\leq x\leq0$) and ($0\leq\nu\leq\pi/2\rightarrow0\leq y\leq\pi$), the scattering function (151) becomes :

$$f(a,b,c)=8\int_0^1\mathrm{d}x\int_0^\pi\mathrm{d}y\,\big(1-3\cos^2[\mu]\big)\,\frac{2\log c-C_{\text{Euler}}-\log\big[D_0(\mu,\nu)\big]}{D_0(\mu,\nu)}\,. \tag{152}$$

Finally, Eq. (152) can be reduced to the single integral representation, see [25] and [28], [3], [24]:

$$f(a,b,c) = 8\pi \int_0^1 \left(2\,\ln\left[\frac{\tilde{C}}{2}\left\{\frac{1}{\sqrt{P(x)}} + \frac{1}{\sqrt{Q(x)}}\right\}\right] - C_{\text{Euler}}\right) \frac{1-3x^2}{\sqrt{P(x)Q(x)}}\,\mathrm{d}x, \tag{153}$$

$$\text{with} \quad P(x) = a^2 + (1-a^2)x^2, \quad Q(x) = b^2 + (1-b^2)x^2, \quad \tilde{C} = \log[c^2] - C_{\text{Euler}}\,.$$

The function $f(a,b,c)$ is the 'new scattering function' (see [25] for a full, clear and detailed derivation). Except for a few cases, its calculation requires numerical integration.

After some more work the IBS growth rates for bunched beams can be cast into the compact form that suits Eq. (147)

$$\begin{pmatrix} \dfrac{1}{\tau_\eta} \\ \dfrac{1}{\tau_x} \\ \dfrac{1}{\tau_z} \end{pmatrix} = \mathcal{A}_{\text{P}} \begin{pmatrix} \dfrac{\sigma_{x_\beta}^2}{\sigma_x^2}\, f(a,b,c) \\ f\left(\dfrac{1}{a}, \dfrac{b}{a}, \dfrac{c}{a}\right) + \dfrac{D_x^2\sigma_\eta^2}{\sigma_{x_\beta}^2} f(a,b,c) \\ f\left(\dfrac{1}{b}, \dfrac{a}{b}, \dfrac{c}{b}\right) \end{pmatrix}. \tag{154}$$

where $\sigma_{x_\beta}^2/\sigma_x^2 = 1 - D_x^2\sigma_\eta^2/\sigma_x^2 = 1 - d^2 \equiv \sigma_h^2/\sigma_\eta^2$ according to (139) and (140). Either Eq. (148) or Eq. (153) can be used for $f(a,b,c)$ but it is faster to evaluate the growth rates with the single integral. See the appendix A for a proof of the equivalence between the two formulations (147) and (154).

3.3 The Bjorken–Mtingwa IBS model

3.3.1 Beam phase space density and emittance

A Gaussian probability distribution is taken to characterize the density of the beam in the six-dimensional phase space $\{\mathbf{r}, \mathbf{p}\}$, with $\mathbf{r} = (x, z, s)$ and $\mathbf{p} = (p_x, p_z, p_s)$, where x, z and s stand for the horizontal, vertical and longitudinal directions. The model is formulated as follows (see [2]), in connection with the work of Piwinski [1] (see also [26] for a more recent and enlightening discussion of the topic):

$$\begin{gathered} P(\mathbf{r},\mathbf{p}) = \frac{N_b}{\Gamma}\exp\big[-S(\mathbf{r},\mathbf{p})\big], \qquad \Gamma = \int \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{p}\exp\big[-S(\mathbf{r},\mathbf{p})\big], \\ S(\mathbf{r},\mathbf{p}) = \frac{1}{2}\sum_{i,j=1}^{3}\Big(A_{ij}\delta p_i\delta p_j + 2B_{ij}\delta p_i\delta r_j + C_{ij}\delta r_i\delta r_j\Big) = S^{(x)} + S^{(z)} + S^{(s)}\,, \end{gathered} \tag{155}$$

where $P(\mathbf{r},\mathbf{p})$ is the phase space density of the beam containing N_b particles, Γ is the phase space 'volume' of the beam and $S(\mathbf{r},\mathbf{p})$ represents the nature of the Gaussian particle beam probability distribution, with $\delta\mathbf{r}$ and $\delta\mathbf{p}$ denoting the position and momentum from the reference values $\mathbf{r}$ and $\mathbf{p}$. Upon working out the coefficients A_{ij}, B_{ij} and C_{ij}, the expression for $S(\mathbf{r},\mathbf{p})$ can be written as

$$\begin{gathered} S(\mathbf{r},\mathbf{p}) = S^{(x)} + S^{(z)} + S^{(s)}, \\ S^{(x)} = \frac{\beta_x}{2\sigma_{x_\beta}^2}\big(\gamma_x x_\beta^2 + 2\alpha_x x_\beta x'_\beta + \beta_x x'^2_\beta\big), \quad S^{(z)} = \frac{\beta_z}{2\sigma_{z_\beta}^2}\big(\gamma_z z_\beta^2 + 2\alpha_z z_\beta z'_\beta + \beta_z z'^2_\beta\big), \\ S^{(s)} = \frac{\eta^2}{2\sigma_\eta^2} + \frac{(s-s_0)^2}{2\sigma_s^2}\,, \end{gathered} \tag{156}$$

where $\alpha_{x,z}$, $\beta_{x,z}$ and $\gamma_{x,z}$ are the Twiss parameters, $\varepsilon_{x,z}$ are the r.m.s. transverse beam emittances, ε_s is the r.m.s. longitudinal beam emittance, σ_η is the r.m.s beam momentum spread, σ_s is the r.m.s. beam

length, and $\sigma_{x,z}$ are the r.m.s. beam width and height. Moreover,

$$\varepsilon_x = \frac{\sigma_{x_\beta}^2}{\beta x}\,, \quad \varepsilon_z = \frac{\sigma_{z_\beta}^2}{\beta z}\,, \quad \varepsilon_s = \sigma_\eta \sigma_s\,, \quad x_\beta = x - D_x\eta\,, \qquad (157)$$
$$z_\beta = z - D_z\eta\,, \quad x'_\beta = x' - D'_x\eta\,, \quad z'_\beta = z' - D'_z\eta$$

where

$$x' = \frac{\Delta p_x}{p}, \quad z' = \frac{\Delta p_z}{p}, \quad \eta = \frac{\Delta p}{p}\,.$$

From now on the Heaviside–Lorentz (HL) units $\epsilon_0 = \hbar = c = 1$ will be used instead of the SI units kg, m, s etc. (with a few exceptions). To convert back to SI units, we must restore the missing factors $\epsilon_0, \hbar$ and c.

Comment: Since $\hbar \stackrel{\text{def}}{=} h/2\pi = 1.054 \times 10^{-34}\,\text{J s}$, in HL units the value of the Planck constant is $h = 2\pi$. Referring to the footnote associated with Eq. (1), the number of 'single-particle states' in the momentum volume $\mathrm{d}^3\mathbf{p}$ is $\mathrm{d}^3\mathbf{p}/h^3$, which in HL units is $\mathrm{d}^3\mathbf{p}/(2\pi)^3$.

3.3.2 Two-body scattering

The Bjorken and Mtingwa approach to IBS modelling is based on the *S-matrix*, which is a time-evolution operator relating the transition from an *initial* (quantum) state $|i\rangle$ to a *final* state $|f\rangle$ of a physical system undergoing a scattering process; the matrix elements of S are inner products denoted by $\langle f|S|i\rangle$.

Comment: The S-matrix is proportional to the *amplitude* $\mathcal{M}$, which represents the physics of the process [31]: $S \propto (2\pi)^4\delta^4(p_f - p_i)\mathcal{M}$ (which we take for granted at this stage), where p_f and p_i are the 4-momenta of the outgoing and incoming states. The δ-function enforces the momentum conservation of the process.

The squared modulus $|\langle f|S|i\rangle|^2$ is interpreted as a *transition probability* $\mathcal{P}$ for a transition from an initial state to a final one. See the appendix B for more details about this topic. Bjorken's formalism provides new insight into the other theories; in particular, the IBS calculations allow for the case of *strong-focusing* lattices.

In a two-body scattering process, particles 1 and 2 with 4-momenta $p^\mu_{1,2}$ (written in brief as $p_{1,2} \stackrel{\text{def}}{=} p^\mu_{1,2}$), i.e. with energy–momentum 4-vectors $p^\mu_{1,2}$, interact with each other to give the following two 4-momenta after collision: $p'_{1,2} \stackrel{\text{def}}{=} p'^\mu_{1,2}$; cf. Eq. (160). The vectors $\mathbf{p}_{1,2}$ are the usual 3-momenta (e.g. $\mathbf{p}_1 + \mathbf{p}_2 \to \mathbf{p}'_1 + \mathbf{p}'_2$ for 3-momenta and $p_1 + p_2 \to p'_1 + p'_2$ for 4-momenta). The *transition rate* for the two-particle scattering process, or equivalently the number of scattering events per unit time, is given by Eq. (2.3) in [2]; see also Eq. (7.42bis) in [32]:

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}t} = \frac{1}{2}\int \mathrm{d}^3\mathbf{r}\frac{\mathrm{d}^3\mathbf{p}_1}{\gamma_1}\frac{\mathrm{d}^3\mathbf{p}_2}{\gamma_2}\,P(\mathbf{r},\mathbf{p}_1)\,P(\mathbf{r},\mathbf{p}_2)\big|\mathcal{M}\big|^2\,\frac{\mathrm{d}^3\mathbf{p}'_1}{\gamma'_1}\,\frac{\mathrm{d}^3\mathbf{p}'_2}{\gamma'_2}\,\frac{\delta^4(p'_1+p'_2-p_1-p_2)}{(2\pi)^2}\,, \qquad (158)$$

where $\gamma_{1,2} = E_{1,2}/m$, with m being the mass of the two particles (assumed to be the same) and $E_{1,2}$ their energies (in HL units), and $\mathcal{M}$ is the Lorentz-invariant Coulomb scattering amplitude.

The transition rate $\mathrm{d}\mathcal{P}/\mathrm{d}t$ caused by a two-particle scattering process, (158), can be reformulated by introducing the exponent $S(\mathbf{r},\mathbf{p})$ of the Gaussian beam phase space distribution $P(\mathbf{r},\mathbf{p})$ in (155), to calculate the rate of change of the emittances ε_u (with $u = x, z, s$); this yields (cf. [2, 33])

$$\begin{aligned}\frac{\mathrm{d}\varepsilon_u}{\mathrm{d}t} = \frac{N_b}{2\Gamma^2}\int \mathrm{d}^3\mathbf{r}\,\frac{\mathrm{d}^3\mathbf{p}_1}{\gamma_1}\,\frac{\mathrm{d}^3\mathbf{p}_2}{\gamma_2}\,\exp\big[-S(\mathbf{r},\mathbf{p}_1)\big]\exp\big[-S(\mathbf{r},\mathbf{p}_2)\big] \\ \times\,\big|\mathcal{M}\big|^2\big(\varepsilon_u(\mathbf{p}'_1)-\varepsilon_u(\mathbf{p}_1)+\varepsilon_u(\mathbf{p}'_2)-\varepsilon_u(\mathbf{p}_2)\big)\,\frac{\mathrm{d}^3\mathbf{p}'_1}{\gamma'_1}\,\frac{\mathrm{d}^3\mathbf{p}'_2}{\gamma'_2}\,\frac{\delta^4(p'_1+p'_2-p_1-p_2)}{(2\pi)^2}\,.\end{aligned} \qquad (159)$$

The goal is now to compute the scattering amplitude $\mathcal{M}$ for a Coulomb interaction between two particles. This will be done by means of the *Feynman rules*, using the *Feynman diagram* representation of a simple scattering process.

Comment: Feynman diagrams are graphical representations of interactions according to a set of rules that allow us to calculate the matrix elements and amplitude $\mathcal{M}$ of a given interaction. With this approach, the scattering cross-sections, decay rates, transition probabilities etc. can be calculated via a function, called the *propagator*, which represents the transfer of momentum from one particle to another. Propagators are obtained by following the prescriptions of the Feynman rules.

Computation of $\mathcal{M}$ for a simplified process is sketched in the following. Let r^μ denote a contravariant vector, which, together with the corresponding covariant vector $r_\mu \stackrel{\text{def}}{=} g_{\mu\nu} r^\nu$, makes the product $g_{\mu\nu} r_\mu r_\nu$ invariant with respect to the Lorentz transform (with $g_{11} = 1$, $g_{22} = g_{33} = g_{44} = -1$, and $g_{\mu\neq\nu} = 0$).

Comment: The metric is 4-momentum^2 = energy^2 − 3-momentum^2 (in HL units), where 4-momenta are written as p and 3-momenta as $\mathbf{p}$ (boldface).

Then we can write (briefly reintroducing $c = 3 \times 10^8 \,\mathrm{m\,s^{-1}}$ in place of its HL value $c = 1$)

$$
\begin{gathered}
r \stackrel{\text{def}}{=} r^\mu \equiv (ct, \mathbf{r}) = (ct, x, z, s), \qquad p \stackrel{\text{def}}{=} p^\mu \equiv \left(\frac{E}{c},\, \mathbf{p}\right) = \left(\frac{E}{c},\, p_x,\, p_z,\, p_s\right), \\
p_1 \cdot p_2 \stackrel{\text{def}}{=} p_1^\mu p_{2_\mu} \equiv \frac{E_1 E_2}{c} - \mathbf{p}_1 \cdot \mathbf{p}_2, \quad p^2 \stackrel{\text{def}}{=} p^\mu p_\mu \equiv \frac{E^2}{c^2} - \mathbf{p}^2 = m^2 c^2, \\
r \cdot p \stackrel{\text{def}}{=} r^\mu p_\mu \equiv tE = \mathbf{r} \cdot \mathbf{p}.
\end{gathered}
\tag{160}
$$

The fourth formula in (160) is the well-known squared 3-momentum $|\mathbf{p}|^2 c^2 = E^2 - m^2 c^4$. With the help of these definitions, the integral of the Dirac delta-functions in (158) expresses the conservation of the 4-momentum:

$$
\int \cdots \delta^4(p_1 + p_2' - p_1' - p_2)\, \mathrm{d}p_1\, \mathrm{d}p_2\, \mathrm{d}p_1\, \mathrm{d}p_1'\, \mathrm{d}p_2' \,.
$$

Here the intention is to sketch the techniques involved in analysing the interactions of charged particles. The approach is a kind of makeshift job based on several approximations, such as non-relativistic elastic scattering in the centre-of-mass frame and the assumption of 'structureless' particles. This allows us to regard proton–proton and electron–electron collisions as being on the same footing. (evidently protons, with spin $1/2$, subject to IBS electromagnetic forces within a circulating beam have nothing to do with high-energy head-on collisions between protons circulating in opposite directions and experiencing chromodynamic quark–quark interactions, with spin $1/2$, mediated by the exchange of gluons, with spin 1.)

Let us consider a Coulomb scattering between two electrons of mass m via the exchange of a virtual photon driving the electromagnetic force, as described by quantum electrodynamics (QED). To quantify this scattering process with the minimal amount of formalism, we drastically simplify the computations of the tricky QED mathematics for the 'real-life' four-body process $\mathrm{e}^- + \mathrm{e}^- \to \mathrm{e}^- + \mathrm{e}^-$ with electrons of spin $1/2$ and a massless photon of spin 1 (carrying the electromagnetic force). Instead, we use a 'toy model' which does not handle particles with spin (cf. [34]). Thus, spinless and point-like charged particles with spinless and massless bosons are used to mimic protons and photons.

Figure 13 illustrates the elastic scattering process $\mathbf{p}_1 + \mathbf{p}_2 \to \mathbf{p}_1' + \mathbf{p}_2'$ in the centre-of-mass frame of two particles, say electrons. Unlike a classical Rutherford Coulomb scattering process, for a QED scattering process, referred to as Møller scattering, the force between two electrons results from the exchange of *virtual photons* (i.e. photons that cannot be directly observed) located at two vertices. Figure 14 depicts a two-electron scattering process described using a Feynman diagram. The two electrons enter from the left of the diagram, exchange a photon and then move away to the right of the

diagram. The nature of the interaction is described by relativistic quantum electrodynamic theory (see e.g. [31, 32, 34, 35]).

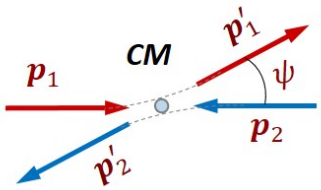


Fig. 13: Kinematics of two-particle scattering in the centre-of-mass system, with scattering angle ψ and 4-momentum $p_{1,2} = (E_{1,2}, \pm\mathbf{p}_{1,2})$; similarly for $p'_{1,2}$ after the collision.

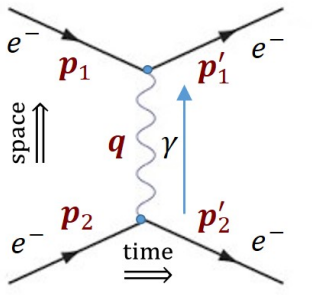


Fig. 14: Feynman diagram for electron–electron scattering in QED; 'time' flows from left to right, the left-hand side of the diagram is the initial state, the right-hand side is the final state, and the wavy line in the middle belongs to neither the initial nor the final state—it shows 'how the interaction happened'. The intermediate photon γ is 'virtual'.

In the framework of the toy model, the electrons are reduced to spin-0 'particles' with mass m, the four *external* lines in the Feynman diagram are associated with the particle 4-momenta p_1, p_2 and p'_1, p'_2 before and after the collision. The photon becomes a spin-0 and zero-mass 'particle', with one *internal* wavy line associated with the internal 4-momentum labelled $\mathbf{q}$.

The amplitude $\mathcal{M}$ for the scattering process is worked out using the Feynman diagram in Fig. 14 together with the Feynman rules. The steps of the Feynman prescriptions for toy models are sketched without proof.

1. **Notation:** Label the incoming and outgoing external 4-momenta p_1, p_2, p'_1, p'_2 and the internal 4-momentum q (with $p_{1,2} \stackrel{\text{def}}{=} p^\mu_{1,2}$ and $q \stackrel{\text{def}}{=} q^\mu$ as above).
2. **Vertex factor:** The two vertices contribute two factors of $-\mathrm{i}g$, where $\mathrm{i} = \sqrt{-1}$. Multiplying these two factors together yields $-g^2$. Note that the coupling constant g in QED specifies the interaction strength between electrons and photons and is related to the *fine structure constant* α through $g = \sqrt{4\pi\alpha}$. In SI units $\alpha = e^2/(4\pi\epsilon_0\hbar c) = 1/137$, and in HL units $\alpha = e^2/(4\pi)$.
3. **Propagator:** The single internal wavy line contributes a factor of $f(q) = \mathrm{i}/(|q|^2 - m_i^2) \equiv \mathrm{i}/|q|^2$ because of the spinlessness and zero mass ($m_i = 0$) of the internal 'boson' (the mimicked photon). Note that a propagator $f(q)$ is associated with the internal wavy line in the Feynman diagram and represents the transfer, or propagation, of momentum from one electron e^- to the other during the interaction time, via a virtual photon γ.
4. **Energy–momentum conservation:** For the two vertices, introduce two Dirac delta-functions $(2\pi)^4\delta^4(p_2 - p'_2 - q)$ and $(2\pi)^4\delta^4(p_1 + q - p'_1)$ and multiply them together. The scattering amplitude is expressed as the product of the coupling constant g and the propagator $f(q)$, namely $\mathcal{M} = gf(q)$. Note that this prescription enforces the conservation of energy and momentum at each vertex, because the delta-functions are zero when the sum of the incoming 4-momenta and that of the outgoing 4-momenta are the same.
5. **Integration over internal 4-momenta:** Then integrate the delta-functions for the unique internal 4-momentum q over the variables $\mathrm{d}^4q/(2\pi)^4$.

Steps 1–5 give the amplitude approximation for non-relativistic scattering of spinless and point-like charged 'particles' (with mass m) and spinless and massless 'bosons' mimicking electrons and photons:

$$\mathcal{M} = -\mathrm{i}\,(2\pi)^4 g^2 \int \frac{1}{|q|^2}\,\delta^4\left(p_2 - p_2' - q\right)\delta^4\left(p_1 + q - p_1'\right)\mathrm{d}^4 q\,. \tag{161}$$

Integration of the delta-function is done by inspection. Choosing the second delta-function to integrate, we substitute $q \mapsto p_1' - p_1$ into (161) and obtain

$$\begin{aligned}\mathcal{M} &= -\mathrm{i}\,(2\pi)^4 g^2 \int \frac{1}{|q|^2}\,\delta^4\big(p_1 + p_2 - p_1' - p_2'\big)\delta^4\big(q - [p_1' - p_1]\big)\,\mathrm{d}^4 q \\ &= -\frac{\mathrm{i}\,g^2}{(p_1' - p_1)^2}\,(2\pi)^4\delta^4\big(p_1 + p_2 - p_1' - p_2'\big)\end{aligned} \tag{162}$$

since $\int \delta^4(q - [p_1' - p_1])\,\mathrm{d}^4 q = 1$.

Discarding the left-out delta-function $(2\pi)^4\delta^4(p_1+p_2-p_1'-p_2')$, which just re-enforces the overall conservation of energy and momentum at the four external lines (i.e. external 'particles', the mimicked electrons), we get the scattering amplitude in terms of the 4-momenta (cf. Eq. (2.3a) in [2]):

$$\mathcal{M} = -\frac{\mathrm{i}\,g^2}{(p_1' - p_1)^2} = \frac{4\pi\alpha}{(p_1' - p_1)^2}\,. \tag{163}$$

To see the link between $\mathcal{M}$ and the collisional process, the internal 4-momentum squared, $|q|^2 \equiv (p_1' - p_1)^2$, can be expanded further. Consider the 4-momentum (and energy) conservation $|p_{1,2}'|^2 = |p_{1,2}|^2$ for elastic collision in the centre-of-mass frame (Fig. 13). We have $E_{1,2}^2 = E_{1,2}'^2$ and $|\mathbf{p}|_{1,2}^2 = |\mathbf{p}|_{1,2}'^2$ since $p_{1,2} = (E_{1,2}, \mathbf{p}_{1,2})$ and $p_{1,2}^2 = (E_{1,2}^2 - \mathbf{p}_{1,2}^2) \equiv m_{1,2} = m$ for particles of the same mass m. Hence

$$\begin{aligned} p_1' - p_1 = (E_1' - E_1) + (\mathbf{p}_1' - \mathbf{p}_1) = \mathbf{p}_1' - \mathbf{p}_1 \quad &\Longrightarrow \\ (\mathbf{p}_1' - \mathbf{p}_1)^2 = \mathbf{p}_1'^2 + \mathbf{p}_1^2 - 2\mathbf{p}_1'\cdot\mathbf{p}_1 &= \mathbf{p}_1'^2 + \mathbf{p}_1^2 - 2|\mathbf{p}_1'|\,|\mathbf{p}_1|\cos\psi \\ = 2|\mathbf{p}|^2(1-\cos\psi) &= |\mathbf{p}|^2\sin[\psi/2]\,, \end{aligned} \tag{164}$$

in which $\mathbf{p} \overset{\text{def}}{=} \mathbf{p}_1$ is the incident momentum of particle 1 and ψ is the scattering angle between the two momenta $\mathbf{p}_1$ and $\mathbf{p}_1'$ before and after collision. Therefore, in HL units, the amplitude $\mathcal{M}$ takes the form

$$\mathcal{M} = \frac{4\pi\alpha}{|\mathbf{p}|^2\sin^2[\psi/2]} \quad\Longrightarrow\quad \big|\mathcal{M}\big|^2 = \left(\frac{e^2}{|\mathbf{p}|^2\sin^2[\psi/2]}\right)^2 \equiv \sigma(|\mathbf{p}|,\psi)\,. \tag{165}$$

The scattering amplitude $\big|\mathcal{M}\big|^2$ is thus identified with the differential cross-section $\sigma(|\mathbf{p}|,\psi)$ of the two-electron collisional process.

Comment: Compare (165) with the cross-section $\sigma(\psi)$ of a Coulomb classical (non-quantal) Rutherford scattering process, Eq. (36).

At this point the calculations are still far from finished, and also they are not easy to perform. After some difficult manipulations the rate of change of the emittances, $\mathrm{d}\varepsilon_u/\mathrm{d}t$ of Eq. (159), can be recast in the form given by Eq. (166) below. See [2] for details of the lengthy calculations used to derive the Bjorken–Mtingwa Eq. (3.4) and then Eqs. (4.5)–(4.7), which yield growth rate expressions that are convenient to use.

3.3.3 Intrabeam scattering growth rates

In deriving Eq. (166), Bjorken and Mtingwa considered a zero vertical dispersion function D_z, so that $\phi_z = 0$ and $H_z = 0$, reducing L_z (168) to a matrix with all components $L_{z_{i,j}} = 0$ except for $L_{z_{3,3}} = \beta_z/\varepsilon_z$:

$$L_z = \frac{\beta_z}{\varepsilon_z}\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix} \text{ (Eq. (2.37d) of [2]) compared to } L_z = \frac{\beta_z}{\varepsilon_z}\begin{pmatrix}0&0&0\\0&\frac{\gamma^2 H_z}{\beta_z}&-\gamma\phi_z\\0&-\gamma\phi_z&1\end{pmatrix} \text{ in Eq. (168)}.$$

Moreover, in solving (166) Bjorken and Mtingwa did an approximation by neglecting the ratios $\beta_{x,z}/\varepsilon_{x,z}$ relative to $\gamma^2 D_x^2/(\varepsilon_x\varepsilon_z)$, $(\beta_x/\varepsilon_z)\gamma^2\phi_x^2$ and γ^2/σ_η.

To conclude, for bunched beams, the IBS growth rates τ_u^{-1} in the horizontal, vertical and longitudinal directions ($u = x, z, s$) are written in the form similar to that obtained by Bjorken and Mtingwa [2],

$$\begin{aligned}\frac{1}{\tau_u} &= \frac{1}{\sigma_u}\frac{d\sigma_u}{dt} = \frac{1}{\sqrt{\langle\varepsilon_u\rangle}}\frac{\mathrm{d}\sqrt{\langle\varepsilon_u\rangle}}{\mathrm{d}t} = \frac{1}{2\,\langle\varepsilon_u\rangle}\frac{\mathrm{d}\langle\varepsilon_u\rangle}{\mathrm{d}t} = \\ &= \mathcal{A}_{\mathrm{BM}}\,C_{\log}\left\langle\int_0^\infty \frac{\mathrm{d}\lambda\,\sqrt{\lambda}}{\sqrt{\det[L+\lambda I]}}\left\{\mathrm{Tr}[L_u]\,\mathrm{Tr}\left[\frac{1}{L+\lambda I}\right] - 3\,\mathrm{Tr}\left[L_u\Big(\frac{1}{L_u+\lambda I}\Big)\right]\right\}\right\rangle,\end{aligned} \tag{166}$$

in which we have introduced the growth rates (134), the rough momentum spread squared $\langle H\rangle \approx \langle\eta\rangle^2 = \sigma_\eta^2$, and the Bjorken-Mtingwa scattering constant $\mathcal{A}_{\mathrm{BM}}$ in analogy with (128):

$$\begin{aligned}\mathcal{A}_{\mathrm{BM}}\,C_{\log} = 4\pi\mathcal{A}_{\mathrm{P}}\,C_{\log} &= \frac{c r_{\mathrm{i}}^2 N_b\,C_{\log}}{16\pi\beta^3\gamma^4\varepsilon_x\varepsilon_z\varepsilon_s} \equiv \frac{\pi^2 N_b\,c_0\,r_{\mathrm{i}}^2\,C_{\log}}{2\gamma\Gamma},\\ \Gamma = (2\pi)^3(\beta\gamma)^3\varepsilon_x\varepsilon_z\varepsilon_s, &\qquad \frac{1}{\tau_s} = \frac{1}{\sigma_\eta}\frac{\mathrm{d}\sigma_\eta}{\mathrm{d}t},\end{aligned} \tag{167}$$

where Γ is the six-dimensional phase volume for bunched beams and $\varepsilon_s \overset{\mathrm{def}}{=} \sigma_\eta\,\sigma_s$. Moreover, the longitudinal growth rate τ_s, has been remodelled in the handy form above, better suited for practical usage (cf. [30] chapter 13.2 and [37] footnote[3]). Unlike Eq. (134) and (166) the growth rates in Ref. [37] are expressed as the time-derivative of the emittances, yielding a factor 2 difference between both versions. Also, the longitudinal emittance can be written instead as $\varepsilon_s = \sigma_s\sigma_E/\beta^2$ using the relative energy spread $\Delta E/E$ since the relationship between the momentum and energy spreads is $\Delta p/p = (\Delta E/E)/\beta^2$, the two relative spreads being equal at high energy.

The present formulation (cf. [37]) includes the vertical dispersion D_z and its derivative D_z'. The matrix $L = L_x + L_z + L_s$ is composed of the 3×3 matrices defined below in (168), I is the identity matrix, and the brackets $\langle\,\cdot\,\rangle$ represent averaging over the lattice period.

$$L_x = \frac{\beta_x}{\varepsilon_x}\begin{pmatrix}1&-\gamma\phi_x&0\\-\gamma\phi_x&\frac{\gamma^2 H_x}{\beta_x}&0\\0&0&0\end{pmatrix},\quad L_z = \frac{\beta_z}{\varepsilon_z}\begin{pmatrix}0&0&0\\0&\frac{\gamma^2 H_z}{\beta_z}&-\gamma\phi_z\\0&-\gamma\phi_z&1\end{pmatrix},\quad L_s = \frac{\gamma^2}{\sigma_\eta^2}\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix} \tag{168}$$

with

$$\phi_{x,z} = \frac{D_{x,z}\alpha_{x,z} + D'_{x,z}\beta_{x,z}}{\beta_{x,z}},\quad H_{x,z} = \frac{D_{x,z}^2 + \beta_{x,z}^2\phi_{x,z}^2}{\beta_{x,z}} = \gamma_{x,z}D_{x,z}^2 + 2\alpha_{x,z}D_{x,z}D'_{x,z} + \beta_{x,z}D_{x,z}'^2\,.$$

In Eqs. (166)-(168), N_b is the number of particles per bunch, c_0 is the speed of light reintroduced here in SI units, r_{i} is the classical ion radius [m] given in (90); it reduces to the classical proton radius r_0 for unit ion mass and charge A, Z (for leptons of charge Z, A is the lepton-to-electron mass ratio), $\gamma = \left(E_0^2 + p^2\right)^{1/2}/E_0$ and $\beta = \left(1-\gamma^{-2}\right)^{1/2}$ are the Lorentz factors, E_0 is the particle rest energy [eV],

and p is the particle momentum [eV/c]. The matrices inside the brackets depend on the optics parameters $\alpha_{x,z}, \beta_{x,z}, D_{x,z}$ and $D'_{x,z}$, on the r.m.s. unnormalized transverse emittances $\varepsilon_{x,z}$ [m], and on the r.m.s. relative momentum spread σ_η and bunch length σ_s [m] (or the r.m.s. longitudinal emittances [m] ε_s defined above). For matched beams, the longitudinal emittance is defined as $\varepsilon_s = \pi p \sigma_\eta \sigma_s (\beta c)^{-1}$ [eVs]. $C_{\log}$ in Eq. (166) is the Coulomb logarithm, with typical values in the range $[10,\ 20]$.

After expansion of the integrand in the brackets in (166) and some lengthy computations, the growth rates are recomputed to include the vertical dispersion function and cast in the form (cf. [36, 37])

$$\frac{1}{\tau_u} = \frac{\pi^2 N_b\, c_0\, r_i^2\, C_{\log}}{2\gamma\Gamma}\, \Delta_u \left\langle \int_0^\infty \frac{\mathrm{d}\lambda \sqrt{\lambda}\,(a_u\lambda + b_u)}{(\lambda^3 + a\lambda^2 + b\lambda + c)^{3/2}} \right\rangle \tag{169}$$

with $u = x, z, s$ and

$$\Delta_x = \frac{\gamma^2 H_x}{\varepsilon_x}, \quad \Delta_z = \frac{\beta_z}{\varepsilon_z}, \quad \Delta_s = \frac{\gamma^2}{\sigma_\eta^2}.$$

The nine coefficients a, b, c, a_x, b_x, a_z, b_z, a_s and b_s depend on the optics parameters and on the approximation used. They are not reproduced here; see [37] for a complete list, including three variants.

Comment: A high energy approximation to Bjorken-Mtingwa theory [2] has been developed by Bane [27] to give formulae for the IBS growth rates that can be rapid and easy to apply. For instance, IBS effects in high energy storage rings, operating above transition, can induce emittance growth during beam coasts. So, numerical simulations of the emittance evolution caused by IBS involves performing the integrals appearing in the growth rates formulae over numerous storage ring turns, averaging the growth rates for each turn around the storage ring and then deriving the emittance variation for this turn. This task may be quite cumbersome and computationally intensive if one has to use plain growth rates formulae without approximation (see also [28, 30, 33]). Although the Piwinski and Bjorken-Mtingwa formulae for the growth rates look dissimilar, the Piwinski model (assuming weak-focussing lattices) and the Bjorken-Mtingwa one (intrinsically ready-made for strong-focussing lattices) are in good agreement has also shown by Bane, with sure assumptions.

For illustration, Fig. 15 plots the evolution of the Coulomb logarithm for the ELENA 100 keV low-energy antiproton decelerator ring, computed with Eqs. (170) and (171) below.

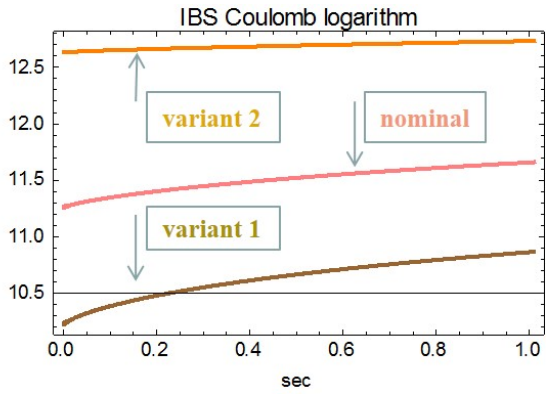


Fig. 15: Evolution of the calculated Coulomb logarithm during 1 s on a 100 keV plateau for the nominal ELENA beam and the first two variants in Table 4.

In [2] the Coulomb logarithm is taken to be the fixed value $C_{\log} = 20$. Here the Coulomb logarithm is defined by the expression (see [23])

$$C_{\log} \equiv \ln\left[r_{\max}/r_{\min}\right] \quad \text{with} \quad r_{\max} = \min[\sigma_x, \lambda_D], \quad r_{\min} = \max\left[r_{\min}^{C}, r_{\min}^{QM}\right], \tag{170}$$

where $r_{\max}$ is the smaller of the mean r.m.s. beam size σ_{x_β} and the Debye length $\lambda_{\rm D}$, and $r_{\min}$ is the larger of the distance of closest approach $r_{\min}^{\rm C}$ and the quantum diffraction limit from the nuclear radius $r_{\min}^{\rm QM}$. These quantities are given by

$$\lambda_{\rm D} = \frac{7.434}{Z}\sqrt{\frac{2E_\perp}{\rho}}, \quad \rho = \frac{N_b \times 10^{-6}}{\sqrt{64\pi^3 \langle\beta_x\rangle \varepsilon_x \langle\beta_z\rangle \varepsilon_z \sigma_s^2}}\,, \quad E_\perp = \frac{(\gamma^2-1)E_0}{2}\frac{\varepsilon_x}{\langle\beta_x\rangle}\,,$$
$$r_{\min}^{\rm C} = \frac{1.4410\times 10^{-9} Z^2}{2E_\perp}, \quad r_{\min}^{\rm QM} = \frac{1.97310\times 10^{-13} Z^2}{\sqrt{8E_\perp E_0}}\,, \tag{171}$$

in which ρ is the particle volume density [m^{-3}] and $E_\perp$ is the transverse beam kinetic energy in the centre-of-mass frame [eV].

4 Examples

4.1 First example: LHC and SLHC at 7 TeV

4.1.1 Nominal LHC, first interaction region upgrade and SLHC

The effects of intrabeam scattering and synchrotron radiation on the expected evolution of the LHC and SLHC beam emittances during coasts at 7 TeV are examined for the nominal beam and beams with reduced emittances. The study was carried out in 2010 for the first 'interaction region' (IR) upgrade and for the SLHC. The nominal beam and LHC parameters and those with reduced emittances have been selected to study the effect of IBS on emittance evolution in a coast [5].

Figure 16 shows the betatron functions along the LHC ring circumference for the LHC 'IR phase 1 triplet' beam and for the lattice parameters of Case 2 in Table 1 with $\beta^* = 0.3$ m, used for all subsequent IBS calculations (small disparities among the other three β^* values were negligible). Except at the interaction points, LHC is a fairly smooth storage ring and so more or less fulfils the *smooth focusing approximation* criteria. The vertical dispersion (see Fig. 17) is generated by the vertical crossing angles at interaction points 1 and 2 and by the detector fields of ALICE and LHCB [37].

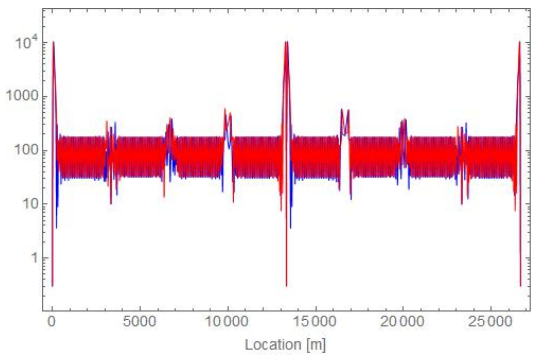


Fig. 16: SLHC betatron functions [m] for $\beta^* = 0.30$ m (at interaction points IP1 and IP5).

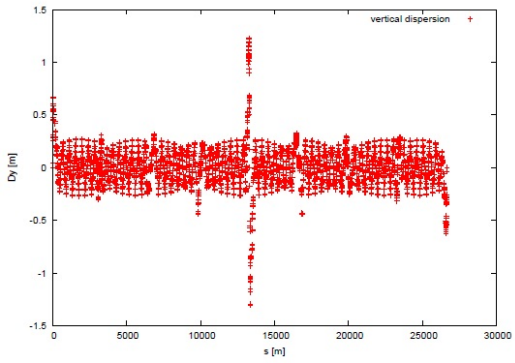


Fig. 17: Vertical dispersion [m] as a function of position [m], for the LHC at 7 TeV.

Table 1 displays a list of beam and lattice parameters related to the luminosity characteristics of the LHC and SLHC. Cases 1 and 2 in the table show the LHC luminosity with a nominal beam intensity of 1.15×10^{11} protons, and Cases 3 and 4 show the SLHC luminosity for the 'highest' ultimate beam intensity of 2.36×10^{11} protons.

Table 1: LHC nominal beam intensity luminosity and SLHC luminosity with improved variants

LHC and SLHC beam parameters	Case 1[a]	Case 2[b]	Case 3[c]	Case 4[d]
N_b $(\times 10^{11})$	1.15	1.15	1.70	2.36
$\varepsilon^n_{H,V} = \gamma\varepsilon$ [μm] (normalized)	3.75	2.54	2.65	2.60
β^* [m]	0.55	0.30	0.25	0.15
$\sigma^*_{H,V}$ [μm]	16.58	10.11	9.40	7.21
σ_L [mm]	75.5	75.5	75.5	75.5
$\sigma_{\Delta p/p}$ $(\times 10^{-4})$	1.13	1.13	1.13	1.13
ε_L r.m.s. [eV s]	0.62	0.62	0.62	0.62
Crossing angle θ [μrad]	285	337	355	452
ΔQ_{bb} head-on[e]	1	1.09	1.43	1.37
Luminosity $\mathcal{L}$ $(\times 10^{-34})$ [cm^{-2} s^{-1}]	2	1.09	4.65	10.29

[a] Initial IR triplet: nominal beam with LHC at top energy gives the nominal luminosity.
[b] IR phase 1 triplet: new optics foreseen for interaction region to give lower β^*, reduced emittance and improved luminosity.
[c] Ultimate N_b: with the ultimate beam intensity and the reduced emittance raise ΔQ_{bb} to still improve the luminosity.
[d] 'Highest' ultimate N_b: top luminosity can be obtained by reducing β^* and raising N_b within an emittance again reduced.
[e] The head-on beam–beam tune shift ΔQ_{bb} is normalized to the value of the nominal beam.

4.1.2 Intrabeam scattering effects in the LHC and SLHC

Figure 18 shows the initial IBS growth times in the LHC and SLHC computed with the beam parameters in Table 1. The simulations were performed with a dedicated *Mathematica* notebook based on the Bjorken–Mtingwa model [2], which handles the vertical dispersion function and its derivative [36].

The *Mathematica* code accounts for the variation of the optical parameters $\beta_{H,V}, \beta'_{H,V}, D_{H,V}$ and $D'_{H,V}$ over the lattice. Table 2 displays the increase in beam emittance due to IBS for a 7 TeV proton

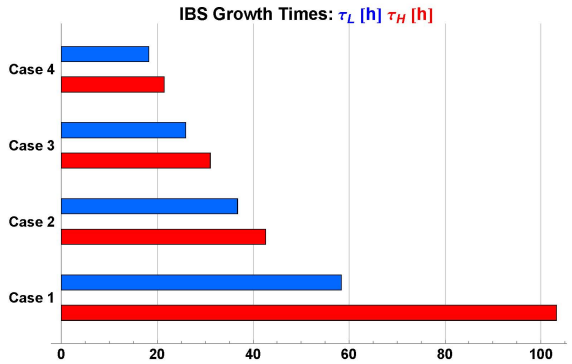


Fig. 18: Initial IBS growth times in hours for the four cases in Table 1. For the nominal LHC parameters (Case 1): $\tau_L = 58$ h, $\tau_H = 103$ h and $\tau_V = -359$ years. The vertical growth times are not shown as they are negative, approximately -100 years, i.e. $\tau_V^{-1} \approx 0$.

beam at the end of 10 hours' storage in the LHC/SLHC.

Figures 19 and 20 plot the emittance evolution over the 10-hour period of coast, assuming a constant beam intensity for the duration of the storage. The *synchrotron radiation* damping effect is not accounted for in the simulations. IBS growth times are calculated iteratively in time steps of 5 minutes.

At each iteration i, the emittances are updated as follows:

$$\varepsilon_{\mathrm{LHV}}(i+1) = \varepsilon_{\mathrm{LHV}}(i) \exp\left[\frac{\Delta t}{\tau_{\mathrm{LHV}}(i)}\right] \quad \rightleftarrows \quad i = i+1, \qquad \frac{1}{\tau_{\mathrm{LHV}}(i+1)} = \frac{1}{\Delta t} \ln\left[\frac{\varepsilon_{\mathrm{LHV}}(i+1)}{\varepsilon_{\mathrm{LHV}}(i)}\right]. \tag{172}$$

Table 2: IBS emittance growth after a 10-hour beam coast, considering only the IBS effect

		$\Delta\varepsilon_L/\varepsilon_L$	$\Delta\varepsilon_H/\varepsilon_H$	$\Delta\varepsilon_V/\varepsilon_V$
Case 1	Initial IR triplet: $\beta^*=0.55$ m	16%	9%	−0.0001%
Case 2	IR phase 1 triplet: $\beta^*=0.30$ m	24%	21%	−0.001%
Case 3	Ultimate N_b: $\beta^*=0.25$ m	32%	27%	−0.001%
Case 4	'Highest' ultimate N_b: $\beta^*=0.15$ m	44%	37%	−0.001%

$$\varepsilon_{\mathrm{L,H,V}}(i+1)=\varepsilon_{\mathrm{L,H,V}}(i)\exp\left[\frac{\Delta t}{\tau_{\mathrm{L,H,V}}(i)}\right] \;\rightleftharpoons\; i=i+1 \quad \frac{1}{\tau_{\mathrm{L,H,V}}(i+1)}=\frac{\ln\varepsilon_{\mathrm{L,H,V}}(i+1)-\ln\varepsilon_{\mathrm{L,H,V}}(i)}{\Delta t} \tag{173}$$

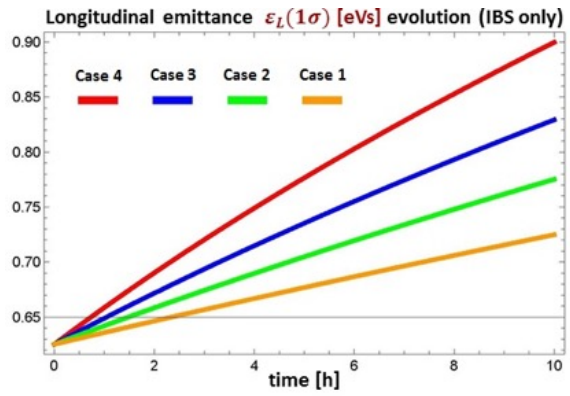


Fig. 19: Evolution of the r.m.s. longitudinal emittance due to IBS effect for the 4 cases of Table 1.

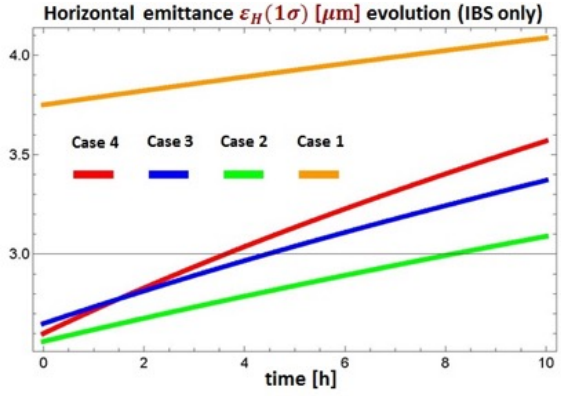


Fig. 20: Evolution of the r.m.s. horizontal emittance due to IBS effect for the 4 cases of Table 1.

4.1.3 Cumulative intrabeam scattering and radiation damping effects in the LHC and SLHC

The *synchrotron radiation* turns into a visible effect for the LHC/SLHC proton beams at 7 TeV collision energy. Emittances shrink with damping times $\tau_{\mathrm{srd_L}}=12.9$ h and $\tau_{\mathrm{srd_{HV}}}=26.0$ h in the transverse planes. Synchrotron radiation damping (SRD) is modelled by replacing $\tau_{\mathrm{LHV}}(i)$ with $(\tau^{-1}_{\mathrm{LHV}}(i)-\tau^{-1}_{\mathrm{srd_{LHV}}})^{-1}$ in Eq. (173).

Figures 21–22 show the evolution of the longitudinal and transverse emittances over a 10-hour beam coast. Synchrotron radiation dominates IBS growth in the longitudinal and vertical planes for all of Cases 1–4; in the horizontal plane the emittance damps continuously during the coast only in Case 1, whereas for Cases 2–4 it expands at some point during the coast. Table 3 displays the increase in beam emittance due to IBS and SRD for a 7 TeV proton beam at the end of 10 hours' storage in the LHC/SLHC.

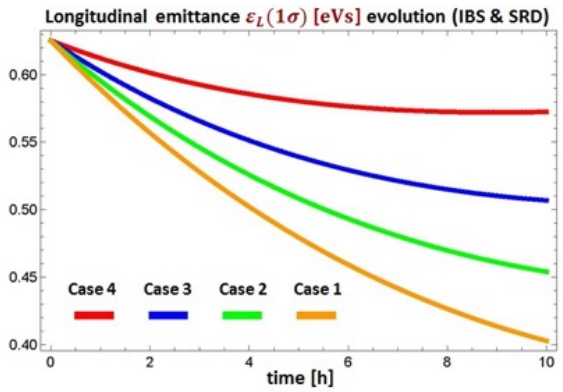


Fig. 21: Evolution of the r.m.s. longitudinal emittance due to IBS and SRD effects for the 4 cases of Table 1.

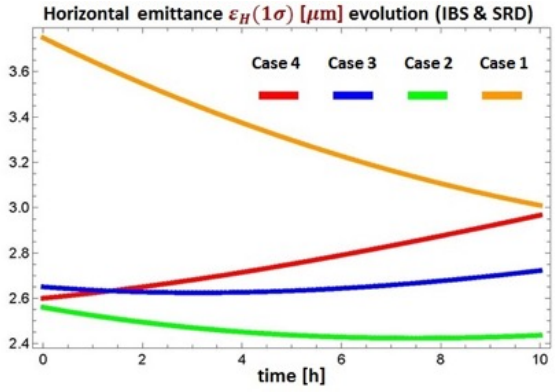


Fig. 22: Evolution of the r.m.s. horizontal emittance due to IBS and SRD effects for the 4 cases of Table 1.

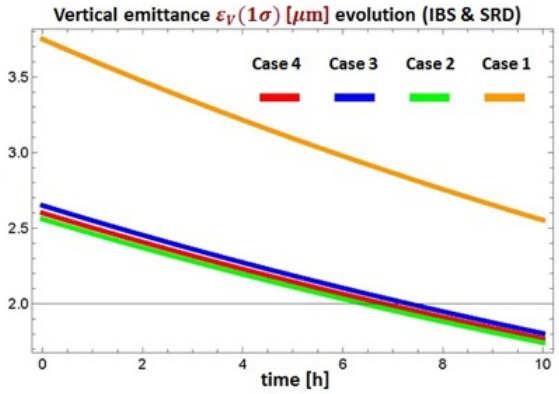


Fig. 23: Evolution of the r.m.s. vertical emittance due to IBS and SRD effects for the 4 cases of Table 1.

Table 3: IBS emittance growth after a 10-hour beam coast, considering both IBS and SRD effects

		$\Delta\varepsilon_L/\varepsilon_L$	$\Delta\varepsilon_H/\varepsilon_H$	$\Delta\varepsilon_V/\varepsilon_V$
Case 1	Initial IR triplet: $\beta^* = 0.55$ m	-36%	-20%	-32%
Case 2	IR phase 1 triplet: $\beta^* = 0.30$ m	-27%	-5%	-32%
Case 3	Ultimate N_b: $\beta^* = 0.25$ m	-19%	3%	-32%
Case 4	'Highest' ultimate N_b: $\beta^* = 0.15$ m	-8%	14%	-32%

In summary, the longitudinal and vertical emittances of all the luminosity scenarios are kept within target specifications for Cases 1–4. Horizontal emittances stay within requirements for Cases 1 and 2; a small blow-up of approximately 3% is anticipated for Case 3, and a larger blow-up of approximately 14% for Case 4. Globally, for most scenarios, the evolution of emittances during the 10-hour coast is kept within the range of design values. Observe that, without the counterbalancing damping effect of the synchrotron radiation, the longitudinal and horizontal emittances would grow continuously during a coast, while the vertical emittance would stay roughly the same. This is because at 7 TeV the LHC is far above the transition energy, as $\gamma = 7461 \gg \gamma_t \approx 53.8$.

4.2 Second example: ELENA at 100 keV

4.2.1 Study of nominal beam parameters and variants

ELENA is a compact hexagonal synchrotron of about 30 m circumference, equipped with an electron cooler. It was designed to allow greater deceleration of the antiprotons ($\bar{p}$) with 5.31 MeV kinetic energy sent by the Antiproton Decelerator (AD), to yield dense beams at 100 keV kinetic energy with a beam population of approximately 2.5×10^7 cooled $\bar{p}$. *Electron cooling* is used to counteract the increase in emittance and momentum due to the deceleration process. The plan is to increase the intensity of the antiprotons delivered to the anti-hydrogen-based experiments at the AD by one to two orders of magnitude [38, 39]. In [38], the joint effects of electron cooling and IBS on the beam equilibrium phase space dimensions are computed using the code BETACOOL [3].

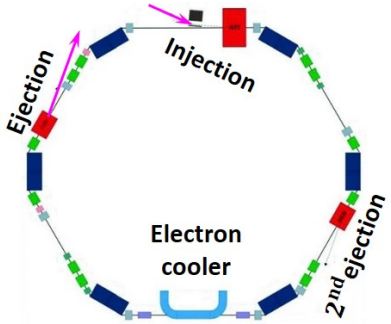


Fig. 24: The ELENA (Extra Low Energy Antiproton) ring (8.6 m×10.0 m, ≈30 m circumference) is a below-transition-energy ring with $\gamma = 1.00001 < \gamma_t \approx 1.9$.

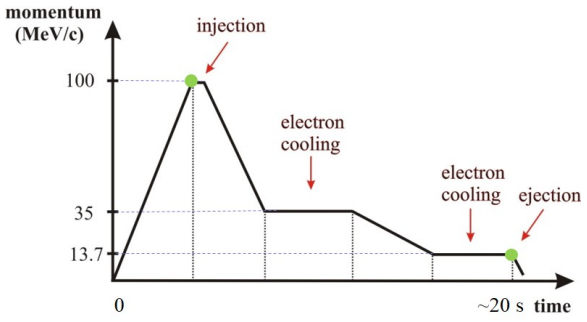


Fig. 25: ELENA cycle, with first injection plateau at 100 MeV/c, second cooling plateau at 35 MeV/c and third cooling plateau at 13.7 MeV/c (100 keV).

The cycle has one injection plateau at 100 MeV/c momentum for the *coasting* $\bar{p}$ beam injection from the AD. Hence, the beam is decelerated down to a first plateau at 35 MeV/c for a cooling period of about 8 s, followed by deceleration down to a second plateau at 13.7 MeV/c for about 3 s extra cooling of the $\bar{p}$ beam. Before the end of the plateau, the beam is *bunched* and further cooled for approximately 0.3 s to the emittances needed for ELENA experiments, up to the extraction of four bunches consisting of approximately 6.25×10^6 antiprotons to the transfer lines between ELENA and the experimental areas.

Figures 24 and 25 show schematically the ELENA ring and its deceleration cycle.

Figure 26 displays the optics of the ELENA ring lattice with working points $Q_H = 2.3$ and $Q_V = 1.3$ (P. Belochitskii, 2012). We see that, unlike the LHC lattice (Fig. 16), ELENA is not a smooth lattice.

Table 4 shows the initial length, momentum spread, and longitudinal and transverse emittance values of the nominal bunch and five variants (prior to IBS), used to evaluate how the bunch parameters behave under IBS effects. This is an important issue, as the bunch length and momentum spread must be kept close to their nominal values. This study was undertaken in 2012.

Table 4: Initial nominal beam emittances on the 100 keV plateau and five variants. The longitudinal emittance is defined as $\varepsilon_L = \pi p \sigma_{BL} \sigma_{\Delta p/p} (\beta c)^{-1}$, where c is the speed of light and σ_{BL} is the r.m.s. bunch length.

	σ_{BL}	$\sigma_{\Delta p/p}$	ε_L^{rms}	ε_{HV}^{rms}
Nominal	0.325 m	7.5×10^{-5}	2.4×10^{-4} eV s	1.0 μm
Variant 1	0.325 m	2.5×10^{-5}	0.8×10^{-4} eV s	0.5 μm
Variant 2	0.325 m	1.25×10^{-4}	4.0×10^{-4} eV s	2.5 μm
Variant 3	0.325 m	2.5×10^{-4}	8.0×10^{-4} eV s	1.0 μm
Variant 4	0.325 m	3.75×10^{-4}	12.0×10^{-4} eV s	1.0 μm
Variant 5	0.325 m	5.0×10^{-4}	16.0×10^{-4} eV s	1.0 μm

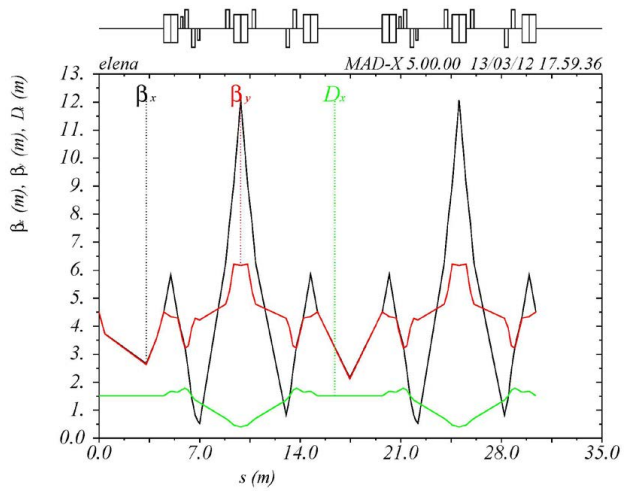


Fig. 26: ELENA ring betatron functions and dispersion function in [m]

4.2.2 Intrabeam scattering simulations for bunched beams on the 100 keV *plateau*

The IBS simulations for ELENA were performed with two different *Mathematica* notebooks for cross-checking. The first consists of the code used for LHC/SLHC simulations. The code in the second notebook is based on an algorithm derived by A. Piwinski [6] and implemented by Ch. Carli [7], which takes the linear coupling into account exactly. The lattice is described in terms of magnets and their strengths, and the program analyses linear properties such as the betatron-oscillation coupling (solenoid and skew quadrupole magnets). The process is applied to the generalized emittances specified via the betatron-oscillation eigenvectors (e.g. as calculated by MADX). Then this information is used to compute IBS growth rates. Simulations using the second code show that the coupling effect has negligible influence on the IBS growth rates and hence on the vertical dispersion function.

Figures 27–29 plot, for bunched beams, the longitudinal τ_{L}^{-1}, horizontal τ_{H}^{-1} and vertical τ_{V}^{-1} growth rates 'without cooling', as functions of the transverse emittance $\varepsilon_{\mathrm{HV}}^{\mathrm{rms}}$ and the relative momentum spread $\sigma_{\Delta\mathrm{p/p}}$. Each bunch contains 6.25×10^6 antiprotons. The plotted growth rates are computed using an initial r.m.s. bunch length of $\sigma_{\mathrm{BL}} = 0.325\,\mathrm{m}$. Alternative plots of τ_{LHV}^{-1} versus $\varepsilon_{\mathrm{H}}^{\mathrm{rms}}$ and $\varepsilon_{\mathrm{V}}^{\mathrm{rms}}$, with nominal $\sigma_{\Delta\mathrm{p/p}} = 7.5 \times 10^{-5}$ and $\sigma_{\mathrm{BL}} = 0.325\,\mathrm{m}$, look similar and so are not shown.

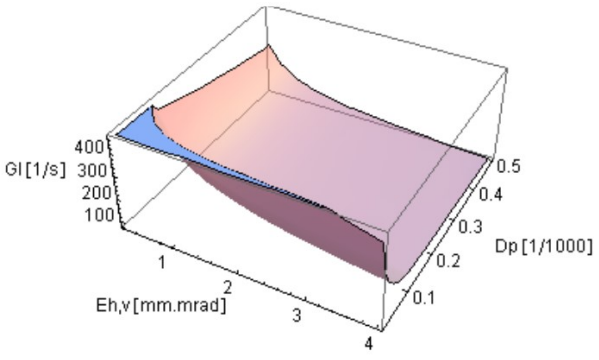


Fig. 27: Longitudinal growth rate τ_{L}^{-1} ($\mathrm{G_L}$) as $\varepsilon_{\mathrm{H,V}}$ and $\sigma_{\Delta\mathrm{p/p}}$ are varied, with $\varepsilon_{\mathrm{H}} \equiv \varepsilon_{\mathrm{V}}$ and $\sigma_{\mathrm{BL}} = 0.325\,\mathrm{m}$.

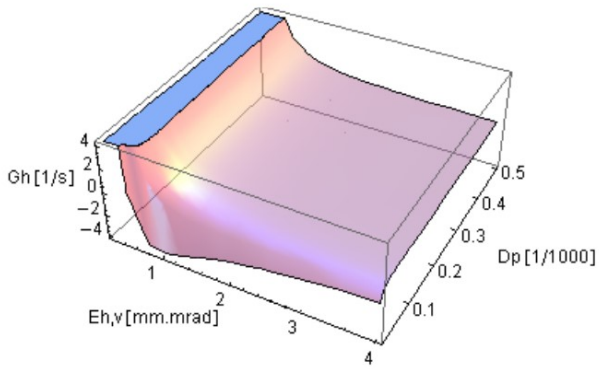


Fig. 28: Horizontal growth rate τ_{H}^{-1} ($\mathrm{G_H}$) as $\varepsilon_{\mathrm{H,V}}$ and $\sigma_{\Delta\mathrm{p/p}}$ are varied, with $\varepsilon_{\mathrm{H}} \equiv \varepsilon_{\mathrm{V}}$ and $\sigma_{\mathrm{BL}} = 0.325\,\mathrm{m}$.

Figures 30 and 31 plot the transverse and longitudinal growth times τ_{LHV} 'without cooling' for bunched beams (where each bunch contains 6.25×10^6 antiprotons). The simulation was performed for the nominal beam and the first two variants given in Table 4.

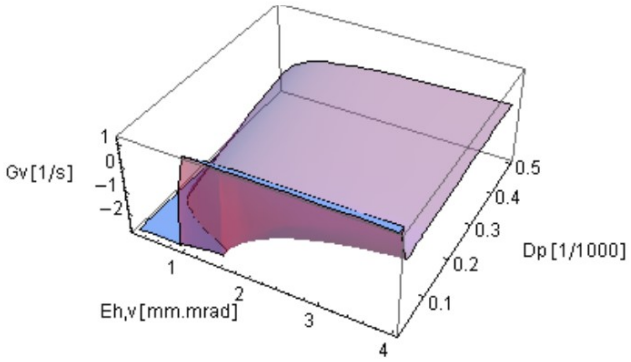


Fig. 29: Vertical growth rate τ_V^{-1} (G_V) as $\varepsilon_{H,V}$ and $\sigma_{\Delta p/p}$ are varied, with $\varepsilon_H \equiv \varepsilon_V$ and $\sigma_{BL} = 0.325$ m

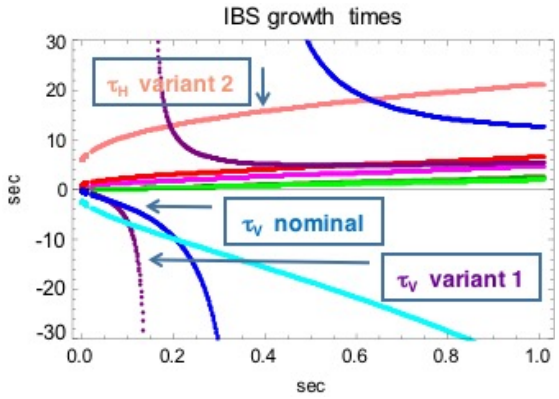


Fig. 30: Evolution of the transverse IBS growth times τ_{HV} (on the linear scale) over 1 s ($\varepsilon_L = \pi p \sigma_{BL} (\beta c)^{-1}$).

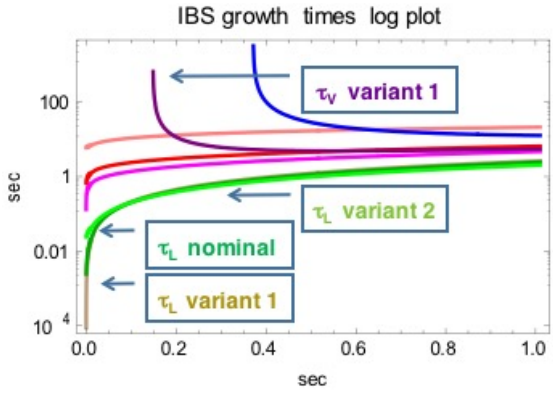


Fig. 31: Evolution of the longitudinal and vertical IBS growth times τ_{LV} (on the logarithmic scale) over 1 s.

4.2.3 Performance of nominal beam and variants for bunched beams on the 100 keV *plateau*

Figures 32 and 33 plot the transverse emittance ε_{HV} 'without cooling' for the nominal beam and the five variants given in Table 4.

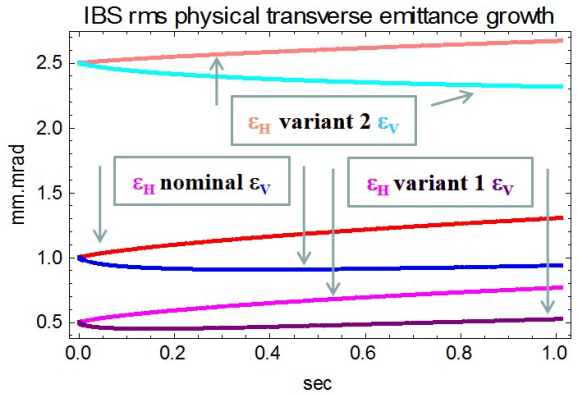


Fig. 32: Evolution of the emittances ε_{HV} over 1 s for the nominal beam and the first two variants in Table 4.

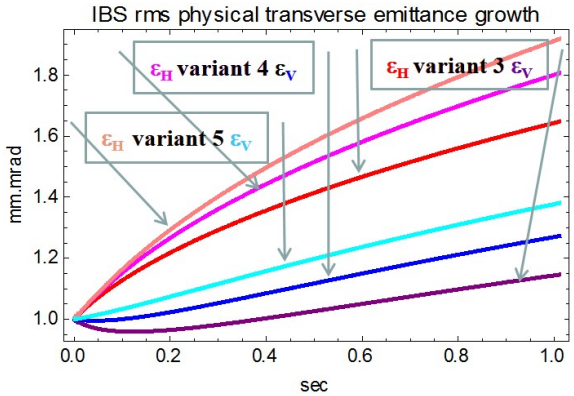


Fig. 33: Evolution of the transverse emittances ε_{HV} over 1 s for variants 3–5 in Table 4.

Figure 34 plots the relative momentum spread $\sigma_{\Delta p/p}$ with IBS effects and without cooling for the nominal beam and the first two variants. Figure 35 plots the bunch length σ_{BL} and the momentum spread

$\sigma_{\Delta p/p}$ for variants 3–5. The bunch length is given in [m] and the momentum spread in [permille], with both units shown on the same vertical scale (e.g. $\sigma_{BL}(1) = 0.42$ m, $\sigma_{\Delta p/p}(1) = 0.65$ permille for variant 5).

Finally, Fig. 36 displays the bunch length σ_{BL} [m] for the nominal beam and the first two variants.

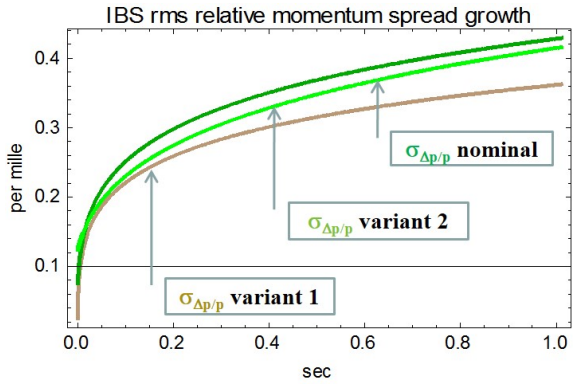


Fig. 34: Evolution of the momentum spread $\sigma_{\Delta p/p}$ over 1 s for the nominal beam and for variants 1–2.

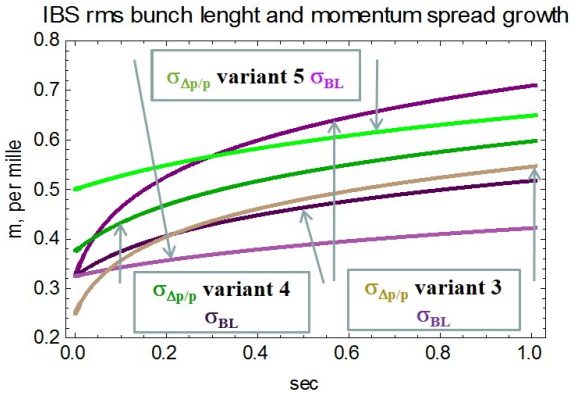


Fig. 35: Evolution of the bunch length σ_{BL} and relative momentum spread $\sigma_{\Delta p/p}$ over 1 s for variants 3–5.

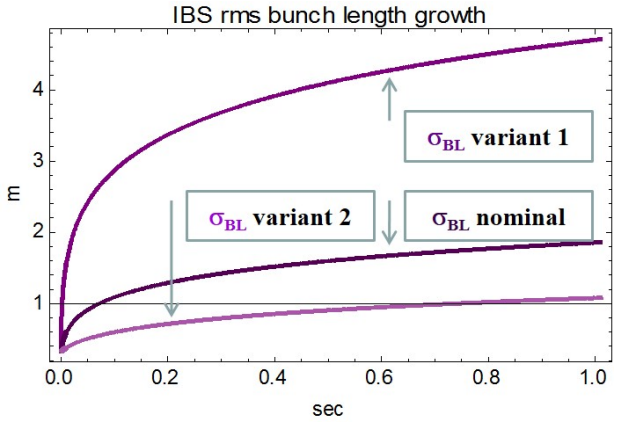


Fig. 36: Evolution of the bunch length σ_{BL} over 1 s for the nominal beam and the first two variants

From Figs. 32–36 we can see that an equilibrium between the transverse and longitudinal emittances does not strictly take place, even though ELENA at 100 keV is below the transition energy, with $\gamma \approx 1 < \gamma_t \approx 1.9$. Unlike the LHC collider, ELENA is a compact decelerator ring made of a non-regular lattice (see Fig. 26) and so does not satisfy the smooth focusing approximation criteria. Table 5 shows the bunch length, relative momentum spread and emittance increase for the nominal bunch and the five variants under IBS effects for 1 s at 100 keV without cooling. This allows us to quantify the bunch parameters with IBS, which is important because bunch length and momentum spread must be kept close to their nominal values.

Table 5: IBS beam growth ratios for the nominal beam and five variants; the ratios are of parameters experiencing the IBS effect for 1 s on the 100 keV plateau without cooling to those same parameters not enduring IBS effects.

	$\sigma_{\mathrm{BL}}(t)/\sigma_{\mathrm{BL}}(0)$	$\varepsilon_{\mathrm{L}}^{\mathrm{rms}}(t)/\varepsilon_{\mathrm{L}}^{\mathrm{rms}}(0)$	$\varepsilon_{\mathrm{H}}^{\mathrm{rms}}(t)/\varepsilon_{\mathrm{H}}^{\mathrm{rms}}(0)$	$\varepsilon_{\mathrm{V}}^{\mathrm{rms}}(t)/\varepsilon_{\mathrm{V}}^{\mathrm{rms}}(0)$
Nominal	5.7	32.5	1.31	0.94
Variant 1	14.5[a]	205.1	1.25	1.05
Variant 2	3.3	11.0	1.07	0.93
Variant 3	2.19	4.78	1.65	1.15
Variant 4	1.59	2.54	1.81	1.27
Variant 5	1.30	1.69	1.92	1.38

[a]Variant 1 is the worst scenario as the bunch length grows by a factor of 14.5 under IBS effects in relation to a bunch without IBS. Likewise, the nominal beam experiences a 5.7 growth factor due to IBS effects. The best scenario is variant 5, as the bunch grows by only a small factor of 1.3, compatible to that of the bunch length before IBS starts.

Figure 37 displays the IBS beam parameter growth factor ratios for the nominal beam and the five variants in Table 5. Compared to the longitudinal emittance, the transverse beam emittances are not as significantly altered by the IBS effects.

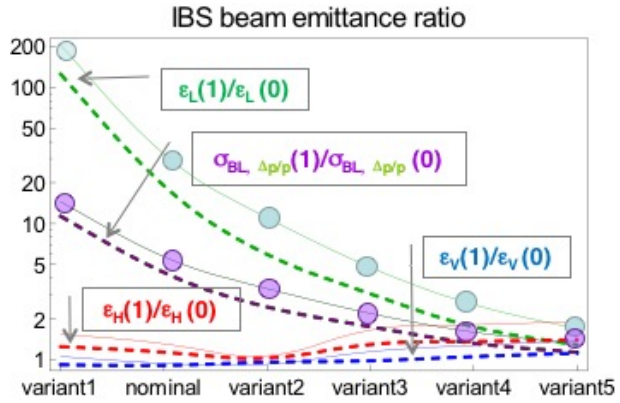


Fig. 37: Longitudinal and transverse emittance growth ratios, $\varepsilon_{\mathrm{LHV}}(t)/\varepsilon_{\mathrm{LHV}}(0)$, and the r.m.s. bunch length growth ratio $\sigma_{\mathrm{BL}}(t)/\sigma_{\mathrm{BL}}(0)$ (equal to that of the relative momentum spread $\sigma_{\Delta_{\mathrm{p/p}}}$) plotted for the nominal beam and the five variants. The growth ratios are shown for IBS simulations over 1 s (large circles represent IBS raw simulation data, while continuous and dashed lines are interpolated).

To finish, let us summarize our observations for the nominal beam and variants 1 and 5 in Table 5.

- **Nominal:** The increase in bunch length and relative momentum spread after the beam circulates for 1 s on the 100 keV plateau is *large*; the r.m.s. bunch length increases to $\sigma_{\mathrm{BL}}(1) = 1.85\,\mathrm{m}$ compared with $\sigma_{\mathrm{BL}}(0) = 0.325\,\mathrm{m}$ before IBS begins (also $\sigma_{\Delta_{\mathrm{p/p}}}(1) = 0.04\%$). This corresponds to a 95% bunch length increase equal to 7.4 m ($\approx 4\sigma_{\mathrm{BL}}(0)$), instead of the nominal 95% bunch length of $1.3\,\mathrm{m}$ at $t = 0$ without IBS effects. The longitudinal emittance varies as the product of the bunch length and the momentum spread since $\varepsilon_{\mathrm{L}} = \pi p \sigma_{\mathrm{BL}} \sigma_{\Delta_{\mathrm{p/p}}} (\beta c)^{-1}$.
- **Variant 1:** The bunch length and momentum spread after 1 s are *huge*. Indeed, the r.m.s. bunch length increases to $\sigma_{\mathrm{BL}}(1\,\mathrm{s}) = 4.7\,\mathrm{m}$, while the 95% bunch length peaks at 18.8 m!
- **Variant 5:** This is the best scenario, because the bunch length and momentum spread will suffer only 30% blow-up due to IBS after circulating for 1 s on the 100 keV plateau; their r.m.s. values are $\sigma_{\mathrm{BL}}(1) = 0.42\,\mathrm{m}$, compared with 0.325 m before IBS starts ($\sigma_{\Delta_{\mathrm{p/p}}}(1) = 0.065\%$). Similarly, the 95% bunch length is 1.7 m.

Appendix A: Complementary proof of growth rate formula

The Piwinski's model [1] derives the formulae for the variations of the mean transverse emittances and momentum spread per time unit due to a scattering event. The derivative of the betatron and dispersion functions and the vertical dispersion function with respect to the longitudinal beam axis are neglected. The growth rates are calculated in accordance with the expression (154) for Gaussian probability law, derived from (147). Here, the aim to show clearly as possible the link between these formulations.

Let us define the functions

$$f_i = 2\,k_i \int_0^{\pi} \mathrm{d}\mu \int_0^{2\pi} \mathrm{d}\nu \int_0^{\infty} \mathrm{d}r\, \sin[\mu]\, g_i[\mu,\nu] \log[r] \exp\Big[-r\, D(\mu,\nu)\Big] \quad \text{for} \quad i = 1,2,3\,. \tag{A1}$$

where $k_1 = \dfrac{1}{c^2}$, $k_2 = \dfrac{a^2}{c^2}$, $k_3 = \dfrac{b^2}{c^2}$, and

$$D(\mu,\nu) = \frac{1}{c^2}\Big(b^2\cos^2[\mu] + \sin^2[\mu]\big(\cos^2[\nu] + a^2\sin^2[\nu]\big)\Big)\,.$$

$D(\mu,\nu)$ is Eq. (145) and $g_1[\mu,\nu]$, $g_2[\mu,\nu]$, $g_3[\mu,\nu]$ are Eqs. (146). Now, we pass from the spherical to Cartesian coordinates $(\sqrt{r},\mu,\nu) \to (u,v,w)$ using (143)

$$\big(u = \sqrt{r}\sin\mu\cos\nu,\ v = \sqrt{r}\sin\mu\sin\nu,\ w = \sqrt{r}\cos\mu\big) \implies u^2+v^2+w^2 = r\,,$$

from which we can write the three functions f_i as follows, since the functions $g_i(\mu,\nu)$ and $D(u,v,w)$ transform into

$$g_1 = \frac{1}{r}(-2u^2+v^2+w^2)\,, \quad g_2 = \frac{1}{r}(u^2-2v^2+w^2)\,, \quad g_3 = \frac{1}{r}(u^2+v^2-2w^2)\,, \tag{A2}$$
$$D(u,v,w) = \frac{1}{r}(u^2-a^2\,v^2+b^2\,w^2)\,, \quad \frac{2}{\sqrt{r}}\,du\,dv\,dw = \sin[\mu]\,\mathrm{d}r\,\mathrm{d}\mu\,\mathrm{d}\nu$$

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = 4 \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \int_{-\infty}^{\infty}\mathrm{d}u \int_{-\infty}^{\infty}\mathrm{d}v \int_{-\infty}^{\infty}\mathrm{d}w \begin{Bmatrix} -2u^2+v^2+w^2 \\ u^2-2v^2+w^2 \\ u^2+v^2-2w^2 \end{Bmatrix}$$
$$\times\, \frac{\ln[u^2+v^2+w^2]}{\big(u^2+v^2+w^2\big)^{3/2}} \exp\left[-\frac{1}{c^2}\big(u^2+a^2v^2+b^2w^2\big)\right]\,, \tag{A3}$$

Next, a trick is used: In f_1 let $u = w, v = u, w = v$, in f_2 let $u = u, v = w, w = v$, do not change f_3 (i.e. $u = u, v = v, w = w$ remain the same). We obtain (cf. [24])

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \frac{4}{c^2} \begin{bmatrix} 1 \\ a^2 \\ b^2 \end{bmatrix} \int_{-\infty}^{\infty}\mathrm{d}u \int_{-\infty}^{\infty}\mathrm{d}v \int_{-\infty}^{\infty}\mathrm{d}w\, \big\{u^2+v^2-2w^2\big\}$$
$$\times\, \frac{\ln[u^2+v^2+w^2]}{\big(u^2+v^2+w^2\big)^{3/2}} \begin{Bmatrix} \exp\left[-\frac{1}{c^2}\big(a^2u^2+b^2v^2+w^2\big)\right] \\ \exp\left[-\frac{1}{c^2}\big(u^2+b^2v^2+a^2w^2\big)\right] \\ \exp\left[-\frac{1}{c^2}\big(u^2+a^2v^2+b^2w^2\big)\right] \end{Bmatrix} \mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,, \tag{A4}$$

Switching back to spherical coordinates leads to, with

$$\frac{u^2+v^2-2w^2}{r^{3/2}} = \frac{1-3\cos^2[\mu]}{r^{1/2}}\,, \quad \text{reintroducing} \quad \rho = \frac{r}{c^2} \quad \Longrightarrow \quad \frac{2}{\sqrt{r}}\,du\,dv\,dw = c^2\sin[\mu]\,\mathrm{d}\rho\,\mathrm{d}\mu\,\mathrm{d}\nu\,,$$

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = 2\int_0^{\pi} \mathrm{d}\mu \int_0^{2\pi} \mathrm{d}\nu \int_0^{\infty} \mathrm{d}\rho\, \sin[\mu]\, \left(1-3\cos^2[\mu]\right) \ln[\rho c^2]$$
$$\times \begin{Bmatrix} \exp\left[-\rho\left(\sin^2[\mu]\left(a^2\cos^2[\nu]+b^2\sin^2[\nu]\right)+\cos^2[\mu]\right]\right. \\ a^2\exp\left[-\rho\left(\sin^2[\mu]\left(\cos^2[\nu]+b^2\sin^2[\nu]\right)+a^2\cos^2[\mu]\right]\right. \\ b^2\exp\left[-\rho\left(\sin^2[\mu]\left(\cos^2[\nu]+a^2\sin^2[\nu]\right)+b^2\cos^2[\mu]\right]\right. \end{Bmatrix} \equiv \begin{pmatrix} f\left(a,b,c\right) \\ f\left(\frac{1}{a},\frac{b}{a},\frac{c}{a}\right) \\ f\left(\frac{1}{b},\frac{a}{b},\frac{c}{b}\right) \end{pmatrix} . \qquad \text{(A5)}$$

in which $f\left(a,b,c\right)$ is the scattering matrix (148) and $D_0(\mu,\nu)$ is (149). The first equation of (A5) can be written

$$f_1 = 2\int_0^{\pi} \mathrm{d}\mu \int_0^{2\pi} \mathrm{d}\nu \int_0^{\infty} \mathrm{d}\rho\, \sin[\mu]\, \left(1-3\cos^2[\mu]\right) \ln[\rho c^2] \exp\left[-\rho\, D_0(\mu,\nu)\right] \equiv f\left(a,b,c\right) ,$$

Referring to Eqs. (147) the IBS growth rates can then given in terms of f_1, f_2, f_3 as

$$\begin{pmatrix} \tau_\eta^{-1} \\ \tau_x^{-1} \\ \tau_z^{-1} \end{pmatrix} = \frac{\mathcal{A}_\mathrm{P}}{c^2} \int_0^{\infty} \mathrm{d}\rho \int_0^{\pi} \mathrm{d}\mu \int_0^{2\pi} \mathrm{d}\nu\, \sin[\mu] \exp\left[-\rho D(\mu,\nu)\right] \ln[\rho c^2]$$
$$\times \begin{Bmatrix} \left(1-d^2\right) g_1[\mu,\nu] \\ a^2 g_2[\mu,\nu] + d^2 g_1[\mu,\nu] \\ b^2 g_3[\mu,\nu]) \end{Bmatrix} \equiv \mathcal{A}_\mathrm{P} \begin{bmatrix} \left(1-d^2\right) f_1 \\ f_2 + d^2 f_1 \\ f_3 \end{bmatrix} \equiv \mathcal{A}_\mathrm{P} \begin{bmatrix} \left(1-d^2\right) f\left(a,b,c\right) \\ f\left(\frac{1}{a},\frac{b}{a},\frac{c}{a}\right) + d^2 f\left(a,b,c\right) \\ f\left(\frac{1}{b},\frac{a}{b},\frac{c}{b}\right) \end{bmatrix} . \qquad \text{(A6)}$$

where $f\left(a,b,c\right) = f_1,\ f\left(\frac{1}{a},\frac{b}{a},\frac{c}{a}\right) = f_2,\ f\left(\frac{1}{b},\frac{a}{b},\frac{c}{b}\right) = f_3$. Eq. (A6) is thus equal to Eq. (154).

Appendix B: Scattering matrix

Decay rates and *cross-sections* are experimentally measurable quantities, predicted in quantum physics through probabilities given by squared inner products of particle quantum states. These inner products can be written as matrix elements $\langle f;t_f|\, i;t_i\rangle$ where $|i;t_i\rangle$ is the *initial* state before the scattering of a physical system and $|f;t_f\rangle$ is the *final* state after the scattering event (cf. [35, 40–42]). The notation $\langle f;t_f|\, i;t_i\rangle$ refers to the so-called *Schrödinger picture* (SP), where the states $|\psi\rangle_\mathrm{S}$ evolve in time while the operators $\mathcal{O}_\mathrm{S}$ are independent of time. In contrast, in the *Heisenberg picture* (HP), the states $|\psi\rangle_\mathrm{H}$ are fixed and the operators $\mathcal{O}_\mathrm{H}(t)$ vary in time:

$$\begin{aligned} &\text{SP:} \quad \mathrm{i}\frac{\mathrm{d}|\psi\rangle_\mathrm{S}}{\mathrm{d}t} = H|\psi\rangle_\mathrm{S} \text{ and } \mathcal{O}_\mathrm{S} \text{ is time-independent} \\ &\text{HP:} \quad |\psi\rangle_\mathrm{H} = \exp(\mathrm{i}Ht)|\psi\rangle_\mathrm{S} \text{ and } \mathcal{O}_\mathrm{H}(t) = \exp(\mathrm{i}Ht)\mathcal{O}_\mathrm{S}\exp(-\mathrm{i}Ht) \end{aligned}$$

(where HL units, with $\epsilon_0 = \hbar = c = 1$, are used). The *interaction picture* (IP) is a hybrid of SP and HP. The Hamiltonian H is split up into a part H_0 that can be handled exactly and a part H_int to be treated as a perturbation: $H = H_0 + H_\mathrm{int}$. Thus, the time dependence of operators $\mathcal{O}_\mathrm{I}$ is driven by H_0 and the time dependence of states $|\psi\rangle_\mathrm{I}$ by H_int so that the corresponding IP state vector and operator are defined by

$$|\psi(t)\rangle_\mathrm{I} = \exp(\mathrm{i}H_0 t)\, |\psi(t)\rangle_\mathrm{S} \ \text{ and } \ \mathcal{O}_\mathrm{I}(t) = \exp(\mathrm{i}H_0 t)\, \mathcal{O}_\mathrm{S} \exp(-\mathrm{i}H_0 t)$$

$$\overset{\mathcal{O}=H_{\text{int}}}{\Longrightarrow} \quad H_{\text{I}} \equiv (H_{\text{int}})_{\text{I}} = \exp(\mathrm{i}H_0 t)(H_{\text{int}})_{\text{S}} \exp(-\mathrm{i}H_0 t) \;\; \text{(IP)} \; .$$

From this we have the equation of motion for $|\psi(t)\rangle_{\text{I}}$,

$$\mathrm{i}\,\frac{\mathrm{d}|\psi(t)\rangle_{\text{I}}}{\mathrm{d}t} = H_I|\psi(t)\rangle_{\text{I}}, \tag{B1}$$

whose solution is given by *Dyson's expansion*:

$$|\psi(t)\rangle_{\text{I}} = U(t,t_0)|\psi(t_0)\rangle_{\text{I}} \;\; \Longrightarrow \;\; U(t,t_0) = \mathcal{T}\left\{\exp\left[-\mathrm{i}\int_{t_0}^{t} H_{\text{I}}(t')\,\mathrm{d}t'\right]\right\} \tag{B2}$$

with $\mathrm{i} = \sqrt{-1}$, where $U(t,t_0)$ is a time-evolution operator satisfying the properties $U(t,t) = 1$ and $U(t_1,t_2)U(t_2,t_3) = U(t_1,t_3)$. The function $\mathcal{T}\{\,\cdot\,\}$ implies a *time-ordering product* acting on operators such that operators evaluated at later times are placed to the left.

Comment: For two operators $\mathcal{O}_{1,2}$ the function $\mathcal{T}$ is defined by $\mathcal{T}(\mathcal{O}_1(t_1)\mathcal{O}_2(t_2)) = \mathcal{O}_1(t_1)\mathcal{O}_2(t_2)$ if $t_1 > t_2$ and $\mathcal{T}(\mathcal{O}_1(t_1)\mathcal{O}_2(t_2)) = \mathcal{O}_2(t_2)\mathcal{O}_1(t_1)$ if $t_2 > t_1$.

In the case where the particle momentum eigenstates evolve from $t = -\infty$ to $t = +\infty$, the time-evolution operator is denoted by S and called the *S-matrix* (for 'scattering matrix'). It is defined via its components S_{fi} as

$$\langle f|S|i\rangle = \langle f; t=\infty|\, i; t=-\infty\rangle_{\text{Schrödinger}} \equiv S_{\text{fi}} \; . \tag{B3}$$

The S-matrix can equivalently be defined in terms of the time-evolution operator describing scattering experiments, as

$$\langle f|S|i\rangle = \lim_{\substack{t_0\to-\infty \\ t\to+\infty}} \langle f|U(t,t_0)|i\rangle \equiv S_{\text{fi}} \; . \tag{B4}$$

The idea behind this definition is that as time evolves, the particles approach each other and may interact briefly before moving away, with each going along its own path so that the states are free of interactions at $t = \pm\infty$. For that reason it is convenient to separate forward scattering by defining $S = I + \mathrm{i}T$, where I is the identity matrix and T is a *transition matrix* (called the T-matrix) which contributes only when particles undergo scattering. The exponential form of Dyson's expansion, the formula on the right in (B2), is not very useful as the integral in the exponent cannot in general be computed exactly. Therefore, provided H_{I} is small compared with H_0, it should be expanded in a power series of the interaction 'Hamiltonian operator', and thus terms can be obtained to a suitable order since (B2) is related to (B4). Since the S-matrix is the limit $S = U(t\to\infty,\, t_0\to-\infty)$, the Dyson's expansion power series of (B2) can be written as

$$S = I - \mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}t_1 H_{\text{I}}(t_1) + \frac{(-\mathrm{i})^2}{2}\int_{-\infty}^{\infty}\mathrm{d}t_1\int_{-\infty}^{t_1}\mathrm{d}t_2\mathcal{T}\big\{H_{\text{I}}(t_2)H_{\text{I}}(t_1)\big\} + \cdots \; . \tag{B5}$$

However, this approach is rather tedious. Instead, it is better to use the graphical tool of *Feynman diagrams* (together with *Feynman rules*) discussed in the main text. Briefly, Feynman diagrams are qualitative and symbolic figures that represent terms in the perturbation expansion of the S-matrix (see e.g. Fig. 14). Particles in space–time are solid lines with arrows, time runs from left to right, the space direction is perpendicular to time (antiparticles travel backwards in time), arrows show the charge flux relative to time, and wavy lines represent 'virtual particles' that live for only a short time, i.e. bosons mediating the interaction which are emitted and absorbed soon after (e.g. photons in QED). 'Loops' are closed patterns of virtual particles (present in high-order diagrams corresponding to high-order terms of the perturbative Dyson's expansion power series).

In the S-matrix, we are interested in the scattering part of a process and not in the part where no scattering occurs, namely where S reduces to an identity matrix I. Therefore, we just want to calculate $\langle f|S-I|i\rangle$, the 'non-trivial' part of $\langle f|S|i\rangle$. Here we follow [35, 40], where it is emphasized that the matrix elements S_{fi} of the S-matrix should reflect 4-momentum conservation, so that S, or $S-1$, should contain a factor $(2\pi)^4\delta^4(p_i-p_f)$. It is then conventional to extract this factor and define the *invariant amplitude* $\mathcal{M}$ via

$$\langle f|(\mathrm{i}T)|i\rangle \equiv \langle f|S-I|i\rangle = (2\pi)^4\delta^4\Big(\sum p_i - \sum p_f\Big)\,\mathrm{i}\mathcal{M} \tag{B6}$$

where $\sum p_f$ and $\sum p_i$ are the sums of the final and initial 4-momenta, and $\mathrm{i}=\sqrt{-1}$, the complex number i is introduced to motivate the relation $\exp[\mathrm{i}T]\approx I+\mathrm{i}T=S$. The reason the δ-function is included in the S-matrix is to ensure observance of conservation of energy and momentum; so it is useful to factor an overall 'momentum-conserving δ-function' to begin with. Detailed proofs of (B6), which are quite difficult, are given in [43] for non-relativistic scattering and in [44] for the relativistic case. Equation (B6) differs from the expressions in those two proofs by its 2π factor. For a two-particle energy–momentum scattering process, this equation reads

$$\langle f_1f_2|\mathrm{i}T|i_1i_2\rangle \equiv \langle f_1f_2|S-I|i_1i_2\rangle = (2\pi)^4\delta^4(p_{1i}+p_{2i}-p_{1f}-p_{2f})\,\mathrm{i}\mathcal{M}, \tag{B7}$$

also written

$$\langle p_{1f}p_{2f}|\mathrm{i}T|p_{1i}p_{2i}\rangle \equiv \langle p_{1f}p_{2f}|S-I|p_{1i}p_{2i}\rangle = (2\pi)^4\delta^4(p_{1i}+p_{2i}-p_{1f}-p_{2f})\,\mathrm{i}\mathcal{M}\,,$$

where we have used the notation $|i\rangle=|i_1\rangle|i_2\rangle=|p_{1i}\rangle|p_{2i}\rangle$ and $|f\rangle=|f_1\rangle|f_2\rangle=|p_{1f}\rangle|p_{2f}\rangle$ for the two-particle state.

Starting from (B6), the transition probability between the initial state $|i\rangle$ and the final state $|f\rangle$ is given by the square of the S-matrix, $|\langle f|S-I|i\rangle|^2$. As is usual in quantum mechanics and quantum field theory, the probabilities of events are the squared (moduli of the) 'quantum amplitudes', such as the entries of the S-matrix for calculating the probabilities of decay rates and cross-section processes. However, to square the matrix elements we have to square the δ-function in (B6). Using the Dirac δ-function property $\{\delta(x-y)\}^2=\delta(x-y)\delta(0)$, we obtain

$$\Big\{\delta^4\Big(\sum p_i-\sum p_f\Big)\Big\}^2 = \delta^4\Big(\sum p_i-\sum p_f\Big)\delta^4(0)=\infty \quad \text{since} \quad \delta(0)=\lim_{\epsilon\to 0}\frac{1}{\epsilon}=\infty\,,$$

which is meaningless! Nevertheless, such infinities must be eliminated from all physical quantities.

One proper way to solve this problem is to confine the whole system of particles in a closed box of volume V and then let the box volume become infinite, while assuming that the interaction is turned on for only a finite time t. For simplicity the particles are described as 'free-particle plane waves' that spread over infinite volumes and time intervals. Such a finite closed box and finite time interval can be defined using the following representation of a Dirac δ-function on a 4-position $r\stackrel{\mathrm{def}}{=}(t,\mathbf{r})$ and a 4-momentum $p\stackrel{\mathrm{def}}{=}(E,\mathbf{p})$. We find that

$$\begin{aligned}\delta^4(p) &= \frac{1}{(2\pi)^4}\int \exp[-\mathrm{i}rp]\,\mathrm{d}^3r \\ &= \frac{1}{(2\pi)^4}\int_V \exp[-\mathrm{i}\mathbf{r}\cdot\mathbf{p}]\,\mathrm{d}^3\mathbf{r}\int_0^t \exp[-\mathrm{i}t'E]\,\mathrm{d}t' \stackrel{\mathrm{p=E=0}}{\Longrightarrow} \delta^4(0)=\frac{Vt}{(2\pi)^4}\,.\end{aligned}$$

Taking the square of (B6) gives, by virtue of the last two formulae,

$$\mathcal{P}_{i\to f}\stackrel{?}{=}|\langle f|(\mathrm{i}T)|i\rangle|^2 \equiv |\langle f|S-I|i\rangle|^2 = Vt(2\pi)^4\delta^4\Big(\sum p_i-\sum p_f\Big)|\mathcal{M}|^2\,. \tag{B8}$$

Although the probabilities are expressed in terms of the squared amplitudes, we shall proceed differently and replace $|\langle f|(\mathrm{i}T)|i\rangle|^2 \equiv |\langle f|S-I|i\rangle|^2$ in (B8) by the 'normalized' probability

$$\mathcal{P}_{i\to f} = \frac{|\langle f|(\mathrm{i}T)|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle} \equiv \frac{|\langle f|(S-I)|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle}, \tag{B9}$$

where the presence of inner products $\langle f|f\rangle$ and $\langle i|i\rangle$ in the denominator of (B9) stems from the fact that they may not be normalized to unity.

As a 'simple' case we envisage (in a toy model framework of structureless, i.e. point-like and spineless, particles) the *decay* of a single initial particle $|i\rangle = |i_1\rangle$, with 4-momentum $|p_{1i}\rangle$, into two distinct final particles $|f\rangle = |f_1\rangle|f_2\rangle$ with 4-momenta $|p_{1i}\rangle|p_{2i}\rangle$ [42]. These inner products can be shown (without proof) to be

$$\langle i|i\rangle \equiv \langle i_1|i_1\rangle = (2E_{1i}V), \qquad \langle f|f\rangle \equiv \langle f_1f_2|f_1f_2\rangle = (2E_{1f}V)(2E_{2f}V), \tag{B10}$$

where $E_{1,2f} = \sqrt{m^2+\mathbf{p}_{1,2f}^2}$ (and E_{1i}) are the particle energies with mass m and V is the volume of the box containing the particles. Observe that $|i_1\rangle$ and $|f_1f_2\rangle$ are not normalized to unity. Therefore, upon substituting (B10) into (B9) the transition probability can be written as

$$\mathcal{P}_{i\to f} = \frac{|\langle f_1f_2|(\mathrm{i}T)|i_1\rangle|^2}{\langle f_1f_2|f_1f_2\rangle\langle i_1|i_1\rangle} = t\frac{(2\pi)^4}{2E_{1i}}\delta^4\big(p_{1i}-p_{1f}-p_{2f}\big)|\mathcal{M}|^2\prod_{n=1}^{2}\frac{1}{2E_{nf}V}\,. \tag{B11}$$

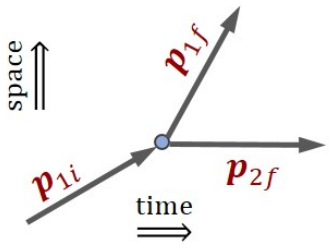


Fig. B.1: Feynman diagram for the decay $\mathbf{p}_{1i} \to \mathbf{p}_{1f} + \mathbf{p}_{2f}$ of one initial structureless particle with 4-momentum $p_{1i} = (E_{1i}, \mathbf{p}_{1i})$ into two final structureless particles with 4-momenta $p_{1f,2f} = (E_{1f,2f}, \mathbf{p}_{1f,2f})$ after decay.

The probability of transition per unit time (called the *transition rate*) between the initial decaying particle state and the final two-particle state created by decay is obtained by dividing (B11) by t, yielding

$$\frac{\mathrm{d}\mathcal{P}_{i\to f}}{\mathrm{d}t} = \frac{1}{2E_{1i}}(2\pi)^4\delta^4\big(p_{1i}-p_{1f}-p_{2f}\big)|\mathcal{M}|^2\prod_{n=1}^{2}\frac{1}{2E_{nf}V} \tag{B12}$$

$$\text{or} \quad \frac{\mathrm{d}\mathcal{P}_{i\to f}}{\mathrm{d}t} = \frac{1}{2E_{1i}}(2\pi)^4\delta^4\big(p_{1i}-p_{1f}-p_{2f}\big)|\mathcal{M}|^2\left(\frac{1}{2E_{1f}V}\right)\left(\frac{1}{2E_{2f}V}\right)$$

which, said differently, is the probability of transition per unit time to one particular final state $|f\rangle = |f_1\rangle|f_2\rangle$ out of all possible final states, made up of two particles with momenta p_{1f} and p_{2f} in that state. In the limit as the box volume $V \to \infty$, the momentum values become continuous, i.e. the $|p_1\rangle + |p_2\rangle$ final states form a continuum, and the *total decay rate*, called Γ, is obtained by integration of $\dot{\mathcal{P}}_{i\to f}$ over the continuum of final states. To determine whether the final momentum of a given particle lies within a domain $\mathrm{d}^3\mathbf{p}$ in the three-dimensional momentum space, Eq. (B12) must be multiplied by the number of states in this domain. To do this, the three-dimensional coordinate space is discretized into cells of size

$(2\pi)^3$ and one state is put into each cell. Hence, in a large box of volume V, the number of single-particle states in a momentum space domain $\mathrm{d}^3\mathbf{p}$ is $V\,\mathrm{d}^3\mathbf{p}/(2\pi)^3$.

Comment: Note that $(2\pi)^3$ is the discretized cell size in HL units, where $\hbar = 1$ and $h \stackrel{\text{def}}{=} 2\pi\hbar = 2\pi$. Expressed in SI units, the cell size is h^3, with $h = 2\pi\hbar = 6.626\times10^{-34}$ J s and $\hbar = 1.054\times10^{-34}$ J s, h being the Planck constant.

Finally, multiplying all the final state particles in Eq. (B12) by the factor $V\,\mathrm{d}^3\mathbf{p}/(2\pi)^3$ and integrating over the two final momenta yields the decay rate Γ (at times the name $\dot{\mathcal{P}}_{i\to f}$ is used for Γ):

$$\Gamma = \frac{1}{2E_{1i}} \int \frac{\mathrm{d}^3\mathbf{p}_{1f}}{(2\pi)^3 2E_{1f}} \int \frac{\mathrm{d}^3\mathbf{p}_{2f}}{(2\pi)^3 2E_{2f}} \, (2\pi)^4 \delta^4\big(p_{1i} - p_{1f} - p_{2f}\big) |\mathcal{M}|^2 \, . \tag{B13}$$

Notice that all the quantities V (and t) have dropped out at the end of the calculations. The amplitude $\mathcal{M}$ will be calculated by means of the Feynman rules (as outlined in the main text). Similar types of calculations are undertaken for the study of scattering *cross-sections*, also referred to as *rate per flux*. An effective application of these powerful computational tools to intrabeam scattering is given by the Bjorken–Mtingwa model. Eq. (B13) agrees with the decay rate formulation [32] (appendix B) and has similarities to the Bjorken's transition rate for the scattering process, Eq. (158).

References

[1] A. Piwinski, Intra-beam scattering, Proc. 9th Int. Conference on High Energy Accelerators, Stanford, CA, 1974 (SLAC, Stanford, CA, 1974), p. 405.

[2] J. Bjorken and S. Mtingwa, *Part. Accel.* **13** (1983) 115.

[3] I. Meshkov *et al.*, BETACOOL Physics Guide, Joint Institute for Nuclear Research, Dubna, Russia (2008).

[4] P. Zenkevich, A. Bolshakov and O. Boine-Frankenheim, *Kinetic effects in multiple intra-beam scattering*, ICFA HB204, Bensheim, 2005 [AIP Conf. Proc. **773** (2005) 425].

[5] M. Martini and A. Vivoli, Effect of intrabeam scattering and synchrotron radiation damping when reducing transverse emittances to augment the LHC luminosity, CERN-sLHC-PROJECT-Report-0032 (2010).

[6] A. Piwinski, in Proceedings of the CAS–CERN Accelerator School and NIKHEF-H, 16–27 June 1991, Amsterdam, the Netherlands, edited by T. Stuart, CERN–1992–001 (CERN, Geneva,1992). http://dx.doi.org/10.5170/CERN-1992-001

[7] Ch. Carli, Private communication, CERN (2012).

[8] M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, Cambridge, 2007). http://dx.doi.org/10.1017/CBO9780511815898

[9] J.A. Bittencourt, *Fundamentals of Plasma Physics* (Springer, New York, 2004). http://dx.doi.org/10.1007/978-1-4757-4030-1

[10] R.L. Liboff, *Kinetic Theory* (Springer, New York, 2003).

[11] L. Landau, E.M. Lifshitz and L.P. Pitaevskii, *Physical Kinetics*, Course of Theoretical Physics, vol. 10 (Butterworth-Heinemann, Oxford, 1981).

[12] D. Tong, *Kinetic Theory*, Graduate Course, University of Cambridge, Cambridge, UK (2012).

[13] K. Huang, *Statistical Mechanics* (Wiley, New York, 1987).

[14] H. Jakobsen, *Chemical Reactor Modeling: Multiphase Reactive Flows* (Springer, New York, 2014). http://dx.doi.org/10.1007/978-3-319-05092-8

[15] N.S. Dikansky and D.V. Pestrikov, *The Physics of Intense Beams and Storage Rings* (AIP Press, Woodbury, New York, 1994).

[16] Ya.S. Derbenev, Collisional relaxation of intense beams of heavy particles in storage rings, Fermilab $\bar{\mathrm{p}}$ Note 176 (1981).

[17] A.H. Sørensen, Intrabeam scattering, US-CERN School on Particle Accelerators, Hilton Head Island, South Carolina, USA (1990).

[18] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley-VCH, Weinheim, Germany, 2008). http://dx.doi.org/10.1002/9783527622047

[19] H. Wiedemann, *Particle Accelerator Physics* (Springer, Berlin, 2007).

[20] A. Piwinski, Intra-beam scattering, Joint US-CERN School on Particle Accelerators, Texas, USA (1986).

[21] J. Freund, *Special Relativity for Beginners* (World Scientific, Singapore, 2008). http://dx.doi.org/10.1142/6601

[22] A. Piwinski, The Touschek effect in strong focusing storage rings, DESY-98-179, Hamburg, Germany (1998).

[23] M. Zisman, S. Chattopadhyay and J. Bisognano, ZAP user's manual, LBL-21270, E.SG-15 (Lawrence Berkeley Laboratory, Berkeley, CA, 1986).

[24] M. Martini, Intrabeam scattering in the ACOL-AA machines, CERN PS/84 9 (1984).

[25] L.R. Evans and B. Zotter, Intrabeam scattering in the SPS, CERN/SPS/80-15-DI (1980).

[26] K. Kubo and K. Oide, Intrabeam scattering in electron storage rings (2001) 124401. http://dx.doi.org/10.1103/PhysRevSTAB.4.124401

[27] K. Bane, A simplified model of intrabeam scattering, in Proceedings of the eighth European Particle Accelerator Conference, Paris, France, 2002 (EPS-IGA and CERN, Geneva, 2002), p. 1443.

[28] K. Kubo, S. Mtingwa and A. Wolski, Intrabeam scattering formulas for high energy beams (2005) 081001. http://dx.doi.org/10.1103/PhysRevSTAB.8.081001

[29] A. Wolski, Space charge, intrabeam scattering and Touschek effects, Fourth International Accelerator School for Linear Colliders, Beijing, China, 2009.

[30] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators* (Imperial College Press, London, 2014).

[31] D. McMahon, *Quantum Field Theory Demystified* (McGraw-Hill, New York, 2008).

[32] J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[33] S. Mtingwa, A New High Energy Approximation of Intrabeam Scattering for Flat Electron and Positron Beams, African Physical Review (2008) 2:0001.

[34] D. Griffiths, *Introduction to Elementary Particles* (Wiley-VCH, Weinheim, Germany, 2010).

[35] I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics. Vol. 1: From Relativistic Quantum Mechanics to QED* (CRC Press, Boca Raton, FL, 2013).

[36] F. Zimmermann, Refined models of intrabeam scattering, Proc. HB2006, Tsukuba, Japan (2006), p. 265.

[37] F. Antoniou and F. Zimmermann, Revision of intrabeam scattering with non-ultrarelativistic corrections and vertical dispersion for MAD-X, CERN-ATS-2012-066 (2012).

[38] J. Resta-López, J.R. Hunt and C.P. Welsch, Intrabeam scattering effects in ELENA, IPAC2015 Conf., Richmond, Virginia, USA (2015).

[39] V. Chohan (Ed.), Extra low energy antiproton (ELENA) ring and its transfer lines, CERN-2014-002 (2014). http://dx.doi.org/10.5170/CERN-2014-002

[40] M.D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, London, 2014).

[41] A. Lahiri and P.B. Pal, *A First Book of Quantum Field Theory* (Alpha Science, Oxford, 2005).

[42] D. Tong, *Quantum Field Theory*, Lecture notes, University of Cambridge, Cambridge, UK (2007).

[43] S. Weinberg, *Lectures on Quantum Mechanics* (Cambridge University Press, New York, 2013).

[44] S. Weinberg, *Quantum Theory of Fields. Vol. 1* (Cambridge University Press, London, 2005).

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Space Charge in Circular Machines

G. Franchetti
GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany

Abstract
Direct Coulomb forces play an important role in beam transport and storage. In this paper, the effect of space charge in beams stored in circular machines is reviewed. Starting from the concept of a matched beam, the most common particle distributions are introduced. Space charge forces are first addressed as 'external frozen' forces and, in this approximation, the modification of the machine optics is discussed, as well as the matching of high-intensity beams. Envelope equations and r.m.s. equivalence are presented as relevant tools for matching high-intensity beams. The space charge limit, space charge tune-shift, tune-spread and their relation with resonances are covered. The more general beam response to Coulomb forces is discussed for mismatched beams, and a derivation of the coherent frequencies is presented for a Kapchinsky–Vladimirsky beam. The concept of free energy is examined in detail, as a source of emittance increase for high-intensity beams. More complex and relevant examples are presented with the fully coupled envelope equations, and the interplay of coherent and incoherent effects is addressed in the Montague resonance. Space charge as driving incoherent resonances is discussed. Longitudinal dynamics in the presence of space charge are briefly presented as well.

Keywords
CERN report; space charge.

1 Introduction

In the following, we will discuss the effect of space charge in circular machines. We will first focus on describing the dynamics of a single particle within a beam. For coasting beams, we use a reference frame co-moving with the reference particle; hence, the co-ordinates x, y used to identify the positions of a particle are meant with respect to the design orbit. Para-axial approximation is used and, as usual, $x' = \mathrm{d}x/\mathrm{d}s, y' = \mathrm{d}y/\mathrm{d}s$, where s is the longitudinal co-ordinate set on the reference orbit.

2 Matched beams at low intensity

We consider in our discussion a beam distribution for two-dimensional (2D) beams, i.e., a coasting beam. Typically, a 2D beam is characterized by a constant charge-line density of $\lambda = \mathrm{d}q/\mathrm{d}s$, and its most general transverse particle density at longitudinal position s is defined as

$$\rho(x,x',y,y',s) = \frac{\mathrm{d}q}{\mathrm{d}x\mathrm{d}x'\mathrm{d}y\mathrm{d}y'\mathrm{d}s} = \lambda n(x,x',y,y',s)\,.$$

The form of the normalized function $n(x,x',y,y',s)$ allows a complete characterization of the transverse beam distribution. The normalization condition on $n(x,x',y,y',s)$ is obtained by the requirement

$$\int n(x,x',y,y',s)\mathrm{d}x\mathrm{d}x'\mathrm{d}y\mathrm{d}y' = 1\,. \tag{1}$$

The variable s in $n()$ says that the beam distribution may change along the machine; this aspect will be discussed later.

https://doi.org/10.23730/CYRSP-2017-003.353

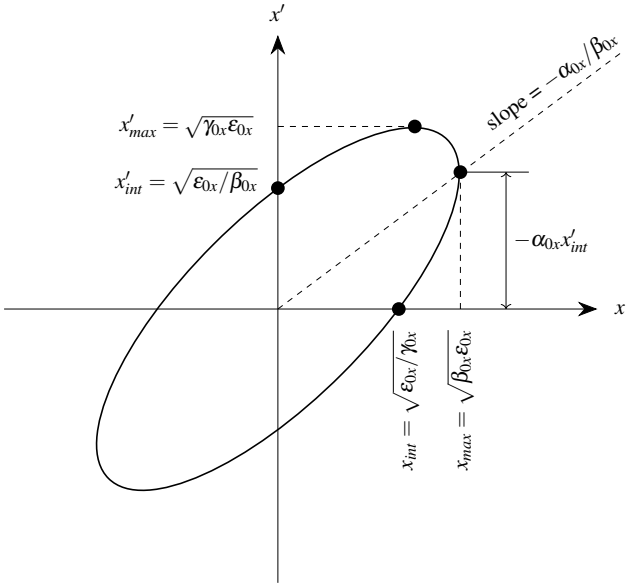


Fig. 1: Courant–Snyder ellipse: ε_{0x}, single-particle emittance; area $= \pi\varepsilon_{0x}$

In a circular accelerator, the transverse dynamical properties of a low-intensity beam at an arbitrary longitudinal position s is determined by the Poincaré section, which for a linear motion is completely described by the Courant–Snyder theory [1]. We will consider 2D beams matched with the optics. A matched beam means that the transverse beam distribution at the section s will be the same after the beam has travelled for one machine turn. In this sense, the property of a particle beam to be matched is related with the Courant–Snyder ellipses, shown in Fig. 1. In the absence of coupling, the planes x–x' and y–y' can be treated independently. At the longitudinal position s in the plane x–x', turn after turn, a particle performs jumps, remaining on a distinct ellipse. Each jump corresponds to a phase jump of $2\pi Q_{0x}$, with Q_{0x} the horizontal machine tune. Similar dynamics happen in the vertical plane. If only one particle is considered, clearly we can distinguish where it is turn after turn. Therefore, in the sense previously defined, a single particle cannot be 'matched' with the machine optics. However, if an ellipse is 'uniformly' populated by particles, then after one turn there is no way for an observer sitting at position s to tell that the beam has made one turn around the machine.

A beam uniformly filling 'one' Courant–Snyder ellipse is then matched. More generally, we can think of a matched beam as a collection of uniformly filled Courant–Snyder ellipses, each with an arbitrary particle density. This means that the most general matched beam distribution is a function of the Courant–Snyder invariants; hence, the normalized function $n()$ is written as

$$n(\varepsilon_{0x}, \varepsilon_{0y}, s)\,. \tag{2}$$

In this formula, the s dependence simply means that the function $n()$ may change along the machine. This mathematical formula (Eq. (2)) says that the particle density depends on the value of the quantities $\varepsilon_{x0}, \varepsilon_{y0}$, which are defined by the expressions:

$$\begin{aligned} \varepsilon_{0x} &= \gamma_{0x}x^2 + 2\alpha_{0x}xx' + \beta_{0x}x'^2\,, \\ \varepsilon_{0y} &= \gamma_{0y}y^2 + 2\alpha_{0y}yy' + \beta_{0y}y'^2\,. \end{aligned} \tag{3}$$

The optical functions

$$\gamma_{0x}, \alpha_{0x}, \beta_{0x}, \gamma_{0y}, \alpha_{0y}, \beta_{0y}\,, \tag{4}$$

are here computed at the longitudinal position s where the Poincaré section is studied. The function $n()$ in Eq. (2) yields the same value when the co-ordinates x,x',y,y' keep $\varepsilon_{0x},\varepsilon_{0y}$ unchanged; therefore, this means that a distribution modelled with a normalized function like Eq. (2), i.e., depending on $\varepsilon_{0x},\varepsilon_{0y}$, will produce a particle distribution matched with the optics at section s because all the Courant–Snyder ellipses are uniformly populated. The index 0 in the optical functions (Eq. (4)) is used to indicate that ε_{0x} and ε_{0y} are computed for weak space charge, virtually absent.

The function of Eq. (2) yields the most general expression for a matched distribution. However, some other consideration has to be used to model beam distributions more realistically: an energy conservation is here invoked, which basically 'incorporates' the physics of the source, all optics manipulations of linacs, and schemes of injection in the ring. As a consequence, a correlation is assumed in the single-particle energy between both planes, and the general matched particle distribution can be written as

$$n(x,x',y,y',s) = \frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \tilde{n}\left(\frac{\varepsilon_{0x}(x,x')}{\mathscr{E}_x} + \frac{\varepsilon_{0y}(y,y')}{\mathscr{E}_y}\right), \tag{5}$$

where $\mathscr{E}_x$, and $\mathscr{E}_y$ are some 'scaling' factors, which define the geometrical extension of the distribution in the 4D phase space. In this expression, we drop the dependence of s in $\tilde{n}()$ because, by our definition, the matched distribution must be periodic, and the periodicity is already included into the optic functions. The absence of s in the function $\tilde{n}()$ also means that the type of distribution does not change and remains the same (at this point of the discussion, this is an ansatz). The normalization condition Eq. (1) applied to Eq. (5) reads

$$\int_0^\infty \tilde{n}(t) t \mathrm{d}t = 1\,.$$

with $\tilde{n}(t)$ defined in $0 \le t < \infty$.

3 Main types of beam distribution

According to the type of function $\tilde{n}$ we use, a different type of matched beam distribution is obtained. The mostly used beam distributions are as follows.

1. Kapchinsky–Vladimirsky:

$$n(x,x',y,y',s) = \frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \delta\left(\frac{\varepsilon_{0x}}{\mathscr{E}_x} + \frac{\varepsilon_{0y}}{\mathscr{E}_y} - 1\right), \tag{6}$$

   where $\delta()$ is the Dirac delta function.

2. Waterbag:

$$n(x,x',y,y',s) = \frac{2}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \Theta\left(1 - \frac{\varepsilon_{0x}}{\mathscr{E}_x} - \frac{\varepsilon_{0y}}{\mathscr{E}_y}\right), \tag{7}$$

   where $\Theta()$ is the Heaviside function.

3. Gaussian:

$$n(x,x',y,y',s) = \frac{1}{4\pi^2 \mathscr{E}_x \mathscr{E}_y} \mathrm{e}^{-\frac{1}{2}\left(\frac{\varepsilon_{0x}}{\mathscr{E}_x} + \frac{\varepsilon_{0y}}{\mathscr{E}_y}\right)}. \tag{8}$$

In all these distributions, it is understood that $\varepsilon_{0x},\varepsilon_{0x}$ are the functions in Eq. (3).

As the space charge forces are determined by the spatial particle distribution, it is convenient to discuss the general form of a 2D matched beam in the x–y plane. This distribution is proportional to the projection of $n(x,x',y,y')$ on the x–y space, which, for convenience, we call $n(x,y)$, and is readily obtained from

$$n(x,y,s) = \int n(x,x',y,y',s)\mathrm{d}x'\mathrm{d}y'\,.$$

Table 1: Functions $\tilde{n}(t)$, and $\hat{n}(t)$ for the beam distribution discussed.

	Kapchinsky–Vladimirsky	Waterbag	Gaussian
$\tilde{n}(t)$	$\delta(t-1)$	$2\Theta(1-t)$	$\frac{1}{4}e^{-t/2}$
$\hat{n}(t)$	$\Theta(1-t)$	$2(1-t)\Theta(1-t)$	$\frac{1}{2}e^{-t/2}$

A direct integration of this equation using $n(x,x',y,y',s)$ given by Eq. (5) yields

$$n(x,y,s) = \frac{1}{\pi a_0(s) b_0(s)} \left[F(\infty) - F\left(\frac{x^2}{a_0^2(s)} + \frac{y^2}{b_0^2(s)} \right) \right], \tag{9}$$

where $F(t) = \int_0^t \tilde{n}(t')\mathrm{d}t'$ is the primitive of $\tilde{n}(t)$. We also define $a_0(s) = \sqrt{\beta_{0x}(s)\mathscr{E}_x}$ and $b_0(s) = \sqrt{\beta_{0y}(s)\mathscr{E}_y}$. These quantities are related to the 'beam sizes' at the longitudinal positions s; in particular, for finite beam distribution, $a_0(s), b_0(s)$ are exactly the beam sizes. It is convenient to rewrite Eq. (9) in the form

$$n(x,y,s) = \frac{1}{\pi a_0(s) b_0(s)} \hat{n}\left(\frac{x^2}{a_0^2(s)} + \frac{y^2}{b_0^2(s)} \right), \tag{10}$$

with $\hat{n}(t) = F(\infty) - F(t)$ defined in $0 \le t < \infty$. From the property of $n(t)$ it follows that $\hat{n}(t)$ must satisfy the normalization condition $\int_0^\infty \hat{n}(t)\mathrm{d}t = 1$. Equation (10) is the most general expression of the normalized spatial 2D beam distribution for a matched beam. The Table 1 shows the functions $\tilde{n}(t)$, and $\hat{n}(t)$ for the main beam distributions.

To clarify the difference between the particle distributions in Eqs. (6)–(8), we plot each of them in several planes. Figure 2 shows the particle distribution in the x–y, x–x', and x profiles for the three types of distributions here discussed. The top row of the figures shows the properties of the Kapchinsky–Vladimirsky distribution. The two particle distributions in the x–y and x–x' profiles show a peculiar characteristic of the Kapchinsky–Vladimirsky distribution, namely that the projection in any plane yields constant particle density. This is a general property of the Kapchinsky–Vladimirsky distribution. The second row of figures refers to a waterbag distribution, and we see that the x–y projection is more dense in the centre. This is also seen by a more pronounced peaked beam profile. The last row of figures shows the properties of a Gaussian distribution, truncated for convenience at $\varepsilon_{0x}/\mathscr{E}_x + \varepsilon_{0y}/\mathscr{E}_y \le 16$.

4 Space charge forces for a frozen distribution

For a static particle distribution in which each particle has position $\vec{r}_i$ and charge q, the electric field at the position $\vec{r}$ is given by Coulomb's law:

$$\vec{E}(\vec{r}) = \frac{q}{4\pi\varepsilon_0} \sum_i \frac{\vec{r} - \vec{r}_i}{|\vec{r} - \vec{r}_i|^3},$$

where ε_0 is the permittivity of vacuum.

Next, we analyse the electric field produced in coasting beams transported in alternating focusing structures. In this case, the electric field is computed with the approximation of a 'local coasting beam': the beam sizes $a_0(s), b_0(s)$ vary because of the optics, and their values change following some wavelength Δs, which depends on the particular optics; see Fig. 3 for an illustration with sizes altered for convenience. If $\Delta s \gg a_0, b_0$, then at the longitudinal position s, the transverse electric field is in good approximation, computed from the coasting beam with constant sizes that mimics the transverse particle distribution at s. In Fig. 3, this coasting beam with constant sizes is shown with black dashed lines. At the section s, the electric field of the actual 2D beam and that of the mimic beam are practically the same. In this approximation, the electric field on a transverse plane at the section s is computed by the coasting beam

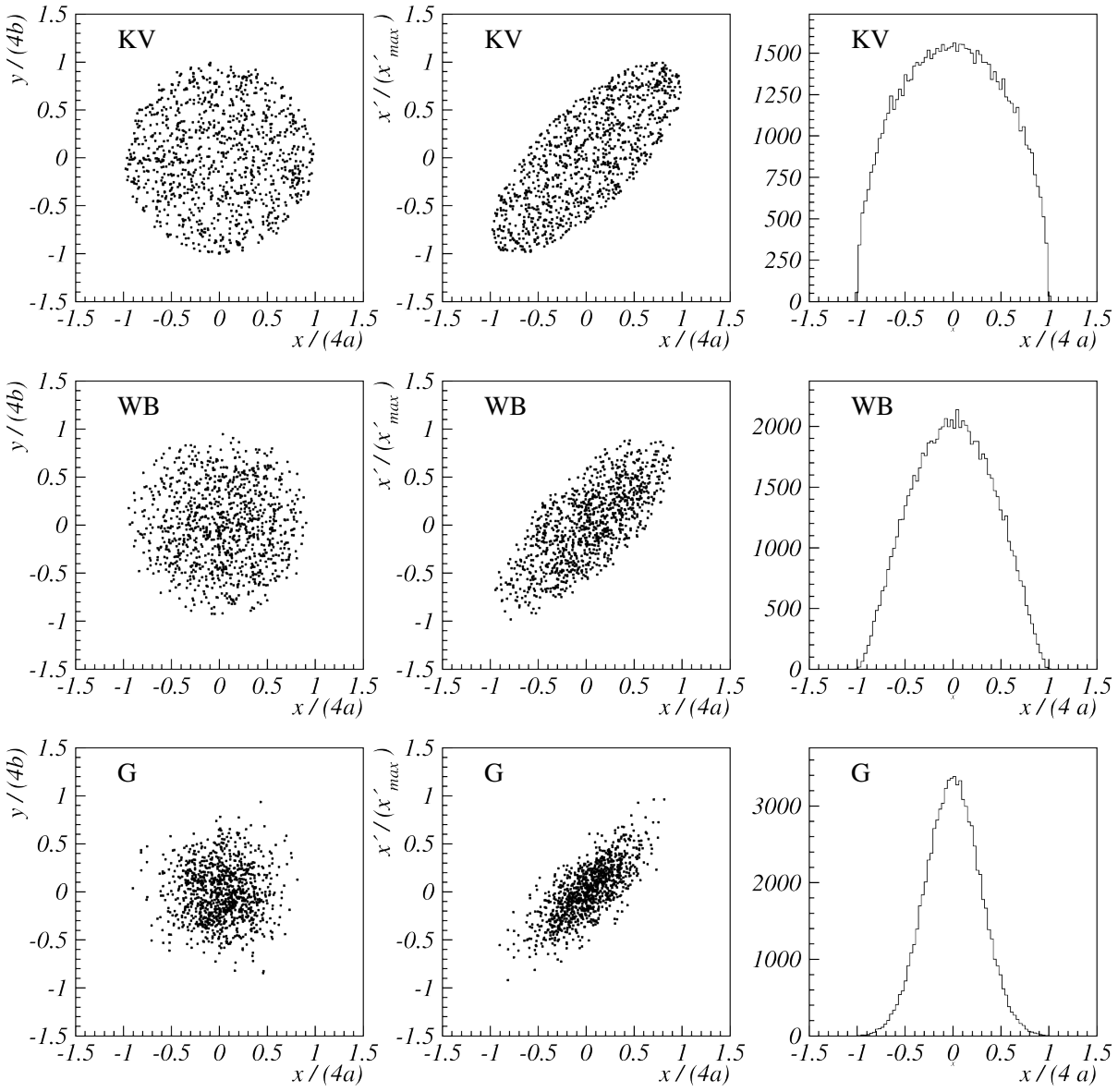


Fig. 2: Projections of (top) Kapchinsky–Vladimirsky (KV), (middle) waterbag (WB), and (bottom) Gaussian distributions. The right-hand column shows the change in the x beam profile.

with the charge distribution $\rho(x,y,s) = \lambda n(x,y,s)$ (see Fig. 3). In the simple case of an axi-symmetric beam of radius $a_0 = b_0$ with a uniform beam density, the radial electric field $E_r(s)$ is given by

$$E_r(s) = \begin{cases} \frac{\rho(s)}{2\varepsilon_0} r & \text{if} \quad r \le a_0\,, \\ \frac{\rho(s) a_0^2(s)}{2\varepsilon_0} \frac{1}{r} & \text{if} \quad r \ge a_0\,. \end{cases} \tag{11}$$

The characteristic shape of this function is shown in Fig. 4. The next more complex situation is with an axi-symmetric non-uniform beam. The general expression of the electric field for an axi-symmetric beam distribution with density $\rho(r,s)$ is readily found by using Gauss's law

$$E_r(r,s) = \frac{1}{\varepsilon_0} \frac{1}{r} \int_0^r \rho(r',s) r' \mathrm{d}r'\,. \tag{12}$$

Note that this expression at section s holds only if $a_0 = b_0$ in Eq. (9). This condition is equivalent to $\beta_{0x}\mathscr{E}_x = \beta_{0y}\mathscr{E}_y$.

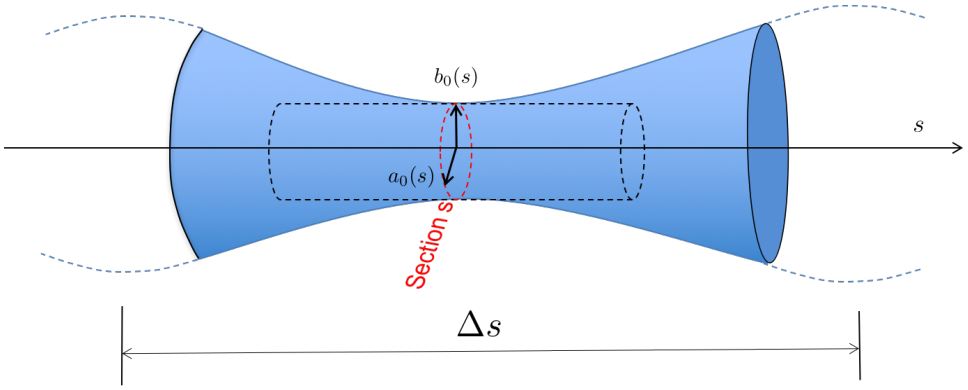


Fig. 3: Local coasting beam approximation

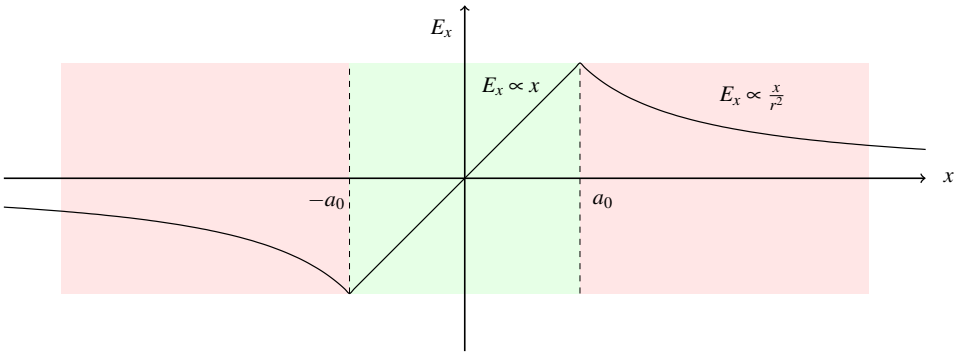


Fig. 4: Radial electric field in an axi-symmetrical beam

If at position s, we have $\beta_{0x}\mathscr{E}_x \neq \beta_{0y}\mathscr{E}_y$ i.e., $a_0 \neq b_0$, then Eq. (12) no longer holds. In this case, the computation of the electric field becomes more difficult, as it has to account for the different axis symmetry. The physical x–y charge density of the beam is given by $\rho(x,y) = \lambda n(x,y)$ with $n(x,y)$ from Eq. (10). It is now convenient to define Eq. (10) as

$$n(x,y) = \frac{\hat{n}(T)}{\pi a_0 b_0}\,, \tag{13}$$

with T an iso-density parameter

$$T = \frac{x^2}{a_0^2} + \frac{y^2}{b_0^2}\,.$$

With the use of these definitions, the transverse electric field is found by the integral

$$E_x = \frac{\lambda}{2\pi\varepsilon_0} x \int_0^\infty \frac{\hat{n}(\hat{T})}{(a_0^2+t)^{3/2}(b_0^2+t)^{1/2}}\mathrm{d}t\,, \tag{14}$$

where the variable $\hat{T}$ is defined as

$$\hat{T} = \frac{x^2}{a_0^2+t} + \frac{y^2}{b_0^2+t}\,. \tag{15}$$

The derivation of this expression is given in Ref. [2]. The electric field E_y is obtained by exchanging $x \leftrightarrow y$ and $a_0 \leftrightarrow b_0$. It is straightforward to verify that Eq. (14) yields Eq. (12) for the case of $a_0 = b_0$; in fact, in this case we obtain

$$E_x = \frac{\lambda}{2\pi\varepsilon_0} x \int_0^\infty \frac{\hat{n}\left(\frac{r^2}{a_0^2+t}\right)}{(a_0^2+t)^2}\mathrm{d}t \tag{16}$$

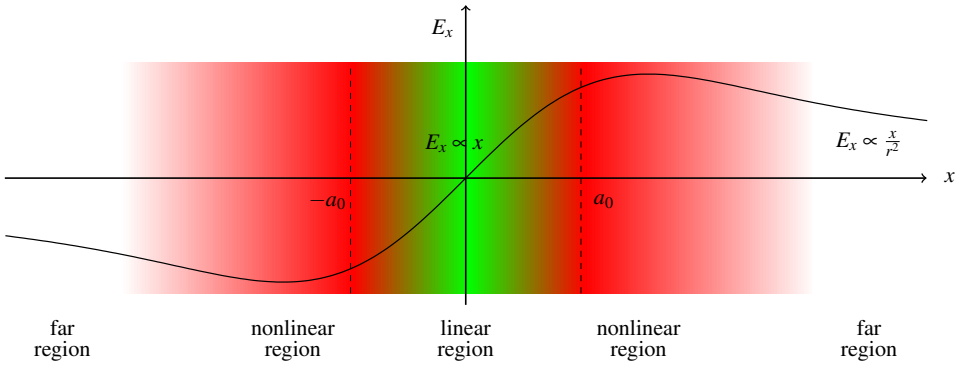


Fig. 5: Electric field generated by a 2D Gaussian particle distribution. For comparison, in the linear region the gradient is the same of that of Fig. 4. The green area is the region of linear electric field. It is clear that the stronger electric field is not found at $\pm a_0$.

with $r^2 = x^2 + y^2$, and by making the variable substitution $r' = ra_0/\sqrt{a_0^2 + t}$, we obtain Eq. (12).

According to the particle amplitude, we distinguish the electric field in three regions: (1) a linear region close to the beam centre; (2) a non-linear region, characterized by a significant change in the electric field in a non-linear way; (3) a far region characterized by a functional dependence as $1/r$. This third dependence is valid for $a_0, b_0 \ll r \ll R$, with R the radius of the accelerator. This characterization of the electric field is shown in Fig. 5.

5 Dynamics in the linear region

The previous section has helped to characterize the electric field generated in the transverse section of a 2D matched beam. The way in which the discussion has been developed assumes that the beam does not change as a result of space charge; hence, the matched beam distribution is following the optics, and does not change for other reasons. We call this distribution 'frozen'. The x–y projection of the frozen beam at the longitudinal position s always creates an ellipsoidal beam distribution described by Eq. (10). We now discuss the effect of the electric field on a particle with a trajectory that remains close to the beam centre. This means that, at any instant during the transport, the co-ordinates x, y of this particle should satisfy the condition $|x|, |y| \ll a_0, b_0$. This condition simplifies Eq. (14). In fact, $\hat{T}$ as defined by Eq. (15) will always satisfy $\hat{T} \leq x^2/a_0^2 + y^2/b_0^2$ for any $t \geq 0$. As $|x|, |y| \ll a_0, b_0$, it follows that $\hat{T}$ will always be very small; hence, to a good approximation, we can replace $\hat{T} = 0$. We therefore find that for $|x|, |y| \ll a_0, b_0$ the electric field is

$$E_x = \frac{\lambda \hat{n}(0)}{2\pi\varepsilon_0} x \int_0^\infty \frac{1}{(a_0^2 + t)^{3/2}(b_0^2 + t)^{1/2}} \mathrm{d}t \,,$$

and the integral is readily evaluated, with the result:

$$E_x = \frac{\lambda \hat{n}(0)}{\pi\varepsilon_0} \frac{1}{a_0(a_0 + b_0)} x \,. \tag{17}$$

This equation says that all the 2D matched frozen distributions create linear forces in the proximity of the origin. The influence on the type of beam distribution appears in the term $\hat{n}(0)$. This quantity changes value according to whether the distribution is Kapchinsky–Vladimirsky, waterbag, or Gaussian. We notice that the electric field in the proximity of the beam centre acts like a defocusing element distributed along the ring.

As we are considering a 'frozen' beam, that is, a beam with sizes following the machine optics, we can at this stage consider the electric force generated by the beam as 'external', just as if it is applied

to the beam particles from a sort of a distributed quadrupole. It therefore makes perfect sense to consider what the single-particle dynamics would be in the presence of such a force.

We remind ourselves here that the single-particle dynamics of a particle in a lattice and in the presence of space charge is

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2} + k_{0x}(s)x = \frac{q}{m\gamma^3 v_s^2}E_x\,, \tag{18}$$

where $k_{0x}(s)$ is the horizontal focusing or defocusing strength exerted by the magnets on one charged particle with charge q and mass m, $\gamma = 1/\sqrt{1-v_s^2/c^2}$, and v_s is the speed of the particle. E_x is the Coulomb electric field, and the term γ^3 includes the effect of the self-magnetic field, which, at high energies, compensates the space charge. If we substitute the electric field close to the origin (Eq. (17)) in the equation of motion (Eq. (18)), then the equation of motion becomes a Hill equation that includes the effect of the linear space charge. This happens in the centre of the beam under the assumption that the beam is coasting, ellipsoidal, and frozen. It is convenient to incorporate all constants in one single coefficient called the perveance, defined as

$$K = \frac{qI}{2\pi\varepsilon_0 m\gamma^3\beta^3c^3}\,, \tag{19}$$

where I is the beam current, and $\beta = v_z/c$. In the following, we take K to be always positive. The equations of motion of the single particle then become

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2} + \left[k_{0x}(s) - \hat{n}(0)\frac{2K}{a_0(a_0+b_0)}\right]x = 0\,, \tag{20}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}s^2} + \left[k_{0y}(s) - \hat{n}(0)\frac{2K}{b_0(a_0+b_0)}\right]y = 0\,. \tag{21}$$

As previously stated, these are the equations of motion for a particle in the centre of a coasting beam. The 2D beam distribution is defined by the Eq. (5), the x–y projection of which is given by Eq. (9). In this discussion, the beam distribution is 'frozen', i.e., the type of distribution does not change, and the sizes a_0, b_0 are 'matched' with the machine optics as $a_0 = \sqrt{\beta_{0x}\mathscr{E}_x}$ and $b_0 = \sqrt{\beta_{0y}\mathscr{E}_y}$. Now we can discuss the optics that a particle will experience in the presence of frozen space charge when its motion is close the origin. From Eqs. (20) and (21), we find that the effective strength of the lattice is now:

$$\begin{aligned} k_{1x}(s) =& k_{0x}(s) - \hat{n}(0)\frac{2K}{a_0(a_0+b_0)}\,,\\ k_{1y}(s) =& k_{0y}(s) - \hat{n}(0)\frac{2K}{b_0(a_0+b_0)}\,. \end{aligned} \tag{22}$$

These two strengths are well defined, and produce the modified optical function $\beta_{1x}, \alpha_{1x}, \beta_{1y}, \alpha_{1y}$. Again, we observe that the memory of the type of distribution is in the factor $\hat{n}(0)$, and the size of the distribution is related to $a_0(s), b_0(s)$.

6 Frozen beam matched with space charge

If we consider the modified lattice in Eq. (22), we are obliged to admit that the effective strengths k_{1x}, k_{1y} will create a different optics from the original lattice. Once we fix $\mathscr{E}_x, \mathscr{E}_y$, given the optics $\beta_{0x}, \alpha_{0x}, \beta_{0y}, \alpha_{0y}$, the frozen beam distribution creates, in the centre of the beam, via space charge, the modified optics $\beta_{1,x}, \alpha_{1,x}, \beta_{1,y}, \alpha_{1,y}$, and this new optics is periodic, as originated by periodic functions: the effect of the frozen space charge is to modify the Twiss parameters.

We could now think to create a beam distribution, which is matched with the optics created by the space charge depressed strength k_{1x}, k_{1y}. This beam will have the following frozen distribution matched with the new optics

$$\frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \tilde{n} \left(\frac{\varepsilon_{1x}}{\mathscr{E}_x} + \frac{\varepsilon_{1y}}{\mathscr{E}_y} \right), \tag{23}$$

where the Courant–Snyder quantities $\varepsilon_{1x}, \varepsilon_{1y}$ are now constructed using the modified optical functions $\beta_{1x}, \alpha_{1x}, \beta_{1y}, \alpha_{1y}$, namely

$$\begin{aligned} \varepsilon_{1x} &= \beta_{1x} x'^2 + 2\alpha_{1x} x x' + \gamma_{1x} x^2 , \\ \varepsilon_{1y} &= \beta_{1y} y'^2 + 2\alpha_{1y} y y' + \gamma_{1y} y^2 . \end{aligned} \tag{24}$$

At this point it is clear that Eqs. (23) and (5) represent two distinct distributions, because the same distribution function makes use of two distinct sets of optics.

To achieve a more realistic modelling of the matching of a 2D intense particle beam, it is necessary that 'both the beam distributions' are the same. This, in fact, is what happens in the real machine: the machine optics is changed by the beam distribution, which in turn 'automatically' guides the beam evolution and creates the same optics. In other words, the matched 2D beam, including space charge, requires the creation of a frozen beam distribution with an optics that already include the effect of that space charge. The problem is that we do not know in advance the modified optics with space charge. Mathematically this means finding $\beta_x, \alpha_x, \beta_y, \alpha_y$ such that $a = \sqrt{\mathscr{E}_x \beta_x}, b = \sqrt{\mathscr{E}_y \beta_y}$ creates a depressed focusing strength

$$\begin{aligned} k_x(s) &= k_{0x}(s) - \hat{n}(0) \frac{2K}{a(a+b)} , \\ k_y(s) &= k_{0y}(s) - \hat{n}(0) \frac{2K}{b(a+b)} , \end{aligned} \tag{25}$$

which in turns creates exactly $\beta_x, \alpha_x, \beta_y, \alpha_y$. The solution to this problem is not obvious and must be discussed with care, according to the circumstances.

Practically, one starts from a 'naked' optics $\beta_{0x}, \alpha_{0x}, \beta_{0y}, \alpha_{0y}$, from which one finds the frozen beam sizes a_0, b_0. Using Eq. (22), we find new optics $\beta_{1,x}, \alpha_{1,x}, \beta_{1,y}, \alpha_{1,y}$. The new modified optics includes the space charge in an 'inconsistent' fashion. However, if the space charge is not too strong, the modified optics is closer than the naked one to the solution. Therefore, we can repeat the process, now creating another frozen beam, matched with the optics $\beta_{1,x}, \alpha_{1,x}, \beta_{1,y}, \alpha_{1,y}$, from which we compute the new beam sizes a_1, b_1 and again, using an equation similar to Eq. (22), we find a new optics $\beta_{2,x}, \alpha_{2,x}, \beta_{2,y}, \alpha_{2,y}$. This procedure of steps defines an infinite sequence by recursion:

$$\beta_{n,x}, \alpha_{n,x}, \beta_{n,y}, \alpha_{n,y} \longrightarrow \beta_{n+1,x}, \alpha_{n+1,x}, \beta_{n+1,y}, \alpha_{n+1,y} ,$$

which starts from $\beta_{0x}, \alpha_{0x}, \beta_{0y}, \alpha_{0y}$. If the sequence converges, the limit is our $\beta_x, \alpha_x, \beta_y, \alpha_y$, which satisfies Eq. (25). Practically if the sequence converges, an adequate approximation is reached in a relatively small number of iterations.

7 Stationary distributions

Till now we have treated the space charge created by a frozen beam. That is, a beam of the sizes which are frozen to the beam optics; in the last section, we found that it is closer to the reality if the frozen beam is matched with the optics that include the space charge (generated by the beam itself). However, another assumption has been implicitly used: we always assumed that the type of distribution remains

the same during the beam evolution. In the framework of the frozen beams, this is perfectly allowed. We now discuss what happens if we allow all particles to evolve according to the actual electric field, in a self-consistent way. We basically want to discuss the evolution of the complete beam. In general, our 'matched' distribution will evolve in a very complex way. If the mechanisms that lead to beam change are slow on the time-scale under consideration, we can assume that the distribution will remain matched, i.e., it keeps a form type

$$n(x,x',y,y',s) = \frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \tilde{n}\left(\frac{\varepsilon_x}{\mathscr{E}_x} + \frac{\varepsilon_y}{\mathscr{E}_y}, s\right),$$

where the appearance of the variable s in the function describing the beam distribution means that the type of distribution may change with time, for example a Kapchinsky–Vladimirsky distribution may change and become a Gaussian. In complete generality, the evolution of a particle distribution is ruled by the Vlasov equation. For convenience, let us call $\vec{v} = (x,x',y,y')$; the Liouville theorem states that:

$$\frac{\partial}{\partial s} n(\vec{v},s) + \sum_i v_i' \frac{\partial}{\partial v_i} n(\vec{v},s) = 0 \tag{26}$$

except that now the quantities v_i' are obtained from the canonical equations, which read:

$$v_i' = S_{ij} \frac{\partial}{\partial v_j} H(\vec{v}), \tag{27}$$

with

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix},$$

and $H(\vec{v})$ the Hamiltonian of the system.

Equations (26) and (27), jointly, form the Vlasov equation, which, in its explicit form, reads:

$$\frac{\partial n}{\partial s} + (\vec{\nabla} n) \cdot (S\vec{\nabla} H) = 0.$$

Note that, with space charge, the Liouville theorem is still applicable if bulk space charge and beam currents can be described by vector potentials $\phi, \vec{A}$ (see Ref. [3]).

This equation tells how the particle density, i.e., the function n, changes in time. This is expressed by the dependence on s, which, where it exists, yields $\frac{\partial}{\partial s} n(\vec{v},s) \neq 0$. If, instead,

$$\frac{\partial}{\partial s} n(\vec{v},s) = 0,$$

this means that the particle distribution does not depend on s; hence, it does not change 'shape' with time. In the absence of space charge, or with very small space charge, the general particle distribution

$$\tilde{n}\left(\frac{\varepsilon_{0x}(x,x',s)}{\mathscr{E}_x} + \frac{\varepsilon_{0y}(y,y',s)}{\mathscr{E}_y}, s\right)$$

is stationary. In fact, it is easy to check that, in the absence of space charge,

$$\frac{\partial}{\partial s} \tilde{n}\left(\frac{\varepsilon_{0x}(x,x',s)}{\mathscr{E}_x} + \frac{\varepsilon_{0y}(y,y',s)}{\mathscr{E}_y}, s\right) = 0.$$

In the presence of space charge, the situation is much more complicated. There is only one known distribution that is stationary with space charge and that is the Kapchinsky–Vladimirsky distribution. In fact, for the Kapchinsky–Vladimirsky distribution,

$$n(\vec{v},s) = \frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \delta\left(\frac{\varepsilon_x}{\mathscr{E}_x} + \frac{\varepsilon_y}{\mathscr{E}_y} - 1\right), \tag{28}$$

if used in Eq. (9), we find that the function $F(t)$ becomes $F(t) = \Theta(t-1)$, where the function $\Theta(u)$ has value $\Theta(u) = 0$ if $u < 0$, and $\Theta(u) = 1$ if $u > 0$. Therefore, we find

$$n(x,y) = \frac{1}{\pi ab}\left[1 - \Theta\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1\right)\right] = \frac{1}{\pi ab}\Theta\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right),$$

which simply means that if (x,y) is inside the beam, i.e., if

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1,$$

the projection of the beam on the x–y plane has constant density $n(x,y) = 1/(\pi ab)$; if (x,y) is outside the beam, then $n(x,y) = 0$, as one expects. As $n(x,y)$ is constant, we find that the electric fields are linear everywhere inside the beam. In fact, Eq. (14) specialized to the Kapchinsky–Vladimirsky distribution yields

$$E_x = \frac{\lambda}{2\pi\varepsilon_0} x \int_0^\infty \frac{1}{(a_0^2+t)^{3/2}(b_0^2+t)^{1/2}} \mathrm{d}t \tag{29}$$

for any (x,y) inside the beam, and by direct integration we find

$$E_x = \frac{\lambda}{\pi\varepsilon_0} x \frac{1}{a_0(a_0+b_0)}.$$

Similarly, for the y plane. We therefore find that the Kapchinsky–Vladimirsky distribution creates a linear electric field everywhere inside the beam. This electric field enters the Hill's equation, and consequently the optics with space charge can be obtained. But now it is straightforward to prove that, for the particle distribution in Eq. (28), we find

$$\sum_i v_i' \frac{\partial}{\partial v_i} n(\vec{v},s) + \frac{\partial n}{\partial \varepsilon_x}\frac{\partial \varepsilon_x}{\partial s} + \frac{\partial n}{\partial \varepsilon_y}\frac{\partial \varepsilon_y}{\partial s} = 0,$$

which implies that

$$\frac{\partial}{\partial s} n\left(\frac{\varepsilon_x(x,x',s)}{\mathscr{E}_x} + \frac{\varepsilon_y(y,y',s)}{\mathscr{E}_y}, s\right) = 0,$$

that is the Kapchinsky–Vladimirsky is a stationary distribution. (The partial derivative is applied only to the last s, i.e., to the functional dependence of the type of distribution). All this means that any initial Kapchinsky–Vladimirsky distribution will remain a Kapchinsky–Vladimirsky distribution, that is, this distribution will not change shape during the beam evolution.

Next, we show an example of evolution of a non-stationary beam distribution. In Fig. 6, we see in green the y profile of a Kapchinsky–Vladimirsky distribution. This distribution is injected into a lattice with $Q_{0x} = Q_{0y} = 4.41$ with emittances $\tilde{\varepsilon}_x = \tilde{\varepsilon}_y = 5$ mm-mrad. A mismatch of $M = 30\%$ is applied in both planes. This means that the x co-ordinates are multiplied by 1.3, and the x' co-ordinates are divided by 1.3, leaving the emittance unchanged, and likewise for the y plane. (The space charge tune-shift is $\Delta Q_x = -0.165$). The distribution is left circulating in a ring for 200 turns, and afterwards the beam profile is shown by the black curve. The particle-in-cell simulation clearly shows that the beam distribution changes shape and readjusts to a different type.

8 Beam characterization

Unless something dramatic happens to the beam, the beam distribution does not change too quickly. Therefore, for a short time-scale, one can ask on how a beam distribution can be characterized.

Experimentally, this task is very difficult, as it necessitates a complete 4D phase space reconstruction, and in circular machines only the measurement of beam profiles is part of normal diagnostics, or the application of 2D tomographic techniques.

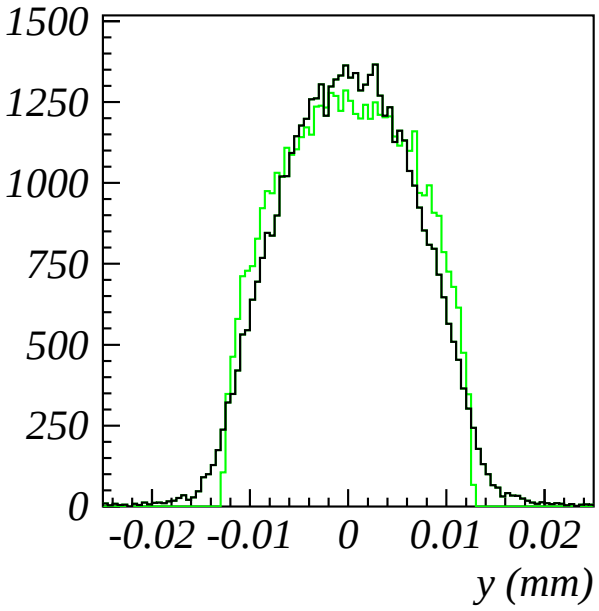


Fig. 6: Evolution of non-stationary beam distribution as obtained from a particle-in-cell simulation

From the point of view of multiparticle simulations, at each step of a simulation, the full multi-particle beam distribution is accessible, and a characterization of the beam is useful. The easier way to characterize the beam is through the second-order moments. If the centre of mass of the beam is in the origin, i.e., if $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i' = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} y_i' = 0$, with N the total number of macro-particles, the second-order moments

$$\begin{aligned} \langle x^2 \rangle &= \frac{1}{N}\sum_{i=1}^{N} x_i^2 , \\ \langle x'^2 \rangle &= \frac{1}{N}\sum_{i=1}^{N} x_i'^2 , \\ \langle xx' \rangle &= \frac{1}{N}\sum_{i=1}^{N} x_i x_i' \end{aligned} \tag{30}$$

enable the following quantities to be defined: the r.m.s. beam size $\tilde{x}$ is defined as

$$\tilde{x} = \sqrt{\langle x^2 \rangle} ,$$

and the r.m.s. beam emittance $\tilde{\varepsilon}_x$ is defined as

$$\tilde{\varepsilon}_x = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2} .$$

Similar moments and definitions apply for the y plane.

For matched beams, as we defined in this paper, there is a direct relation between the quantities $\mathscr{E}_x, \mathscr{E}_y$, the beam distribution type, and second order moments. By assuming that the 2D beam distribution has the general analytic shape

$$n(x,x',y,y') = \frac{1}{\pi^2 \mathscr{E}_x \mathscr{E}_y} \tilde{n}\left(\frac{\varepsilon_{0x}(x,x',s)}{\mathscr{E}_x} + \frac{\varepsilon_{0y}(y,y',s)}{\mathscr{E}_y} \right)$$

Table 2: Main properties of the three types of distribution. For KV, and WB distribution, the spatial distribution $n(x,y)$ is meant in the domain $x^2/a^2+y^2/b^2\le 1$.

	$n(x,x',y,y')$	$n(x,y)$	$a/\tilde{x}$	$\mathscr{E}_x/\tilde{\varepsilon}_x$
Kapchinsky–Vladimirsky	$\frac{1}{\pi^2\mathscr{E}_x\mathscr{E}_y}\delta\left(\frac{\varepsilon_x}{\mathscr{E}_x}+\frac{\varepsilon_y}{\mathscr{E}_y}-1\right)$	$\frac{1}{\pi ab}$	2	4
Waterbag	$\frac{2}{\pi^2\mathscr{E}_x\mathscr{E}_x}\Theta\left(1-\frac{\varepsilon_x}{\mathscr{E}_x}-\frac{\varepsilon_y}{\mathscr{E}_y}\right)$	$\frac{2}{\pi ab}\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right)$	$\sqrt{6}$	6
Gaussian	$\frac{1}{4\pi^2\mathscr{E}_x\mathscr{E}_y}\exp\left[-\frac{1}{2}\left(\frac{\varepsilon_x}{\mathscr{E}_x}+\frac{\varepsilon_y}{\mathscr{E}_y}\right)\right]$	$\frac{1}{2\pi ab}\exp\left[-\frac{1}{2}\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)\right]$	1	1

and ignoring at this point the possible s dependence of the type of distribution, we find characteristic relations between r.m.s. sizes $\tilde{x},\tilde{y}$, and sizes a,b, also between $\mathscr{E}_x,\mathscr{E}_y$ and the r.m.s. emittances $\tilde{\varepsilon}_x,\tilde{\varepsilon}_y$. These relations are shown in Table 2 for the three types of distribution here presented. The ratios $a/\tilde{x}$, and $\mathscr{E}_x/\tilde{\varepsilon}_x$ are important; in fact, we see that the relation between the edge of a beam and the r.m.s. size depends on the type of distribution that we are dealing with.

9 Envelope equation

As discussed in the previous section, the Kapchinsky–Vladimirsky distribution is stationary, that is, a Kapchinsky–Vladimirsky distribution will evolve while remaining a Kapchinsky–Vladimirsky type. From the previous section, we find that, once the type of a certain matched distribution is known, we need to know only two parameters to completely define all its properties. These parameters can be $\mathscr{E}_x,\mathscr{E}_y$, or a,b. Alternatively, r.m.s. quantities are also helpful to find these two parameters, for example, $\tilde{\varepsilon}_x,\tilde{\varepsilon}_y$, or $\tilde{x},\tilde{y}$.

For linear lattices, the Kapchinsky–Vladimirsky distribution is stationary; hence, we can describe the evolution of the beam distribution by characterizing the evolution of the chosen parameters. The r.m.s. emittances have the interesting property that they remain constant under linear forces. In fact, from the definition $\tilde{\varepsilon}_x^2=\langle x^2\rangle\langle x'^2\rangle-\langle xx'\rangle^2$, it is easy to verify that

$$\frac{\mathrm{d}\tilde{\varepsilon}_x^2}{\mathrm{d}s}=\frac{\mathrm{d}}{\mathrm{d}s}(\langle x'^2\rangle\langle x^2\rangle-\langle xx'\rangle^2)=2(\langle x'x''\rangle\langle x^2\rangle-\langle xx'\rangle\langle xx''\rangle)\,. \tag{31}$$

If the forces acting on each particle are linear; hence, if the equation of motion is $x''=\xi(s)x$ with $\xi(s)$ an arbitrary function, then using it in Eq. (31) yields $\mathrm{d}\tilde{\varepsilon}_x^2/\mathrm{d}s=0$.

Therefore, for linear forces $\tilde{\varepsilon}_x$ cannot be used to describe the evolution of the beam. We have to consider the r.m.s. beam sizes instead. From the definition of $\tilde{x}$ we find

$$\tilde{x}'=\frac{\langle xx'\rangle}{\tilde{x}}$$

and

$$\tilde{x}''=\frac{\langle xx''\rangle\tilde{x}^2+\langle x'^2\rangle\tilde{x}^2-\langle xx'\rangle^2}{\tilde{x}^3}\,. \tag{32}$$

By using the definition of the r.m.s. emittance, we find

$$\tilde{x}''=\frac{\langle xx''\rangle}{\tilde{x}}+\frac{\tilde{\varepsilon}_x^2}{\tilde{x}^3}\,. \tag{33}$$

As the equations of motion are linear,

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2}+\left[k_{0x}(s)-\frac{2K}{a(a+b)}\right]x=0\,, \tag{34}$$

we then find

$$\tilde{x}'' + k_{0x}(s)\tilde{x} - \frac{2K}{a(a+b)}\tilde{x} - \frac{\tilde{\varepsilon}_x^2}{\tilde{x}^3} = 0\,.$$

As we are discussing a Kapchinsky–Vladimirsky distribution, we can rewrite a, b as functions of the r.m.s. sizes and find

$$\tilde{x}'' + k_{0x}(s)\tilde{x} - \frac{K}{2(\tilde{x}+\tilde{y})} - \frac{\tilde{\varepsilon}_x^2}{\tilde{x}^3} = 0\,.$$

A similar equation holds for the y plane inverting $\tilde{x} \leftrightarrow \tilde{y}$. The fact that $\tilde{\varepsilon}_x$ and $\tilde{\varepsilon}_y$ are constant for a Kapchinsky–Vladimirsky distribution allows to use the r.m.s. envelope equations,

$$\tilde{x}'' + k_{0x}(s)\tilde{x} - \frac{K}{2(\tilde{x}+\tilde{y})} - \frac{\tilde{\varepsilon}_x^2}{\tilde{x}^3} = 0\,, \tag{35}$$

$$\tilde{y}'' + k_{0y}(s)\tilde{y} - \frac{K}{2(\tilde{x}+\tilde{y})} - \frac{\tilde{\varepsilon}_y^2}{\tilde{y}^3} = 0\,, \tag{36}$$

$$\tag{37}$$

to predict the evolution of $\tilde{x}, \tilde{y}$. As the Kapchinsky–Vladimirsky distribution does not change type, we can compute the evolution of the envelopes a, b, as $a = 2\tilde{x}, b = 2\tilde{y}$.

10 R.m.s. equivalent beams

If the time-scales considered are short enough, so that the distribution does not significantly change type, we can wonder if there exists a generalization of the r.m.s. envelope equations. We face here the problem of how to describe the evolution of a non-Kapchinsky–Vladimirsky distribution, which, therefore, necessarily yields non-linear space charge forces. We start here with the assumption that the beam distribution, i.e., the type of distribution, does not change. Again, the evolution of the beam is described by the two quantities that characterize the beam. The evolution of the r.m.s. moments requires evaluation of the quantity $\langle xx'' \rangle$ in Eq. (32), but now the equation of motion of a single particle becomes

$$\frac{\mathrm{d}^2 x}{\mathrm{d}s^2} + k_{0x}(s)x - Kx\int_0^\infty \frac{\hat{n}(\hat{T})}{(a^2+t)^{3/2}(b^2+t)^{1/2}}\mathrm{d}t = 0\,, \tag{38}$$

$$\tag{39}$$

where $\hat{T}$ is given by Eq. (15). Therefore,

$$\langle xx'' \rangle = -k_{0x}\langle x^2 \rangle + K\left\langle x^2 \int_0^\infty \frac{\hat{n}(\hat{T})}{(a^2+t)^{3/2}(b^2+t)^{1/2}}\mathrm{d}t \right\rangle,$$

and the last term becomes rather complicated to compute. However, Sacherer has proved [4] that for the class of beam distributions

$$\rho = \lambda \frac{1}{\pi ab}\hat{n}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right),$$

which are our matched beams, the following remarkable result holds:

$$K\left\langle x^2 \int_0^\infty \frac{\hat{n}(\hat{T})}{(a^2+t)^{3/2}(b^2+t)^{1/2}}\mathrm{d}t \right\rangle = \frac{K}{2}\frac{\tilde{x}}{\tilde{x}+\tilde{y}}\,.$$

This is valid independently of the function $\hat{n}(t)$. Therefore, Eq. (33) becomes

$$\tilde{x}'' = -k_{0x}(s)\tilde{x} + \frac{K}{2}\frac{1}{\tilde{x}+\tilde{y}} + \frac{\tilde{\varepsilon}_x^2}{\tilde{x}^3}\,. \tag{40}$$

This equation shows that the effect of the space charge in the term $\langle xx'' \rangle$ has always the same functional form, as for a Kapchinsky–Vladimirsky beam. This result is remarkable because it means that, independently of the type of beam distribution, if a beam is matched and the distribution type does not change, the r.m.s. envelope will evolve in the same way: this also means that if two beams have the same r.m.s. moments at given s, their r.m.s. envelopes will evolve in the same manner. These beams are called *r.m.s. equivalent*.

Clearly, this result is valid as long as the type of distribution does not change. If the r.m.s. emittance does not change, then the equations can be directly integrated; however, if r.m.s. emittance changes, it is then necessary to provide its evolution, as Eq. (40) alone will not be sufficient to predict the evolution of $\tilde{x}$.

11 Incoherent tune-shift

The tune of a single particle for a linear machine is defined as

$$Q_{0x} = \frac{1}{2\pi}\int_0^L \frac{1}{\beta_{0x}(s)}\mathrm{d}s\,,$$

and the Courant–Snyder theory shows that a localized gradient error of integrated strength Δk_x produces a change of tune as

$$\Delta Q_{0x} = \frac{1}{4\pi}\beta_{0x}(s)\Delta k_x\,.$$

The space charge in the centre of a beam acts exactly as a small defocusing gradient error. If a particle has small transverse amplitude (i.e., is mainly close to the beam centre), then the effect of the space charge on this particle when it goes from s to $s+\Delta s$, is given a local kick

$$\Delta x' = \hat{n}(0)\frac{2K}{a(s)[a(s)+b(s)]}x\Delta s\,,$$

in the same way, a local gradient Δk_x produces a kick $\Delta x' = -\Delta k_x x$. The positive sign of the kick stems from the space charge force, which is always defocusing. Therefore, the change of tune is then

$$\Delta Q_{0x} = -\frac{1}{4\pi}\beta_{0x}(s)\hat{n}(0)\frac{2K}{a(s)[a(s)+b(s)]}\Delta s\,.$$

By integration of all the distributed effect of space charge along the circumference, we obtain

$$\Delta Q_{0x} = -\frac{1}{4\pi}\int_0^L \beta_{0x}(s)\hat{n}(0)\frac{2K}{a(s)[a(s)+b(s)]}\mathrm{d}s\,.$$

Now we rewrite this equation as

$$\Delta Q_{0x} = -\frac{1}{4\pi}\frac{1}{\mathscr{E}_x}\int_0^L \frac{\beta_{0x}(s)}{\beta_x(s)}\hat{n}(0)\frac{2K}{1+\sqrt{\frac{\mathscr{E}_y\beta_y(s)}{\mathscr{E}_x\beta_x(s)}}}\mathrm{d}s\,.$$

If the space charge is not too strong, then $\beta_{0x}/\beta_x(s) \simeq 1$, and

$$\Delta Q_{0x} = -\frac{1}{4\pi}\frac{1}{\mathscr{E}_x}\hat{n}(0)2K2\pi R\left\langle \frac{1}{1+\sqrt{\frac{\mathscr{E}_y\beta_y(s)}{\mathscr{E}_x\beta_x(s)}}}\right\rangle_s\,,$$

where $\langle\rangle_s$ is the average over the machine length, and R is the average machine radius.

We next consider the decomposition of the beta functions around their average $\langle\beta_x\rangle_s$, and $\langle\beta_y\rangle_s$, as

$$\begin{aligned}\beta_x(s) &= \langle\beta_x\rangle_s + \Delta\beta_x(s)\,,\\ \beta_y(s) &= \langle\beta_y\rangle_s + \Delta\beta_y(s)\,,\end{aligned} \tag{41}$$

and we also consider the following auxiliary function

$$\mathscr{A}(x,y) = \frac{1}{1+\sqrt{\frac{\mathscr{E}_y y}{\mathscr{E}_x x}}}\,.$$

It is clear that

$$\frac{1}{1+\sqrt{\frac{\mathscr{E}_y \beta_y(s)}{\mathscr{E}_x \beta_x(s)}}} = \mathscr{A}(\beta_x(s),\beta_y(s))\,,$$

and $\mathscr{A}$ can be Taylor expanded around $\langle\beta_x\rangle_s, \langle\beta_y\rangle_s$, as

$$\mathscr{A}(\beta_x(s),\beta_y(s)) = \sum_{n,m} \partial_x^n \partial_y^m \mathscr{A}(x,y)|_{x=\langle\beta_x\rangle_s, y=\langle\beta_y\rangle_s} \frac{[\Delta\beta_x(s)]^n [\Delta\beta_y(s)]^m}{n!m!}\,.$$

Therefore,

$$\begin{aligned}\langle\mathscr{A}(\beta_x(s),\beta_y(s))\rangle_s &= \sum_{n,m} \partial_x^n \partial_y^m \mathscr{A}(x,y)|_{x=\langle\beta_x\rangle_s, y=\langle\beta_y\rangle_s} \frac{1}{n!m!}\langle[\Delta\beta_x(s)]^n[\Delta\beta_y(s)]^m\rangle_s \\ &= \mathscr{A}(\langle\beta_x\rangle_s,\langle\beta_y\rangle_s) + O(\langle\Delta\beta_x^2\rangle_s,\langle\Delta\beta_y^2\rangle_s)\,,\end{aligned} \tag{42}$$

because by definition $\langle\Delta\beta_x\rangle_s = \langle\Delta\beta_y\rangle_s = 0$. This means that

$$\left\langle \frac{1}{1+\sqrt{\frac{\mathscr{E}_y \beta_y(s)}{\mathscr{E}_x \beta_x(s)}}} \right\rangle_s = \frac{1}{1+\sqrt{\frac{\mathscr{E}_y \langle\beta_y\rangle_s}{\mathscr{E}_x \langle\beta_x\rangle_s}}} + O(\langle\Delta\beta_x^2\rangle_s,\langle\Delta\beta_y^2\rangle_s)\,.$$

This formula states that if the beta functions do not oscillate too much away from the average beta, then we can approximate the L.H.S. with the first term of the expansion and neglect the higher order terms in $\Delta\beta_x$, and $\Delta\beta_y$. This can be seen by the second term of the expansion, which is the quadratic in the average deviation from the average beta, and which becomes small for not too wild oscillation of the beta. Therefore, under these conditions, we obtain

$$\Delta Q_{0x} = -\frac{1}{\mathscr{E}_x}\hat{n}(0)K\frac{R}{1+\sqrt{\frac{\mathscr{E}_y\langle\beta_y\rangle_s}{\mathscr{E}_x\langle\beta_x\rangle_s}}} = -\hat{n}(0)K\langle\beta_x\rangle_s\frac{R}{\sqrt{\mathscr{E}_x\langle\beta_x\rangle_s}(\sqrt{\mathscr{E}_x\langle\beta_x\rangle_s}+\sqrt{\mathscr{E}_y\langle\beta_y\rangle_s})}\,.$$

In a similar way, we can now expand the beta function in the depressed tune formula,

$$Q_x = \frac{1}{2\pi}\int_0^L \frac{1}{\beta_x(s)}\mathrm{d}s\,,$$

and we find

$$Q_x = \frac{R}{\langle\beta_x\rangle_s} + \frac{R}{\langle\beta_x\rangle_s^3}\langle\Delta\beta_x^2\rangle_s + \ldots$$

Under the same approximation of not too wild beta oscillation, we can approximate $Q_x = R/\langle\beta_x\rangle_s$, and by substitution, we finally obtain

$$\Delta Q_{0x} = -\frac{R^2}{Q_x}\hat{n}(0)K\frac{1}{\sqrt{\mathscr{E}_x\langle\beta_x\rangle_s}(\sqrt{\mathscr{E}_x\langle\beta_x\rangle_s}+\sqrt{\mathscr{E}_y\langle\beta_y\rangle_s})}\,. \tag{43}$$

This formula is very intuitive, and has a deep meaning: the incoherent space charge tune-shift appears to be insensitive to the machine optics, but to depend to an 'equivalent' smooth accelerator structure with the same tunes as the original one.

Table 3: Factor f as function of beam distribution.

	Kapchinsky–Vladimirsky	Waterbag	Gaussian
f	1	4/3	2

12 Space charge tune-shift for r.m.s. equivalent beams

The last formula gives the tune-shift as a function of the quantities $\mathscr{E}_x, \mathscr{E}_y$. For practical uses, it is useful to modify Eq. (43) using r.m.s. emittances. First, we observe that for a matched beam

$$\tilde{x}^2 = \mathscr{E}_x \beta_x(s) \frac{1}{2} \int_0^\infty t\hat{n}(t)\mathrm{d}t\,,$$

and the r.m.s. emittance is

$$\tilde{\varepsilon}_x = \mathscr{E}_x \frac{1}{2} \int_0^\infty t\hat{n}(t)\mathrm{d}t\,.$$

Therefore, we can rewrite the tune-shift as

$$\Delta Q_{0x} = -\frac{R^2}{Q_x}\hat{n}(0)K\frac{1}{2}\int_0^\infty t\hat{n}(t)\mathrm{d}t \frac{1}{\sqrt{\tilde{\varepsilon}_x\langle\beta_x\rangle_s}(\sqrt{\tilde{\varepsilon}_x\langle\beta_x\rangle_s}+\sqrt{\tilde{\varepsilon}_y\langle\beta_y\rangle_s})}\,,$$

and define a peak tune-shift as

$$\Delta Q_{0x} = f\Delta\tilde{Q}_{0x} \tag{44}$$

where

$$\Delta\tilde{Q}_{0x} = -\frac{R^2}{Q_x}\frac{K}{4}\frac{1}{\sqrt{\tilde{\varepsilon}_x\langle\beta_x\rangle_s}(\sqrt{\tilde{\varepsilon}_x\langle\beta_x\rangle_s}+\sqrt{\tilde{\varepsilon}_y\langle\beta_y\rangle_s})}\,, \tag{45}$$

is the tune-shift of a Kapchinsky–Vladimirsky beam r.m.s. equivalent to the beam we are considering. The factor f incorporates the type of the distribution, and is

$$f = 2\hat{n}(0)\int_0^\infty t\hat{n}(t)\mathrm{d}t\,.$$

The values of f according to the type of distribution are shown in Table 3. These results show that r.m.s. equivalent beams having the same perveance will produce different peak tune-shift according to the type of distribution.

13 Space charge limit

The previous results allow us to discuss which is the maximum current or, equivalently, the maximum number of particles a coasting beam may have in an accelerator. The limiting factor is the maximum allowed incoherent space charge tune depression, which is dictated by the necessity of avoiding the overlap of the tune-spread with machine resonances. We proceed by discussing the horizontal plane, but similar formulae are obtained on exchanging the x and y planes. As the resonances up to the fourth order may be dangerous for beam survival, it is then assumed that the space charge limit is set by the condition

$$|\Delta Q_{0x}| \leq |\Delta Q_{xl}| \simeq 0.25\,,$$

see Ref. [3]. This constraint can be used in Eq. (44); inverting this equation, we obtain the maximum longitudinal particle density, which reads

$$\frac{\mathrm{d}N}{\mathrm{d}s} = \frac{\lambda}{eZ} = \frac{8\pi\varepsilon_0 m_u A\gamma^3 v^2}{e^2 Z^2}\frac{|\Delta Q_{xl}|}{f}\frac{\tilde{\varepsilon}_x}{R}\left(1+\sqrt{\frac{\langle\beta_y\rangle_s\tilde{\varepsilon}_y}{\langle\beta_x\rangle_s\tilde{\varepsilon}_x}}\right)\,, \tag{46}$$

where A is the mass number, m_u the nucleon mass, Z is the charge state of the particle, and e is the elementary charge.

The same argument applies to bunched beams although we did not discuss the space charge field of a bunched beam. In fact, if a bunch is very long with respect to the transverse size, then the local coasting beam approximation enables the 'local' transverse tune-shift to be computed according to the local charge-line density $\lambda(z)$. In complete analogy, we can speak of the local current $I(z)$, which is largest where the local particle density is the largest. Therefore, one finds the peak current I_{peak} in the centre of the bunch ($z = 0$). To compare how different a bunched beam is from a coasting beam, it is customary to compare the peak current of a bunched beam I_{peak} with the current of a coasting beam composed of the same number of particles I_{average}. The ratio $B_f = I_{\text{average}}/I_{\text{peak}}$, called bunching factor, quantifies the difference between bunched beams and coasting beams. Clearly, the bunching factor is depending on the longitudinal beam distribution, and on how many bunches are accommodated in a circular machine. If $B_f = 1$ then the beam is un-bunched; the more the longitudinal bunch distribution is peaked, the smaller is B_f. As the largest transverse tune-spread is at the bunch centre, for that transverse section, the space charge limit is given by Eq. (46). The maximum number of particles in the ring to reach the space charge limit (hence the maximum average current), is found by multiplying Eq. (46) by B_f and integrating along the circumference:

$$N_{\text{tot}} = B_f \frac{4(2\pi)^2 \varepsilon_0 m_u A \gamma^3 v^2}{e^2 Z^2} \frac{|\Delta Q_{xl}|}{f} \tilde{\varepsilon}_x \left(1 + \sqrt{\frac{\langle \beta_y \rangle_s \tilde{\varepsilon}_y}{\langle \beta_x \rangle_s \tilde{\varepsilon}_x}}\right) . \tag{47}$$

From Eq. (47), it appears evident which are the critical parameters to increase the total number of particles stored in a ring: the bunching factor B_f, and the type of transverse distribution, i.e., the parameter f. By increasing B_f, and making a beam closer to a Kapchinsky–Vladimirsky distribution, if possible, N_{tot} will be increased. Alternatively one should compensate all natural resonances so that $|\Delta Q_{xl}|$ can be taken larger, or increase the energy of the beam to reach higher γ.

14 Oscillation of mismatched beams

Consider a low-intensity 2D beam matched with the machine optics at the longitudinal position s_0, see left panel of Fig. 7. Suppose now that we distort the beam, as shown in the right panel of Fig. 7. The black ellipse in the right panel of Fig. 7b marks the initial edge of the beam before distortion. The stretched phase space now has two 'wings', which we call A and B. If we follow the beam evolution, each particle rotates with instantaneous phase advance $1/\beta_x(s)$. While the beam moves ahead (increases s), the beam ellipse rotates in the co-moving reference frame preserving its area. The two beam wings out of the initial beam black ellipse, marked A and B, will also rotate as s increases, as indicated by the red curved arrows.

The beam envelope is found by taking the outermost particle, and with this definition for this type of mismatched beam the envelope cannot distinguish between the two beam wings. Consequently, if the beam ellipse rotates through 180°, the two wings will swap (A $\leftrightarrow$ B), but the beam envelope will remain the same. Therefore, the beam envelope oscillates with a frequency that is the double the frequency of the single particle, i.e., the beam envelope will oscillate with a phase $2\pi(2Q_{x0})$ per turn.

This argument merely considers the single-particle dynamics, but the effect of space charge on a mismatched beam is more subtle.

To visualize the effect of the mismatch, we consider the quantity $\Delta x(s) = a(s) - a_0(s)$, where with $a_0(s)$ we indicate the matched envelope at location s, whereas with $a(s)$ we indicate the 'mismatched envelope' at location s. Figure 8 shows the evolution of $\Delta x(s)$ as obtained from a multiparticle simulation for two beams, one with low intensity (green), and one with high intensity (red). Both beams are mismatched by the same amount. The picture shows clearly that the high-intensity mismatched beam has a longer wavelength than the low-intensity one, which seems to indicate that the depressed tunes play a role in creating this shift. Therefore, at first sight, one would use the same argument as illustrated

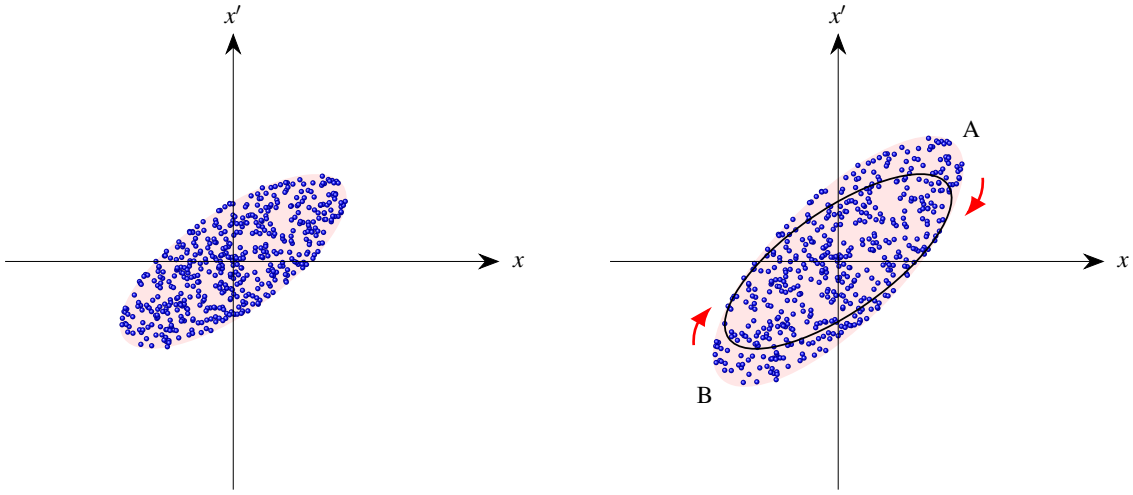


Fig. 7: Dynamics of mismatched beam. Left-hand side: initial matched beam distribution. Right-hand side: mismatched beam. The mismatch is produced by creating the 'wings' A and B.

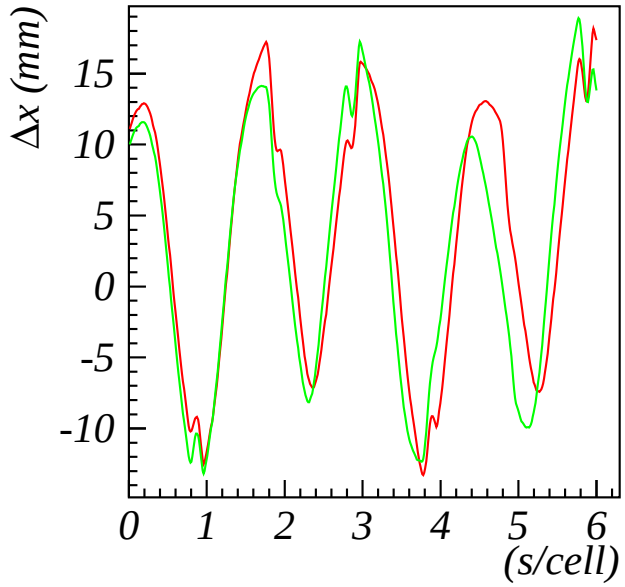


Fig. 8: Dynamics of mismatched beam. Green: beam with low intensity. Red: beam with high intensity

in Fig. 7 to discuss the dynamics of a mismatched high-intensity beam. As the space charge will depress the single-particle tune, one would conclude that the beam envelope will make $2Q_x$ oscillations per turn, where Q_x is the depressed tune. However, this is not true.

15 Coherent frequencies

To explain the effect of space charge on a mismatched beam, we need to start from the envelope equations. For this discussion, it is easier to use a constant-focusing lattice, and for convenience we rescaled the r.m.s. envelope to the matched solution as $\tilde{x} = \hat{x}\tilde{x}_0, \tilde{y} = \hat{y}\tilde{y}_0$. Here $\tilde{x}_0, \tilde{y}_0$ are the r.m.s. matched

envelopes for a Kapchinsky–Vladimirsky beam with r.m.s. emittances $\tilde{\varepsilon}_x, \tilde{\varepsilon}_y$ and perveance K. In these terms, the envelope equations take the form

$$\hat{x}'' + k_{0x}(s)\hat{x} - \frac{K}{2}\frac{1}{\tilde{x}_0(\tilde{x}_0\hat{x}+\tilde{y}_0\hat{y})} - \frac{\tilde{\varepsilon}_x^2}{\tilde{x}_0^4\hat{x}^3} = 0\,, \tag{48}$$

$$\hat{y}'' + k_{0y}(s)\hat{y} - \frac{K}{2}\frac{1}{\tilde{y}_0(\tilde{x}_0\hat{x}+\tilde{y}_0\hat{y})} - \frac{\tilde{\varepsilon}_y^2}{\tilde{y}_0^4\hat{y}^3} = 0\,. \tag{49}$$

If $\hat{x} = \hat{y} = 1$, the beam is matched. If the beam is mismatched, then we can define $\delta\hat{x} = \hat{x} - 1$ and $\delta\hat{y} = \hat{y} - 1$, and Eqs. (48) and (49) can be expanded around the matched solution. After some algebra, and neglecting quadratic terms of the incoherent tune-shift, we find:

$$\begin{aligned} \delta\hat{x}'' + a_{xx}\delta\hat{x} + a_{xy}\delta\hat{y} = 0\,,\\ \delta\hat{y}'' + a_{yx}\delta\hat{x} + a_{yy}\delta\hat{y} = 0\,, \end{aligned} \tag{50}$$

with

$$\begin{aligned} a_{xx} &= 4\left(\frac{Q_{0x}}{R}\right)^2 - \frac{2Q_{0x}\Delta Q_x}{R^2}\left(\frac{\tilde{x}_0}{\tilde{x}_0+\tilde{y}_0} - 3\right),\\ a_{xy} &= -\frac{2Q_{0x}\Delta Q_x}{R^2}\frac{\tilde{y}_0}{\tilde{x}_0+\tilde{y}_0},\\ a_{yx} &= -\frac{2Q_{0y}\Delta Q_y}{R^2}\frac{\tilde{x}_0}{\tilde{x}_0+\tilde{y}_0},\\ a_{yy} &= 4\left(\frac{Q_{0y}}{R}\right)^2 - \frac{2Q_{0y}\Delta Q_y}{R^2}\left(\frac{\tilde{y}_0}{\tilde{x}_0+\tilde{y}_0} - 3\right). \end{aligned} \tag{51}$$

These are two linearly coupled second-order differential equations with constant coefficients. By making a proper linear co-ordinate transformation $(\delta\hat{x}, \delta\hat{y}) \to (\xi, \nu)$, the system in Eq. (50) can be decoupled into two independent differential equations,

$$\begin{aligned} \xi'' + \lambda_+\xi = 0\,,\\ \nu'' + \lambda_-\nu = 0\,, \end{aligned} \tag{52}$$

with $\lambda_\pm = Q_{\pm,\mathrm{coh}}^2/R^2$ the eigenvalues of the matrix a_{ij} defining Eq. (50). For $\tilde{x}_0 \neq \tilde{y}_0$, in good approximation

$$\begin{aligned} Q_{+,\mathrm{coh}} &= 2Q_{0x} - \frac{1}{2}\Delta Q_x\left(\frac{\tilde{x}_0}{\tilde{x}_0+\tilde{y}_0} - 3\right),\\ Q_{-,\mathrm{coh}} &= 2Q_{0y} - \frac{1}{2}\Delta Q_y\left(\frac{\tilde{y}_0}{\tilde{x}_0+\tilde{y}_0} - 3\right). \end{aligned}$$

Therefore, we find that the oscillations of the envelopes around the matched solutions are obtained by the compositions of the modes with tunes $Q_{\pm,\mathrm{coh}}$ because of inverting $(\xi, \nu) \to (\delta\hat{x}, \delta\hat{y})$. We observe that $Q_{\pm,\mathrm{coh}}$ can be re-cast in a different form as

$$Q_{+,\mathrm{coh}} = 2Q_x - \frac{1}{2}\Delta Q_x\left(\frac{\tilde{x}_0}{\tilde{x}_0+\tilde{y}_0} + 1\right), \tag{53}$$

$$Q_{-,\mathrm{coh}} = 2Q_y - \frac{1}{2}\Delta Q_y\left(\frac{\tilde{y}_0}{\tilde{x}_0+\tilde{y}_0} + 1\right), \tag{54}$$

where Q_x, Q_y are the depressed tunes by the incoherent space charge tune-shift, as previously discussed. These last two equations show that the modes oscillate with an extra term to the $2Q_x, 2Q_y$ which has a

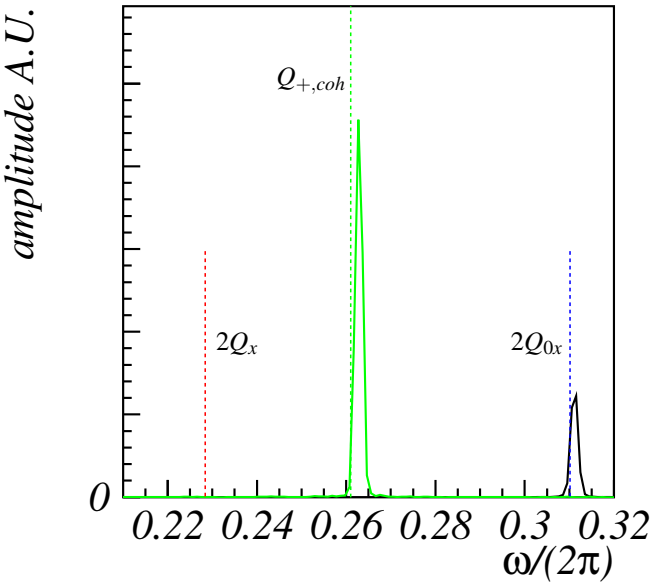


Fig. 9: Spectrum of oscillations of $\tilde{x}$. Black: low intensity. Green: high intensity. Dashed lines are drawn to theoretical predictions.

coherent nature, and act in the opposite direction of the incoherent tune-shift, increasing the frequency of the mode. (The sign of $\Delta Q_x, \Delta Q_y$ is negative).

Figure 9 shows the Fourier spectrum of the r.m.s. envelope of a mismatched beam. The beam is tracked in an arbitrarily chosen constant-focusing lattice of length $L = 110$ m, with tunes $Q_{0x} = 2.655, Q_{0y} = 2.823$. The incoherent space charge tune-shift is $\Delta Q_{0x} = -4 \times 10^{-2}$, $\Delta Q_y = -5.6 \times 10^{-2}$. The dashed lines show the theoretical predictions. The black spectrum is obtained by tracking a mismatched beam with no space charge, and we find it peaked on $2Q_{0x}$, as expected from the argument of the previous section. However, when the space charge is activated, the spectrum shown by the green solid curve is now peaked on the coherent frequency as predicted by the theory of Eq. (53). Note that the frequency $2Q_x$ is not excited.

This analysis is based on the r.m.s. envelope equations, and more modes of oscillation exist. A full discussion of these modes based on Vlasov equation can be found in Ref. [5]; a visual representation is shown in Fig. 10.

16 Tune-spread: the non-linear region

The previous analysis of the oscillations of the mismatched beam relies on the envelope equations. However, these equations do not consider the actual tune that each particle experiences. We already discussed, in Section 5, the dynamics in the linear region of the space charge field, and found a deviation of the single-particle tune from the bare tunes, which we called the incoherent space charge tune-shift $\Delta Q_{0x}, \Delta Q_{0y}$. The effect on the dynamics of the *non-linear field regime* introduced in Section 4 is still to be discussed. A first effect of the non-linear field will be to change the single-particle tunes by a different amount from $\Delta Q_{0x}, \Delta Q_{0y}$. The best way to visualize this effect is to plot, for each beam particle, its tune's deviation from the machine bare tunes. This may be computed with a standard procedure by analysing the frequency of each particle using a turn-by-turn data method [6]. The set of all these incoherent tunes is called space charge *tune-spread*. Figure 11 shows the space charge tune-spread for the three distributions presented.

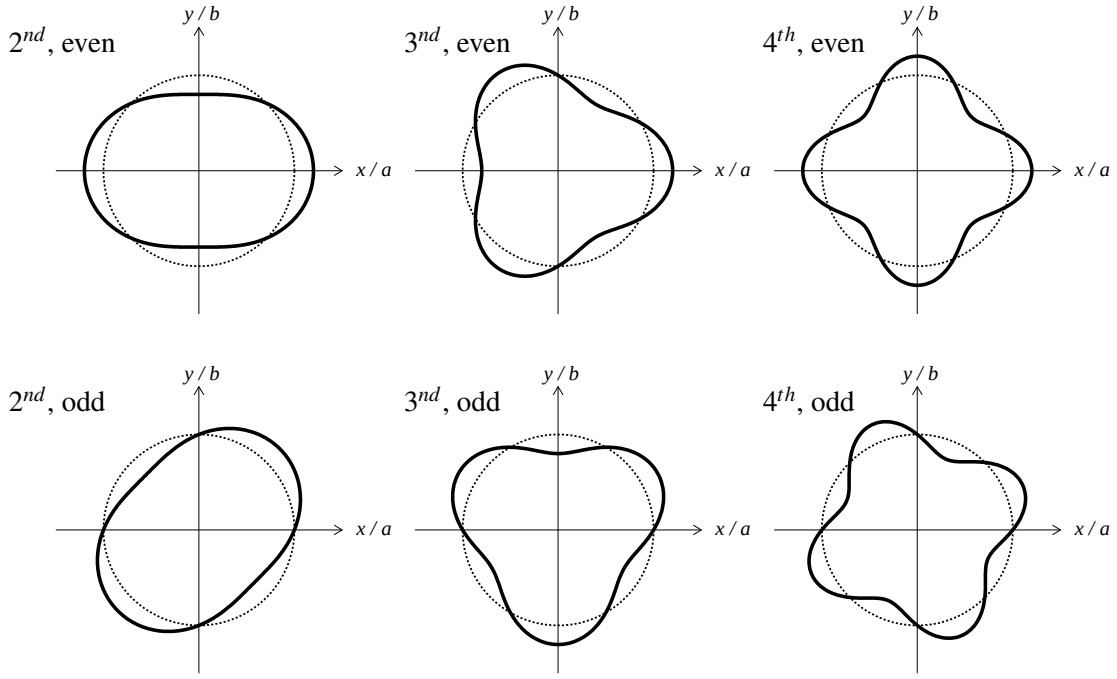


Fig. 10: Even and odd modes for the second, third, and fourth orders

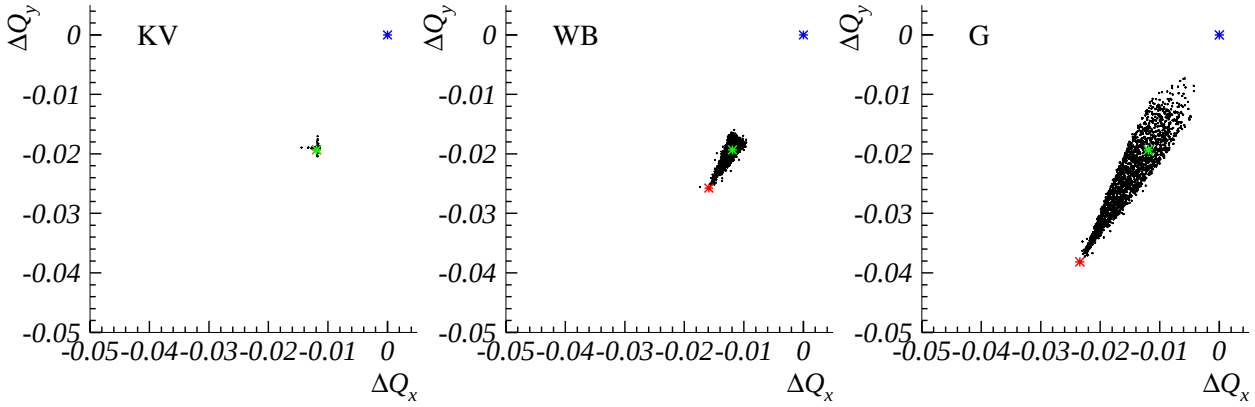


Fig. 11: Tune-spread of main types of particle distributions. Left-hand side: Kapchinsky–Vladimirsky (KV). Centre: waterbag (WB). Right-hand side: Gaussian (G). Blue: bare tune. Green: Kapchinsky–Vladimirsky tune-shift. Red: tune-shift for (centre) waterbag or (right-hand side) Gaussian.

The left-hand panel shows the tune-spread of a Kapchinsky–Vladimirsky distribution. Each single-particle tune is a black dot. The blue marker shows the machine bare tune, and the green marker shows incoherent detuning, as predicted by the theoretical formula of Eq. (43). The figure clearly shows that all single-particle tunes are, to a good approximation, overlapping with the theoretical tune. This happens because the electric field generated by the Kapchinsky–Vladimirsky is linear everywhere inside the beam distribution.

The panel in the centre shows the tune-spread for a 2D waterbag beam, this beam has the same perveance as the Kapchinsky–Vladimirsky beam, but the sizes are chosen so that it is r.m.s. equivalent. The set of tunes is now not overlapping and is spread over a triangular shape. The red marker shows the incoherent space charge tune-shift, as computed using Eq. (43). The factor f is now 4/3, and is responsible for the increase of the tune-shift.

The panel on the right shows the space charge tune-spread for a Gaussian distribution, still r.m.s. equivalent to the Kapchinsky–Vladimirsky beam. This graph shows that the single-particle tunes are spread over a larger area, and the density of tunes decreases as the tunes approach the machine bare tune. Now $f = 2$, and the graph shows that the distance between the green and blue markers is equal to the distance between the red and the green marker.

17 Amplitude-dependent detuning

Figure 11 shows that all particles experience the same space charge tune-shift only for the Kapchinsky–Vladimirsky beam. For the other distributions, waterbag and Gaussian, the non-linear field creates different tunes according to how particles are distributed. It, therefore, makes sense to ask if there is any relation between the particle position and the detuning experienced by that particle. An intuitive argument helps qualitatively in understanding what happens: when a particle has a large single-particle emittance $\varepsilon_x, \varepsilon_y$, i.e., large oscillation amplitudes, this particle following the betatron motion travels periodically through the beam core and also through the beam tails; therefore, on 'average', this particle experiences a weaker electric field than those particles with small amplitudes that always stayed close the beam centre. Hence, the space charge tune-shift of a particle with large amplitudes will be smaller than $\Delta Q_{0x}, \Delta Q_{0y}$. Therefore, the larger the particle amplitude, the closer the particle tune is to the bare machine tune because it is as if the space charge is not contributing to the single-particle dynamics. In the left panel of Fig. 12, we plot all single-particle tunes of the right panel of Fig. 11 as a function of the X amplitude re-scaled with the beam r.m.s. size. It can be seen that all the black dots are well contained within a red curve of simple functional dependence as

$$\Delta Q_x \simeq \Delta Q_{x,\max} \frac{1}{1 + \left(\frac{X}{2\sigma_x}\right)^2} .$$

The particle tunes are also found above the red curve, because of the vertical transverse amplitudes: even if a particle has zero horizontal amplitude, the vertical amplitude can be large, hence the horizontal transverse electric field will be diminished, with consequent reduction of the space charge tune-shift. The right-hand panel of Fig. 12 shows the tunes of a few particles with initial $\varepsilon_y \simeq 0$. These black markers overlap the red curve to good approximation.

There are more complex formulae to describe the relation between single-particle emittances (or particle action) and space charge tune-shift. For example, see Ref. [7].

18 Mismatched beams: free energy and emittance growth

Till now we have discussed the effect of space charge on the frequencies of particle oscillations in a beam, and on the coherent oscillation frequencies of the beam envelope. In those analyses, the starting point is from a matched beam whose properties do not change with time: this also means that all the different

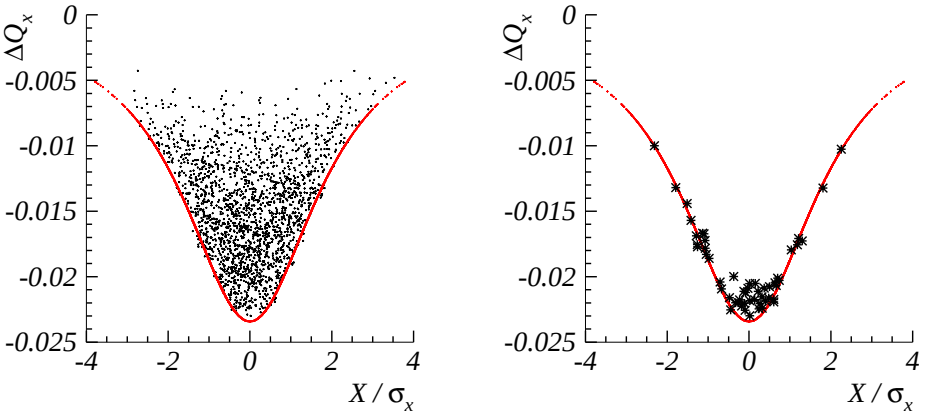


Fig. 12: Distribution of single-particle incoherent tune-shift in Gaussian 2D beam as function of horizontal particle amplitude. Right-hand side: only particles with $\varepsilon_y = 0$ are shown.

forms of beam energy are stationary, i.e., the beam is in 'equilibrium', and it has been assumed that any perturbation of the beam creates an oscillatory behaviour around the stationary solution. However, the possibility that the perturbation might modify the original stationary state of the beam has not been addressed.

We now discuss the possibility that, when brought out of equilibrium, a beam may evolve to a new stationary state. This discussion is developed using the concept of free energy (see in Ref. [3]). When a beam is in a stationary state, there is no free energy but when the beam is significantly perturbed, free energy is created, and this energy, assisted by other mechanisms, will re-distribute between all forms of energy, taking the beam into a new stationary state.

We make a discussion for a Kapchinsky–Vladimirsky circular beam transported in an axi-symmetric constant-focusing lattice. In this case, $\tilde{x} = \tilde{y}$ and, for convenience, we use the envelope $a = 2\tilde{x}$. As the beam is circular, we drop the index x or y. For convenience we also use the full beam emittance $\mathscr{E} = 4\tilde{\varepsilon}$.

We recall that a stationary a beam distribution is characterized by the condition $a'' = 0$, which means

$$k_0 a - \frac{K}{a} - \frac{\mathscr{E}^2}{a^3} = 0\,.$$

It is also convenient to define the depressed focusing strengths as in Eq. (25), which now reads

$$k = k_0 - \frac{K}{a^2}\,. \tag{55}$$

We now compute the energy content of a beam when it is in a stationary state. There are three different form of energy:

1. transverse kinetic energy;
2. potential energy of particles in the potential of the lattice;
3. potential energy of the Coulomb field created by the particle distribution.

18.1 Transverse kinetic energy

The transverse kinetic energy of one particle is $\frac{1}{2}m\gamma(v_x^2 + v_y^2)$; therefore, the average kinetic energy of one particle is

$$E_{\rm k} = \frac{1}{2}m\gamma(\langle v_x^2\rangle + \langle v_y^2\rangle)\,,$$

and by using the definitions $v_x = vx'$, $v_y = vy'$ we find

$$E_\mathrm{k} = \frac{1}{2} m\gamma v^2 (\langle x'^2 \rangle + \langle y'^2 \rangle) .$$

As the beam is axi-symmetric, $\langle x'^2 \rangle = \langle y'^2 \rangle$,

$$E_\mathrm{k} = m\gamma v^2 \langle x'^2 \rangle .$$

As the beam is stationary, the equation of motion of a single particle in the beam is $x'' + kx = 0$, and consequently 'depressed optics' is generated, as discussed in Section 6. For this lattice, we find that the modified optics is given by the beta function, $\beta_x = 1/\sqrt{k}$, and $\alpha_x = 0$. For a matched beam, the second-order moments can be related to the depressed optics according to

$$\begin{aligned} \langle x^2 \rangle &= \beta_x \tilde{\varepsilon}_x , \\ \langle x'^2 \rangle &= \gamma_x \tilde{\varepsilon}_x , \\ \langle xx' \rangle &= -\alpha_x \tilde{\varepsilon}_x , \end{aligned} \tag{56}$$

with $\tilde{\varepsilon}_x$ the r.m.s. emittance. In particular, for the constant-focusing case we find $\langle x'^2 \rangle = \frac{1}{\beta_x} \tilde{\varepsilon}_x = \frac{1}{\beta_x^2} \beta_x \tilde{\varepsilon}_x = \frac{1}{\beta_x^2} \langle x^2 \rangle = k \langle x^2 \rangle$. Therefore, the kinetic energy per particle is

$$E_\mathrm{k} = m\gamma v^2 k \langle x^2 \rangle . \tag{57}$$

18.2 Potential energy

The transverse potential energy due to the lattice alone is computed by writing the equation of motion in the time domain; in one plane the equation becomes

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} m\gamma x = m\gamma v^2 k_0 x ,$$

a similar equation is found for the vertical plane with the substitution $x \to y$. Therefore, the potential energy of one particle is then

$$m\gamma v^2 k_0 \left(\frac{x^2}{2} + \frac{y^2}{2} \right) .$$

From this relation, we compute the average potential energy per particle as

$$E_\mathrm{p} = m\gamma v^2 k_0 \langle x^2 \rangle .$$

18.3 Field energy

The field energy is the energy necessary to create the beam, i.e., to create the configuration of charges and currents. It consists of two densities of field energy,

$$\frac{\varepsilon_0}{2} \vec{E}^2 , \qquad \frac{1}{2\mu_0} \vec{B}^2 ,$$

the electric and the magnetic field energy. The total field energy is computed by integrating the density of field energy over a volume of length l and extending integration to the beam pipe, of radius R_p. The electric field for a circular Kapchinsky–Vladimirsky coasting beam is given by Eq. (11),

$$E_r(s) = \begin{cases} \frac{\rho(s)}{2\varepsilon_0} r & \text{if} \quad r \le a , \\ \frac{\rho(s) a^2(s)}{2\varepsilon_0} \frac{1}{r} & \text{if} \quad r \ge a . \end{cases} \tag{58}$$

Therefore,

$$\int \frac{\varepsilon_0}{2}\vec{E}^2 \mathrm{d}V = \frac{\rho^2 a^4 \pi l}{4\varepsilon_0}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

By using the beam current I, we find $I = Av\rho = \pi a^2 v\rho$, which yields

$$\int \frac{\varepsilon_0}{2}\vec{E}^2 \mathrm{d}V = \frac{I^2 l}{4\pi\varepsilon_0 v^2}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

The magnetic field B created by a circular 2D Kapchinsky–Vladimirsky beam is given by the Biot–Savart law, namely

$$B_\theta(s) = \begin{cases} \frac{\mu_0 I}{2\pi a^2} r & \text{if} \quad r \le a\,, \\ \frac{\mu_0 I}{2\pi}\frac{1}{r} & \text{if} \quad r \ge a\,. \end{cases} \tag{59}$$

Therefore,

$$\int \frac{1}{2\mu_0}\vec{B}^2 \mathrm{d}V = \frac{\mu_0 I^2 l}{4\pi}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

Now using the classical relation $c^2 = 1/(\varepsilon_0\mu_0)$, we find

$$\int \frac{1}{2\mu_0}\vec{B}^2 \mathrm{d}V = \frac{I^2 l}{4\pi\varepsilon_0 c^2}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

The total field energy in the volume V is given by the difference of the electric and magnetic energy (see Ref. [3] for more details),

$$E_\mathrm{fV} = \frac{I^2 l}{4\pi\varepsilon_0 c^2}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right]\left(\frac{1}{\beta^2} - 1\right).$$

Now the number of particles in our integration volume is

$$N_\mathrm{p} = \frac{\rho}{q}\pi a^2 l = \frac{Il}{qv};$$

therefore, the field energy per particle is

$$E_s = qv\frac{I}{4\pi\varepsilon_0 c^2}\left[\frac{1}{4} + \ln\left(\frac{R_\mathrm{p}}{a}\right)\right]\left(\frac{1}{\beta^2} - 1\right). \tag{60}$$

Recalling the definition of perveance, Eq. (19) ,

$$K = \frac{qI}{2\pi\varepsilon_0 m\gamma^3\beta^3 c^3},$$

and replacing qI with the perveance in Eq. (60), we find

$$E_s = Km\gamma\frac{v^2}{8}\left[1 + 4\ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

Now using the fact that the beam is stationary, from Eq. (55), we obtain $K = (k_0 - k)a^2$; that is, we find that the field energy per particle is

$$E_s = (k_0 - k)a^2 m\gamma\frac{v^2}{8}\left[1 + 4\ln\left(\frac{R_\mathrm{p}}{a}\right)\right].$$

18.4 Total energy, free energy, and emittance growth

The total energy per particle of a stationary beam is the sum of these three types of energy as obtained in the previous three subsections, namely

$$E_n = E_\mathrm{k} + E_\mathrm{p} + E_s$$

which yields

$$E_n = \frac{1}{4} m\gamma v^2 \left\{ ka^2 + k_0 a^2 + (k_0 - k) a^2 \frac{1}{2} \left[1 + 4 \ln\left(\frac{R_\mathrm{p}}{a} \right) \right] \right\}, \tag{61}$$

where here we substituted $\langle x^2 \rangle = 4a^2$.

Let us consider a stationary beam characterized by an *initial* envelope a_i and emittance $\mathscr{E}_i$, hence, by an energy per particle $E_{n,i}$. Now let's add to this beam an extra energy per particle ΔE_n: this free energy will make the the beam un-stationary, and the beam envelope will perform oscillations. The principle here invoked is that, due to such additional mechanisms as non-linear or stochastic forces, the free energy ΔE_n will thermalize. This means that it will spontaneously become 'equally' distributed in all allowed energy forms; hence, the beam will relax into a new *final* stationary state characterized by a new beam size a_f, and a new emittance $\mathscr{E}_f$, hence in a new stationary state with energy per particle E_{nf}. We therefore find that $E_{nf} = E_{ni} + \Delta E_n$, where E_{nf} is obtained by substituting $k \to k_f, a \to a_f$ into Eq. (61), and E_{ni} is obtained by substituting $k \to k_i, a \to a_i$ into Eq. (61). It is convenient to give the free energy per particle as a function of a dimensionless parameter h, as follows

$$\Delta E_n = \frac{1}{2} m\gamma v^2 k_0 a_i^2 h\,. \tag{62}$$

The relation $E_{nf} = E_{ni} + \Delta E_n$ then becomes

$$\frac{a_f^2}{a_i^2} - 1 - \left(1 - \frac{k_i}{k_0} \right) \ln \frac{a_f}{a_i} = h\,, \tag{63}$$

this relation yields a_f/a_i as a function of h. Now we can find the emittance growth from the envelope growth using the straightforward relation

$$\frac{\mathscr{E}_f}{\mathscr{E}_i} = \frac{a_f}{a_i} \left[1 + \frac{k_0}{k_i} \left(\frac{a_f^2}{a_i^2} - 1 \right) \right]^{1/2}. \tag{64}$$

Figure 13 shows the emittance growth as a function of the free-energy parameter h. There are ten curves for ten different tune depressions: from $Q_x/Q_{0x} = 0.1$, to $Q_x/Q_{0x} = 1$ (i.e., for no space charge). The graph shows that the same free-energy parameter h yields a larger emittance growth, the larger the tune depression.

We now discuss how the free energy is created by mismatching a stationary beam. Consider a matched beam with envelope a_i, and now let's mismatch it so that the envelope becomes a_m but preserving the initial beam emittance; let's now call $M = a_m/a_i$ the mismatch factor. Identifying $k_i/k_0 = (Q_{ix}/Q_{0x})^2$, we find that the free-energy parameter h reads

$$h = \frac{1}{2} \left(\frac{Q_{ix}}{Q_{0x}} \right)^2 \left(\frac{1}{M^2} - 1 \right) - \frac{1}{2} \left(1 - M^2 \right) + \left[1 - \left(\frac{Q_{ix}}{Q_{0x}} \right)^2 \right] \ln \left(\frac{1}{M} \right). \tag{65}$$

This relation allows us to compute h. We make an example of how to use it. Figure 14 shows a simulation with $Q_{ix}/Q_{0x} = 0.962$, with a beam mismatched by a factor $M = 1.3$. By using Eq. (65), we find the free energy parameter $h = 0.1347$, and the free-energy limit of emittance growth for this value of h is obtained from Eqs. (63) and (64). We find $\mathscr{E}_f/\mathscr{E}_i = 1.145$, which is drawn in Fig. 14 as a dashed green horizontal line. The black and red curves are the horizontal and vertical re-scaled r.m.s. emittances, as obtained from a particle-in-cell simulation. The green curve is the average between them. The simulation shows that the green curve approaches the free-energy limit to a good approximation.

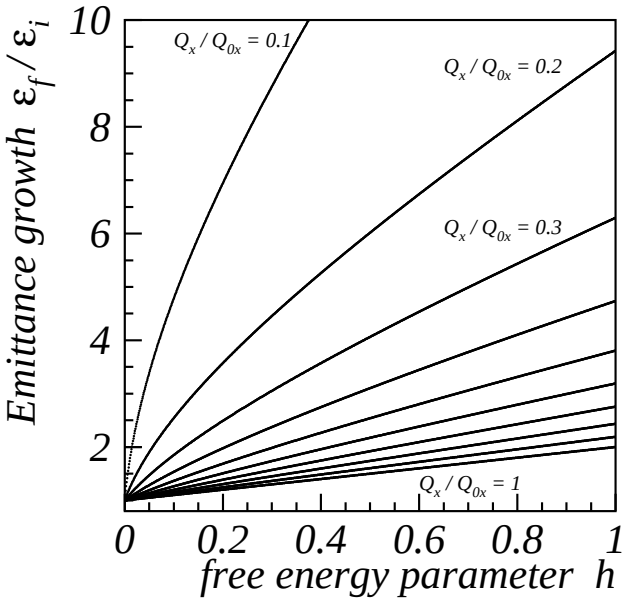


Fig. 13: Emittance growth as function of free parameters *h*

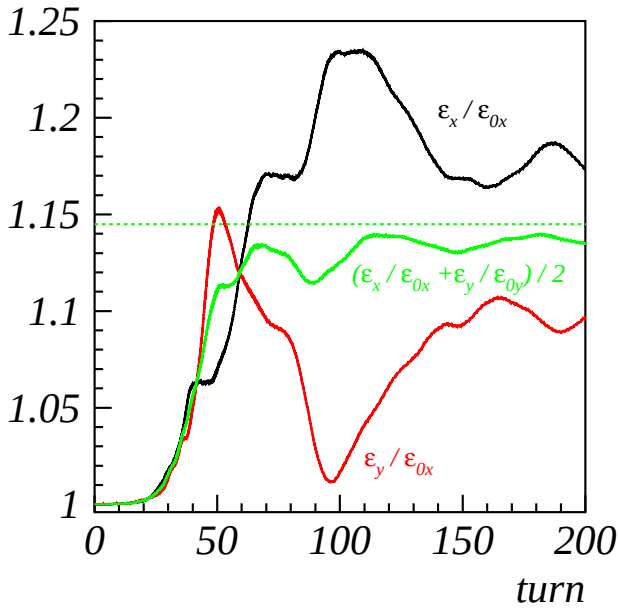


Fig. 14: Evolution of beam emittance for mismatched beam in axi-symmetrical constant-focusing lattice. Green: free-energy limit.

19 Linear coupling: Chernin equations

In the previous section, we discussed the effect of the free energy on a stationary beam. The mismatch type discussed is obtained by changing the beam size from a_i to a_m, and this process creates free energy, which is converted in emittance growth. In doing this, we have addressed only one special type of mismatch, namely the simultaneous mismatch in x–x' and y–y'. However, the mismatch of a beam can be created in much more complex ways by really perturbing the 4D particle distribution. There is, however, one special type of mismatch of relevance when high-intensity beams are considered, and that is the x–y mismatch. This is relevant because the space charge forces are directly controlled by the spatial position of the beam distribution.

We observe that in all the previous discussions, we always implicitly assumed the spatial beam profile to be upright. The simplest type of x–y mismatch is a rotation in the x–y plane of the full beam by an angle θ. The space charge forces respond only to the spatial beam distribution; hence, these forces in the rotated beam will be the same as in the upright distribution, but they will just be rotated as well. The situation is illustrated in Fig. 15.

The left panel shows an upright 2D beam, and the scaled space charge forces as they appear on one particle in the equations of motion are

$$F_x = \frac{2K}{a(a+b)}X, \qquad F_y = \frac{2K}{b(a+b)}Y, \tag{66}$$

where X,Y are the co-ordinates of the test particle in the reference frame where the beam is upright. The right panel of Fig. 15 shows the beam tilted. The axes and forces of the original reference frame, where the beam is upright, are depicted in black. The axis of the new reference frame is depicted in blue. The test particle now has new co-ordinates x,y, and the forces acting on the particle are the projection of the scaled space charge force vector (red) along the x and y axes (blue). It is therefore possible to find the X,Y co-ordinates in the upright reference frame from the particle co-ordinates x,y in the frame of the tilted beam, and there the space charge forces are given by Eq. (66). At that point, we decompose F_x,F_y along the new axes, and find the new components of the scaled space charge force f_x,f_y along the axis x,y. By applying this procedure mathematically, we obtain the following scaled forces:

$$\begin{aligned} f_x &= \left[\frac{2K}{a(a+b)}\cos^2\theta + \frac{2K}{b(a+b)}\sin^2\theta\right]x + \sin\theta\cos\theta\left[\frac{2K}{a(a+b)} - \frac{2K}{b(a+b)}\right]y, \\ f_y &= \left[\frac{2K}{b(a+b)}\cos^2\theta + \frac{2K}{a(a+b)}\sin^2\theta\right]y + \sin\theta\cos\theta\left[\frac{2K}{a(a+b)} - \frac{2K}{b(a+b)}\right]x. \end{aligned} \tag{67}$$

We observe immediately that the tilting of the beam produces a coupling between the horizontal and vertical planes. It is interesting that the coefficient of the coupling terms in f_x,f_y is the same. This appears exactly as if there would be a skew quadrupole acting on the test particle.

The situation is now more complex than that discussed in Section 9. In fact, the former equation of motion of a single particle, Eq. (34), now takes a more complex form, namely

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2} + k_{0x}(s)x - \frac{2K}{a+b}\left(\frac{1}{a}\cos^2\theta + \frac{1}{b}\sin^2\theta\right)x - \sin\theta\cos\theta\frac{2K}{a+b}\left(\frac{1}{a} - \frac{1}{b}\right)y = 0\,. \tag{68}$$

A similar equation can be derived for the motion in the y plane. The coupling term has an important consequence if we try to compute the r.m.s. envelope equation. In fact, in Eq. (32), the second-order moment $\langle xx''\rangle$ will now produce one additional term proportional to the second-order moment $\langle xy\rangle$. Note that the second-order moment $\langle xy\rangle$, is directly proportional to $\sin(2\theta)$. This simply means that the evolution of $\tilde{x}$ depends on coupled second-order moments, and it is therefore not possible to write two separate equations of motion for horizontal and vertical planes.

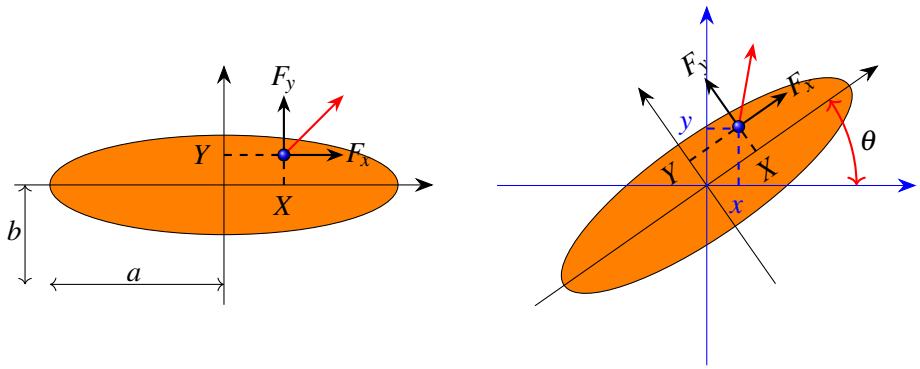


Fig. 15: Decomposition of forces in tilted beam. Left-hand side: upright beam and test particle in it, with space charge forces. Right-hand side: the same beam, now tilted by θ. The space charge force is now decomposed along new axes.

The problem has to be discussed in full generality by describing the evolution of all second-order moments, which forms the 4×4 matrix

$$\Sigma_{i,j} = \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle \,,$$

where $\vec{v} = (x, x', y, y')$. The evolution of Σ is derived by Chernin [8] as

$$\Sigma' = M\Sigma + (M\Sigma)^{\mathrm{T}}, \tag{69}$$

where the symbol $^{\mathrm{T}}$ refers to the operation to transpose a matrix. The matrix M has the form

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\tilde{k}_x & 0 & \tilde{j} & 0 \\ 0 & 0 & 0 & 1 \\ \tilde{j} & 0 & -\tilde{k}_y & 0 \end{pmatrix}, \tag{70}$$

where

$$\tilde{k}_x = k_{0x} - q_{xx} \,, \qquad \tilde{k}_y = k_{0y} - q_{yy} \,, \qquad \tilde{j} = j_0 + q_{xy} \,, \tag{71}$$

and

$$q_{xx} = \frac{K}{2}\frac{S_y}{S_0(S_x+S_y)} \,, \qquad q_{yy} = \frac{K}{2}\frac{S_x}{S_0(S_x+S_y)} \,, \qquad q_{xy} = -\frac{K}{2}\frac{\Sigma_{13}}{S_0(S_x+S_y)} \,, \tag{72}$$

and

$$S_x = \Sigma_{11} + S_0 \,, \quad S_y = \Sigma_{33} + S_0 \,, \quad S_0 = \sqrt{\Sigma_{11}\Sigma_{33} - \Sigma_{13}^2} \,.$$

The quantities k_{0x}, k_{0y} are the usual focusing strength, j_0 is the skew strength produced by skew quadrupoles. The terms q_{xx}, q_{yy} are the space charge defocusing effect in the x and y planes, and their form resembles Eq. (66). When $\Sigma_{13} = 0$, then q_{xx}, and q_{yy} become exactly equal to Eq. (66). The term q_{xy} is a self-coupling created by the space charge. It is now possible that the tilted beam produces a self-linear coupling that affects the beam evolution in the x–y plane. This coupling derives from the last term in Eq. (68).

Figure 16 shows the evolution of the beam emittances in a lattice with a skew quadrupole calculated using Eq. (69). The simulation is performed in proximity to the linear coupling resonance

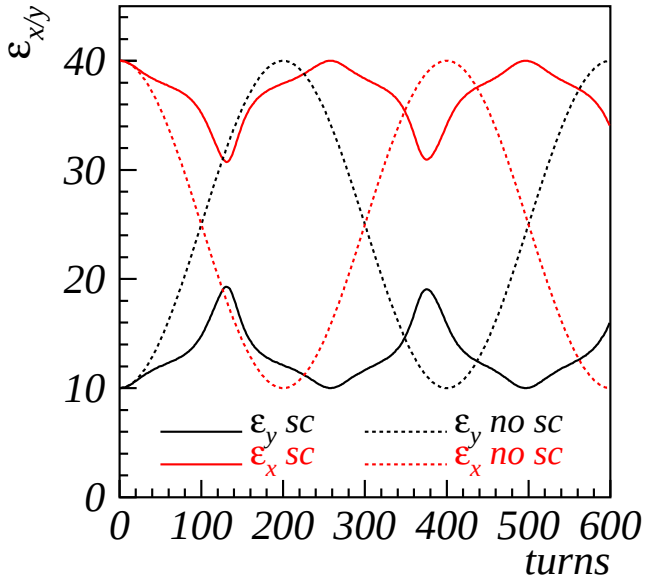


Fig. 16: Beam emittance evolution for low-intensity beam (dashed curves), and high-intensity beam (solid curves)

$Q_{0x} - Q_{0y} = N$ (see also Ref. [9] for further details). The graph shows a simulation of a beam with initial Kapchinsky–Vladimirsky emittances of $\varepsilon_x = 40$ mm-mrad and $\varepsilon_y = 10$ mm-mrad. The dashed curve shows the case without space charge when the tunes sit on the linear coupling resonance. The skew quadrupole exciting the resonance produces an emittance exchange in 200 machine turns. The solid curves show the emittance evolution as computed with Chernin equations when the high intensity yields $\Delta Q_y = -0.2$. The tunes are $Q_{0y} = 3.2$ and $Q_{0x} \simeq 4.16$. It can be seen that the space charge prevents the full emittance exchange. However, the sum of the two emittance is preserved.

20 The Montague resonance

The possibility of a beam affecting itself via space charge is quite a relevant topic. In the previous section we discussed the linear coupling, which in Eq. (68) is shown by an extra coupling term in y, the strength of which depends on the tilting of the beam.

However, more complex effects may be created by space charge. In particular, the space charge may influence particle motion even for an upright beam. In fact, for a 2D Gaussian upright beam, the scaled force on a beam particle is

$$\begin{aligned} F_x(x,y) &= K\left[\frac{1}{a(a+b)}x - \frac{2a+b}{6a^3(a+b)^2}x^3 - \frac{1}{2ab(a+b)^2}xy^2 + \ldots\right], \\ F_y(x,y) &= K\left[\frac{1}{b(a+b)}y - \frac{2b+a}{6b^3(a+b)^2}y^3 - \frac{1}{2ba(a+b)^2}yx^2 + \ldots\right], \end{aligned} \tag{73}$$

where a, b are the r.m.s. sizes of the distribution with

$$n(x,y,s) = \frac{1}{2\pi ab}\exp\left[-\frac{1}{2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\right].$$

For a frozen beam, the forces in Eq. (73) are of incoherent type, and act on a single particle as if they are given by an external element. The main difference between the scaled forces in Eq. (73), and those in Eq. (67) is that for an upright beam $\theta = 0$; hence, the coupling strength in Eq. (67) disappears, while

in Eq. (73) a coupling term with strength

$$-\frac{K}{2ba(a+b)^2}$$

remain unaffected. The consequences of this coupling term go beyond the linear motion as the forces that it generates are non-linear.

According to the theory of the resonances [10, 11], the excitation of a resonance depends on the strength of the harmonics of the driving term, which can be expressed from the potential of the forces. In particular the potential of the coupling term in Eq. (73) reads

$$\frac{K}{4ba(a+b)^2}y^2x^2\,.$$

This term excites the resonances $2Q_x \pm 2Q_y = N$. Of particular interest is the resonance $2Q_x - 2Q_x = 0$, which is quite unusual in single-particle dynamics, as the zero-order harmonics are very weak because non-linear components are typically localized in distinct spots around the machine. The strength of the zero-order harmonics is proportional to the integral

$$K\int_0^L \frac{\beta_{0x}(s)\beta_{0y}(s)}{a(s)b(s)[a(s)+b(s)]^2}\exp\left\{i2\left[\phi_x(s)-2\pi Q_{0x}\frac{s}{L}\right]-i2\left[\phi_y(s)-2\pi Q_{0y}\frac{s}{L}\right]\right\}\mathrm{d}s\,,$$

where it is evident that the driving term is proportional to the perveance. In the formula are used the phase advances defined as $\phi_x(s) = \int_0^s ds/\beta_{0x}(s), \phi_y(s) = \int_0^s ds/\beta_{0y}(s)$. This coupling resonance is found in any ring near the diagonal $Q_x = Q_y$, but it is driven by the fourth-order term in the space charge potential. The first study of this space charge resonance was made by Montague [12]. A simple demonstration of the effect of this resonance is obtained by integrating the dynamics of the following simplified system

$$\begin{aligned} x'' + \left(\frac{Q_{0x}}{R}\right)^2 x &= F_x(x,y)\,,\\ y'' + \left(\frac{Q_{0y}}{R}\right)^2 y &= F_y(x,y)\,, \end{aligned} \tag{74}$$

with F_x, F_y given by Eq. (73). These equations describe the dynamics of particles under the effect of the second- and fourth-order terms of the frozen potential. The left panel of Fig. 17 shows the beam response as a function of the working point of a Gaussian beam with r.m.s. emittances of $\tilde{\varepsilon}_x = 40$ mm-mrad and $\tilde{\varepsilon}_y = 20$ mm-mrad. The red curve shows the average r.m.s. emittances $\tilde{\varepsilon}_y$, and the black curve shows the average r.m.s. emittances $\tilde{\varepsilon}_x$. These averages are computed with all the r.m.s. beam emittance from turn 250 to turn 1000. The strength of the perveance K produces a tune-shift of $\Delta Q_{0x} = -0.05$ in the Gaussian beam. The fourth-order coupling term in the potential creates a dynamics which brings the two emittances closer so that the depressed tunes are closer to the resonance $2Q_x - 2Q_y = 0$.

In this first discussion, which follows the tracks of Montague, the space charge force is frozen (i.e., $a_0(s), b_0(s)$ of Eq. (10) follow the machine optics as $a_0(s) = \sqrt{\beta_{0x}(s)\mathscr{E}_x}$ and $b_0(s) = \sqrt{\beta_{0y}(s)\mathscr{E}_y}$ if the intensity is low, otherwise follow the depressed optics as discussed in Section 6).

The results of the left panel of Fig. 17 are obtained by including only the first two terms of the space charge potential; when all the terms of the frozen potential are included, the picture changes, as shown in the right panel of Fig. 17. Here, we see that the emittances are exchanged more, and the stop-band becomes narrower. Surely, this modification stems from the inclusion of all resonant terms and all non-resonant terms in the dynamics, to produce a more realistic detuning.

We observe that in Fig. 17 the r.m.s. emittances of the test particles exchange by a significant amount, owing to the effect of the resonance; consequently, we expect that the beam sizes a, b will also change accordingly. However, in Fig. 17 this effect is removed as we have kept the sizes a, b in

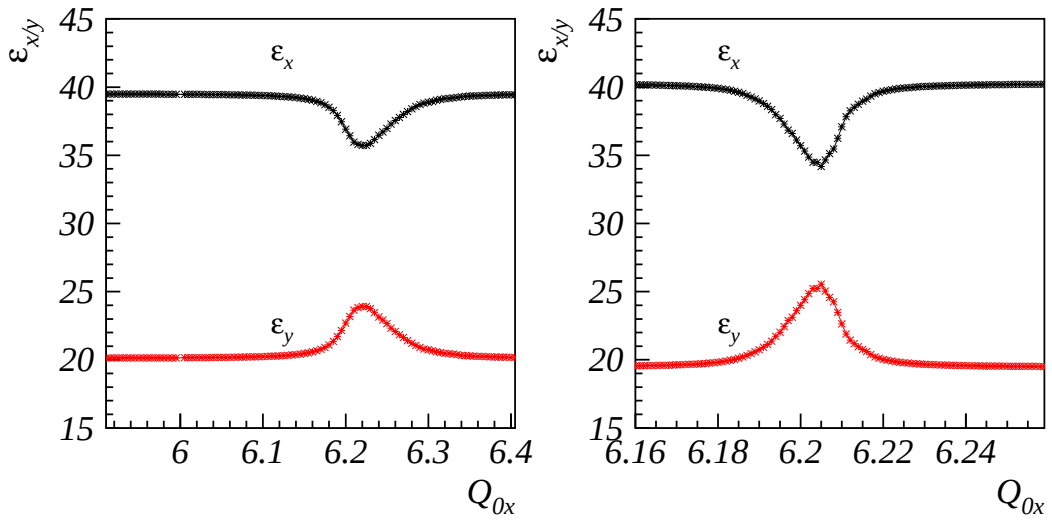


Fig. 17: Average $\tilde{\varepsilon}_x, \tilde{\varepsilon}_y$ in 750 turns after beam. Left-hand side: space charge computed using the second- and third-order potentials. Right-hand side: all terms of the potential are included in the calculation.

the simulations frozen to the optics. If we let the beam size a, b vary with time, the change of beam size will feed back to the space charge; consequently, the dynamics will not produce the results of the right panel of Fig. 17, as the variation of a, b will be included in the beam evolution. This effect of self-consistency will affect the degree of emittance exchange, as shown in Fig. 18. The vertical tune is set to $Q_{0y} = 6.21$, and the incoherent tune-shift is $\Delta Q_y = -0.05$. A detailed discussion of these studies is reported in Ref. [13]. The complexity of the Montague resonance, however, exceeds the analysis so far presented, as the condition of the Montague resonance is also the condition of the instability of the beam collective modes, introduced in Sections 14 and 15. In fact, it can be shown, by a perturbative Vlasov analysis of the collective modes, that some of them become unstable under certain conditions. Figure 19 shows an example of the dependence of the growth rates on Q_{0x} for the fourth-order collective modes in a coasting beam with a transverse Kapchinsky–Vladimirsky distribution (from Ref. [13]). These modes are excited and contribute to the dynamics of the Montague resonance, which becomes a mix of incoherent and coherent effects. The dominance of the coherent effects is found for distributions closer to a Kapchinsky–Vladimirsky distribution, where the dynamics of the emittance exchange is driven by the growth of the collective modes rather than the fourth-order potential à la Montague. For a Gaussian distribution, the collective modes are instead damped by the non-linear field, but the very same field acts according to the original study of Montague and triggers an emittance exchange anyway. Experimental and simulation benchmarking studies are found in Ref. [14].

21 Space charge as incoherent force

If we look again at Eq. (73), the expansion shows the existence of non-coupled terms. Therefore, the third-order components of the force can excite fourth-order resonances if the corresponding driving term is excited by the lattice structure. This is shown in Fig. 20, where a multiparticle particle-in-cell simulation shows that space charge creates the characteristic four islands of a fourth-order resonance. In this example, the horizontal tune is set above the fourth-order resonance, which is excited by a lattice composed of a number of FODO cells equal to the harmonics of the resonance: specifically, here we used a lattice formed with 25 identical FODO cells; the tunes are $Q_{0x} = 6.26$, $Q_{0y} = 6.73$, and the beam is Gaussian with r.m.s. emittances $\tilde{\varepsilon}_x = 40$ mm-mrad, $\tilde{\varepsilon}_y = 20$ mm-mrad. The space charge incoherent

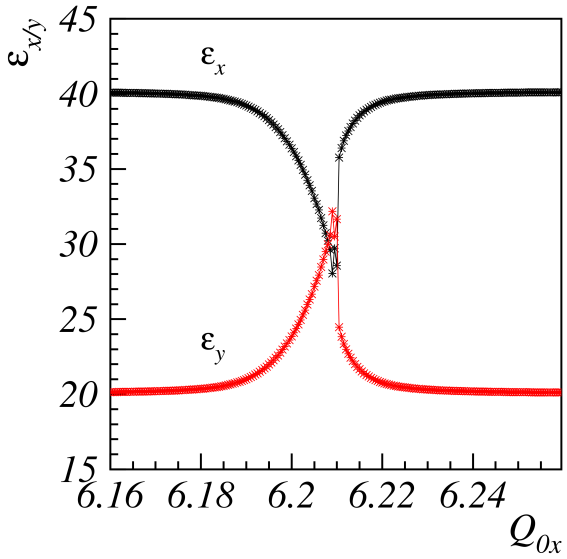


Fig. 18: Emittance exchange obtained via fully self-consistent simulation. It is clear that the symmetry in the exchange is broken by the change of beam sizes.

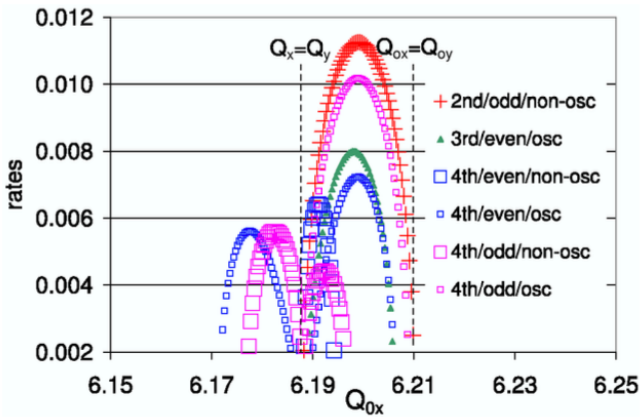


Fig. 19: Example of growth rate of the fourth-order collective modes in a Kapchinsky–Vladimirsky distribution (from Ref. [13]). The stop-band of the modes coincides with the Montague stop-band found in the self-consistent simulations.

tune-shifts are $\Delta Q_{0x} \simeq -0.05$ and $\Delta Q_{0y} \simeq -0.065$. These are relatively modest, but the effect is still remarkable. The original work on the excitation of fourth-order resonances by space charge is found in Ref. [15]. More generally, the non-linear space charge force will exhibit all the odd-order non-linear components. Each of these components can excite a structure resonance if the harmonic number of a non-linear force component satisfies the resonance condition for that non-linear term. A spectral analysis of the x–y components of the space charge force will reveal which harmonics are strongly excited.

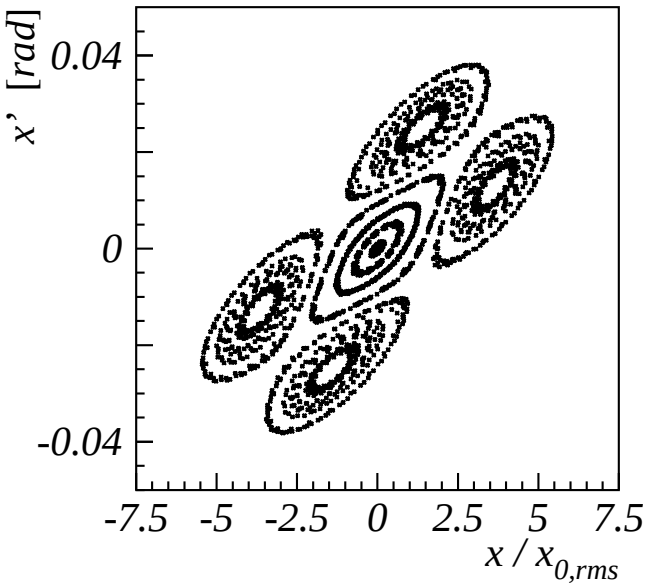


Fig. 20: Poincaré map of a few test particles in a full beam tracked with a particle-in-cell code through a lattice formed by FODO cells. The four islands clearly show that the forth-order resonance is excited. The x axis is plotted in 'normalized' co-ordinates to highlight that the four islands are located, in this example, beyond the tails of the beam distribution.

22 The longitudinal envelope equation

The discussion made for the transverse effects of space charge can be repeated for the longitudinal plane. We recall that the equation of motion in a linearized RF bucket is

$$z'' + k_{0z} z = 0\,, \tag{75}$$

with

$$k_{0z} = \frac{qV\eta h}{2\pi R^2 mc^2 \gamma_0 \beta_0^2}$$

the longitudinal focusing [3]. The derivative has the usual meaning $()' = \mathrm{d}()/\mathrm{d}s$. Here, q is the charge state; V is the maximum voltage applied to the cavity; $\eta = \alpha - 1/\gamma_0^2$ is the slip factor and α the momentum compaction; h is the harmonic number of the cavity, for which the angular frequency is $\omega_{\mathrm{rf}} = 2\pi h/\tau_0$, with τ_0 the revolution time, R the average accelerator radius, and m the particle mass. The stability of the longitudinal oscillations requires $k_{0z} > 0$; hence, V must be selected according to the sign of the slip factor η.

In this notation, we can describe the particle dynamics in terms of the co-ordinates (z, z'), as we did for the transverse plane. The relation of z' with the off momentum of a particle is $z' = -\eta\,\delta p/p$. In analogy with the discussion made for the transverse dynamics, we now consider a matched distribution with the longitudinal optics created by Eq. (75), for a small intensity. Any function

$$\rho_{zz'}(z,z') = Qn(z,z') = Q\frac{1}{\pi \mathscr{E}_z}\tilde{n}\left(\frac{\varepsilon_{0z}(z,z')}{\mathscr{E}_z}\right)$$

represents a matched distribution in the longitudinal plane. Q is the total charge in the bunch, and $\mathscr{E}_z$ is a 'scaling' factor that defines the geometrical extension of the distribution in the phase space. $n(z,z')$ is

the normalized distribution function. The normalization condition requires that the function $\tilde{n}(t)$ satisfies $\int_0^\infty \tilde{n}(t)\mathrm{d}t = 1$. In analogy with the transverse plane discussion,

$$\varepsilon_{0z}(z,z') = \varepsilon_{0z} = \sqrt{k_{0z}}z^2 + \frac{1}{\sqrt{k_{0z}}}z'^2 \tag{76}$$

is the single-particle emittance. If we identify $\beta_{0z} = 1/\sqrt{k_{0z}}$, Eq. (76) is the Courant–Snyder invariant for a constant-focusing channel. Following the approach used in the transverse plane, we may now consider several types of function. Of particular interest is the distribution

$$n(z,z') = \frac{1}{\pi\mathscr{E}_z}\frac{3}{2}\sqrt{1-\frac{\varepsilon_{0z}}{\mathscr{E}_z}}\,.$$

The sizes of this distribution are $z_m = \sqrt{\beta_{0z}\mathscr{E}_z}$, $z'_m = \sqrt{1/\beta_{0z}\mathscr{E}_z}$ and the longitudinal charge-line density, i.e., the projection of $\rho_{zz'}$ to z axis becomes

$$\rho_L(z) = \int_{-z'_m\sqrt{1-z^2/z_m^2}}^{z'_m\sqrt{1-z^2/z_m^2}} Qn(z,z')\mathrm{d}z' = Q\frac{3}{4z_m}\left(1-\frac{z^2}{z_m^2}\right). \tag{77}$$

As expected, the particle distribution will create a longitudinal electric field. Unlike the derivation from the transverse plane, the longitudinal space charge electric field in the laboratory frame is computed as

$$E_z = -\frac{g}{4\pi\varepsilon_0\gamma_0^2}\frac{\partial\rho_L(z)}{\partial z}\,, \tag{78}$$

see Ref. [3] for a derivation. The factor g is a geometric factor; it incorporates the effect of the image charge on the longitudinal electric field. For long bunches, $g \simeq 0.67 + 2\ln(r_{\mathrm{pipe}}/a)$, with a the transverse beam size [3].

If we include the longitudinal space charge electric field in the equation of motion, we obtain

$$z'' + k_{0z}z = -\frac{\eta q}{mc^2\beta_0^2\gamma_0}E_z\,,$$

and for the particular electric field of Eq. (78) we find

$$z'' + k_{0z}z = \frac{Zr_{\mathrm{p}}g\eta}{eA\beta_0^2\gamma_0^3}\frac{\partial\rho_L(z)}{\partial z}\,.$$

where $q = Ze$, $m = Am_{\mathrm{p}}$, with e the electron charge, and m_{p} the proton mass; $r_{\mathrm{p}} = 1/(4\pi\varepsilon_0)e^2/(m_{\mathrm{p}}c^2)$ is the classical radius of the proton. Now we consider the frozen parabolic longitudinal charge-line density described by Eq. (77), and the equation of motion becomes

$$z'' + k_{0z}z = -\frac{3}{2}\frac{NZ^2r_{\mathrm{p}}g\eta}{A\beta_0^2\gamma_0^3}\frac{z}{z_m^3}\,. \tag{79}$$

In analogy with the discussion in the transverse plane, we define a longitudinal perveance K_{L} as

$$K_{\mathrm{L}} = -\frac{3}{2}\frac{NZ^2r_{\mathrm{p}}g\eta}{A\beta_0^2\gamma_0^3}\,,$$

and the equation of motion takes the following form

$$z'' + k_{0z}z - K_{\mathrm{L}}\frac{z}{z_m^3} = 0\,. \tag{80}$$

This equation is the equivalent of Eqs. (20) and (21) for the transverse plane.

We now can define an r.m.s. envelope as $\tilde{z} = \sqrt{\langle z^2 \rangle}$, and through straightforward algebra we find

$$\tilde{z}'' + k_{0z}\tilde{z} - K_{\mathrm{L}}\frac{\tilde{z}}{z_m^3} - \frac{\varepsilon_{zz'}^2}{\tilde{z}^3} = 0\,,$$

where we define the r.m.s. longitudinal emittance as

$$\tilde{\varepsilon}_{zz'}^2 = \langle z^2 \rangle \langle z'^2 \rangle - \langle z z' \rangle^2\,.$$

For the parabolic distribution, we find that the relations between the r.m.s. sizes and the edge of the particle distribution are

$$z_m = \sqrt{5}\sqrt{\langle z^2 \rangle}, \quad z'_m = \sqrt{5}\sqrt{\langle z'^2 \rangle}\,,$$

from which we can compute the emittance of the full distribution as $\varepsilon_{\mathrm{L}}^2 = z_m^2 (z'_m)^2$, and hence $\tilde{\varepsilon}_{zz'}^2 = \langle z^2 \rangle \langle z'^2 \rangle = \varepsilon_{\mathrm{L}}^2/25$. The complete longitudinal envelope equation reads

$$z_m'' + k_{0z} z_m - K_{\mathrm{L}}\frac{1}{z_m^2} - \frac{\varepsilon_{\mathrm{L}}^2}{z_m^3} = 0\,.$$

22.1 Effect of space charge and self-consistency

We observe that the electric field generated by the parabolic distribution is linear. Therefore in Eq. (80), a depressed longitudinal focusing strength is well defined as

$$k_z = k_{0z} - \frac{K_{\mathrm{L}}}{z_m^3}\,.$$

Therefore, we can again define a beam matched with the space charge, and define

$$\rho_{zz'}(z,z') = Q\frac{1}{\pi \mathscr{E}_z}\tilde{n}\left(\frac{\varepsilon_z(z,z')}{\mathscr{E}_z}\right),$$

where now

$$\varepsilon_z(z,z') = \varepsilon_z = \sqrt{k_z}z^2 + \frac{1}{\sqrt{k_z}}z'^2\,.$$

This longitudinal particle distribution creates linear space charge forces. Linear forces are consistent with Courant–Snyder invariants, which means that the type of longitudinal particle distribution will remain unchanged. This is equivalent to what happened in the transverse Kapchinsky–Vladimirsky distribution.

The direct proof that the parabolic distribution satisfies the stationary Vlasov equation, hence that the distribution type does not change, was made by Neuffer [16].

We conclude with the observation that the results reported here are very general and apply even if the particle distribution is not upright. The parabolic distribution will still generate a linear electric field, although the longitudinal 'ellipses' are no longer upright. See an example in Ref. [17].

23 Conclusion

This paper is meant to provide a quick overview of space charge effects. This field is more broad than what is here presented. Topics presented are complementary to other CERN Accelerator School proceedings, for example Ref. [18] for further discussions on beam transport in presence of space charge. The topic of tune-shift from self-field and image charge, not covered here, is treated in Ref. [19]. Other derivations, also including a discussion of the space charge limit are found in Ref. [20]. More recent topics, such as the interplay of space charge and machine resonances, are omitted, as are topics of current research.

Acknowledgements

I thank Frederik Kesting, Yuan Yaoshuo, and Ingo Hofmann for reading and commenting on this paper.

References

[1] E.D. Courant and H.S. Snyder, *Ann. Phys. (N.Y.)* **3** (1958) 1. https://doi.org/10.1016/0003-4916(58)90012-5

[2] O.D. Kellog, *Foundation of Potential Theory* (Dover Publications, New York, 1953).

[3] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley-VCH, Weinheim, 2004).

[4] F.J. Sacherer, *IEEE Trans. Nucl. Sci.* **18** (1971) 1015. https://doi.org/10.1109/TNS.1971.4326265

[5] I. Hofmann, *Phys. Rev. E* **57** (1998) 4713. https://doi.org/10.1103/PhysRevE.57.4713

[6] R. Bartolini *et al.*, *Part. Accel.* **52** (1996) 147.

[7] K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, Singapore, 2006).

[8] D. Chernin, *Part. Accel.* **24** (1988) 29.

[9] G. Franchetti *et al.*, *Phys. Rev. Lett.* **94** (2005) 194801. https://doi.org/10.1103/PhysRevLett.94.194801

[10] G. Guignard, A general treatment of resonances in accelerators, CERN-1978-011 (CERN, Geneva, 1978), http:/dx.doi.org/10.5170/CERN-1978-011.

[11] A. Schoch, Theory of linear and non-linear perturbations of betatron oscillations in alternating gradient synchotrons, CERN-1957-021 (CERN, Geneva, 1957), http://dx.doi.org/10.5170/CERN-1957-021.

[12] B.W. Montague, Fourth-order coupling resonance excited by space-charge forces in a synchrotron, CERN-1968-038 (CERN, Geneva, 1968), http://dx.doi.org/CERN-1968-038.

[13] I. Hofmann and G. Franchetti, *Phys. Rev. Spec. Top. Accel. Beams* **9** (2006) 054202. https://doi.org/10.1103/PhysRevSTAB.9.054202

[14] J. Qiang *et al.*, Numerical simulation study of the montague resonance at the CERN proton synchrotron, 3rd Int. Particle Accelerator Conf., New Orleans, 2012, p. WEPPR011.

[15] S. Machida, *Nucl. Instrum. Methods A* **A309** (1991) 43. https://doi.org/10.1016/0168-9002(91)90091-4

[16] D. Neuffer, *IEEE Trans. Nucl. Sci.* **26** (1979) 3031. https://doi.org/10.1109/TNS.1979.4329929

[17] G. Franchetti *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **3** (2000) 084201. https://doi.org/10.1103/PhysRevSTAB.3.084201

[18] I. Hofmann, in Proceedings of the CAS-CERN Accelerator School: 5th Advanced Accelerator Physics Course, Rhodes, Greece, 20 September-1 October 1993, edited by S. Turner, CERN-1995-006 (CERN, Geneva, 1995), pp. 941–954, http://dx.doi.org/10.5170/CERN-1995-006.941.

[19] A. Hoffmann, in Proceedings of the CAS-CERN Accelerator School: 5th General Accelerator Physics Course, Jyvaskyla, Finland, 7-18 September 1992, CERN-1994-001 (CERN, Geneva, 1994), pp. 329–348, http://dx.doi.org/10.5170/CERN-1994-001.329.

[20] K. Schindl, in Proceedings of the CAS-CERN Accelerator School on Accelerator Physics, Zeuthen, Germany, 15-16 September 2003, CERN-2006-002 (CERN, Geneva, 2006), pp. 305–320, http://dx.doi.org/10.5170/CERN-2006-002.305.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Coherent Beam–Beam Effects

X. Buffat
CERN, Geneva, Switzerland

Abstract

The models used to consistently describe the dynamics of two charge particle beams in circular colliders are discussed, along with relevant examples and observations. The need to treat the dynamics of both beams in a consistent manner is introduced with emphasis on conditions in which the resulting effects may become an intensity limitation for the collider.

Keywords

Beam dynamics; orbit effects; dynamic β effect; coherent beam-beam modes; beam coupling impedance; Landau damping.

1 Introduction

In order to describe the dynamics of a high-energy particle beam it is convenient to make a distinction between external forces, e.g. magnets, that affect the motion of the individual particles and so-called coherent forces. The coherent forces depend on the beam properties as a whole; therefore, not only is the single-particle motion affected, but the coherent forces are affected in return. The treatment of such systems require a consistent description of the single-particle dynamics and the coherent forces. The space-charge force is a well-known example of such a coherent force, the aspects of which are discussed in Ref. [1]. The fact that the beam–beam interactions are localized and result from the interaction of two distinct beams reveals several differences with respect to space-charge effects, which we discuss in this paper.

There exists configurations for which the coherent effects of beam–beam interactions remain negligible, i.e. when the beam–beam forces are not strong enough to generate a significant distortion of the other beam's particle distribution. Such configurations are usually referred to as weak–strong, since at least one of the two beams is weak enough not to affect the dynamics of the strong beam. Here we shall consider so-called strong–strong configurations, when the effect of the two beams on each other is strong enough to have an effect on the beam–beam force itself, and therefore require a consistent treatment of the two beams and their electromagnetic interactions. In Section 2 we shall derive an expression for the coherent beam–beam force, and in Section 3 we shall evaluate its effect on the orbit and optics of the two beams in the strong–strong regime. Section 4 is dedicated to the study of oscillatory solutions around the equilibrium and their stability in the presence of beam coupling impedance.

2 Coherent beam–beam kick

The beam–beam kick on a point-like particle, called the incoherent beam–beam kick, can be obtained by integration of Poisson's equation, see Ref. [2]. Using a Gaussian distribution of particles, with r.m.s. transverse beam size $\sigma = \sigma_x = \sigma_y$, one obtains the kick felt by a test particle at a position (x, y) with respect to the other beam's centroid, see Ref. [3]:

$$\Delta x' = -\frac{2r_0 N}{\gamma_r}\frac{x}{r^2}\left(1 - \mathrm{e}^{-r^2/2\sigma^2}\right) , \tag{1}$$

where we have introduced N, the number of charges in the beam, r_0 the classical radius and $r = \sqrt{x^2 + y^2}$. Since the opposing beam is not point-like, the total beam–beam kick, called coherent

https://doi.org/10.23730/CYRSP-2017-003.391

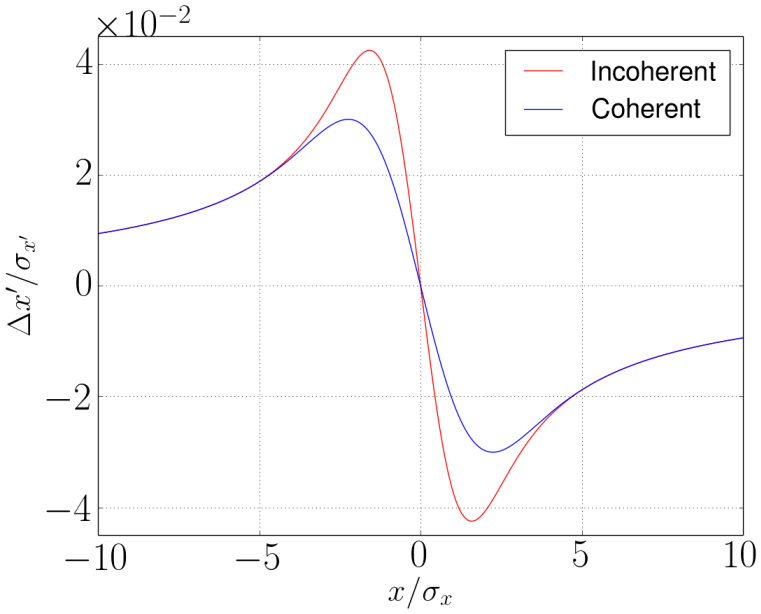


Fig. 1: Comparison between the incoherent and coherent beam–beam kick for round Gaussian beam distributions

kick, is obtained by integration of the single-particle kicks over the beam distribution $\Psi(x, y)$.

$$\Delta x'_{\text{coh}}(x, y) = \int_{-\infty}^{\infty} \mathrm{d}X\mathrm{d}Y \Delta x'(X, Y)\Psi(X - x, Y - y) \, . \tag{2}$$

Assuming a round Gaussian distribution and using Eq. (1), we have, see Ref. [4],

$$\Delta x'_{\text{coh}}(x, y) = -\frac{2r_0 N}{\gamma_r}\frac{x}{r^2}\left(1 - \mathrm{e}^{-r^2/4\sigma^2}\right), \; r = \sqrt{x^2 + y^2} \, . \tag{3}$$

For $x, y \gg \sigma$, we find that the coherent kick is half the single-particle kick, whereas for large separation, i.e for long-range interactions, the difference between coherent and incoherent vanishes (Fig. 1).

3 Self-consistent solutions

In the strong–strong regime, the orbit and optics function of the two beams are dependent on each other through the beam–beam interactions. Therefore a self-consistent calculation of these parameters, including the beam–beam forces, is necessary in order to obtain an accurate description of the two beams. Often, the self-consistent treatment leads to small modifications with respect to a weak–strong treatment, where the beam–beam forces are computed based on the unperturbed optics. In configurations where the beam–beam forces are strong enough, significant dynamic variations of the optics function can occur. We shall illustrate the need for self-consistent solutions through two practical examples: the orbit effect due to multiple long-range beam–beam interactions, and the dynamic β effect.

3.1 Orbit effect

Modern colliders based on synchrotrons usually feature several bunches per beam, potentially leading to several beam–beam interactions per turn. Since head-on collisions other than at the experiment location are unwanted, the beams need to be separated in the rest of the ring. The remaining beam–beam interactions are of long-range type. From Eq. (3), we observe that the coherent kick for beams colliding with

an offset is non-zero, resulting in a modification of the closed orbit, given perturbatively by Ref. [5]:

$$\delta x = \Delta x'_{\text{coh}}(d)\beta \cot(\pi Q) \ , \tag{4}$$

for two beams separated by a distance d in the horizontal plane; for simplicity we assumed $y = 0$. β is the corresponding optics function at the location of the interaction and Q is the unperturbed tune. If the beam–beam interaction is too strong to be treated as a perturbation, i.e. $\Delta_{\text{coh}}(d, 0) \not\approx \Delta_{\text{coh}}(d + \delta x, 0)$, the non-linear equation

$$\delta x = \Delta x'_{\text{coh}}(d + \delta x)\beta \cot(\pi Q) \tag{5}$$

has to be solved. The solution of this equation is, however, not self-consistent, since the effect on the other beam is not taken into account. Considering the two beams self-consistently, one obtains the following system of non-linear equations:

$$\left\{ \begin{array}{rcl} \delta x_1 & = & \Delta x'_{\text{coh}}(d + \delta x_1 + \delta x_2)\beta_1 \cot(\pi Q_1) \\ \delta x_2 & = & \Delta x'_{\text{coh}}(d + \delta x_1 + \delta x_2)\beta_2 \cot(\pi Q_2) \ , \end{array} \right. \tag{6}$$

where the indices refer to the two beams. Such a description is sufficient for configurations where all bunches experience the same beam–beam interactions along the whole ring. This is, however, not the case in machines operated with bunch trains, i.e. where the longitudinal distribution of the bunches along each beam is not uniform. The resulting PACMAN bunches i.e. bunches experiencing different sets of long-range beam–beam interaction, will circulate on different closed orbits (Fig. 2). The algorithm required for the self-consistent evaluation of the orbit of the different bunches is, in essence, identical to the algorithm used to find the closed orbit for a single beam; the dimension of the problem is, however, increased. Starting with a single beam, the closed orbit is defined as the first-order fixed point of a non-linear equation of the following type:

$$\bar{x}_{k+1} = M_0 \cdot \bar{x}_k \ , \tag{7}$$

where M_0 is the one-turn map acting on the phase-space coordinates at a given turn k: $\bar{x_k} = (x_k, x'_k)$. The algorithm can easily be extended to two beams ($i = 1, 2$) with one-turn matrices given by M_1 and M_2 respectively:

$$\bar{x}_{i,k+1} = M_i \cdot \bar{x}_{i,k} \ . \tag{8}$$

In the absence of beam–beam interactions, the one-turn map is identical for all the bunches of each beam:

$$\bar{x}_{i,j,k+1} = M_i \cdot \bar{x}_{i,j,k} \ , \tag{9}$$

with $j = 1, ..., N_b$, the index corresponding to the different bunches of each beam. We can now introduce the effect of beam–beam interactions. The one-turn map of each beam is split into transfer maps between beam–beam interactions, and we write $M_{i,l}$ for the transfer matrix of beam i from beam–beam interaction l to interaction $l+1$, with $l = 1, ..., N_{\text{BB}}$ the index corresponding to the different beam–beam interactions. The effect of the concatenation of the maps must be equal to the original one-turn map:

$$M_i = \prod_{l}^{N_{\text{BB}}} M_{i,l} \ . \tag{10}$$

Then one can write the one-turn matrix including the beam–beam interactions:

$$M'_{i,j} = \prod_{l}^{N_{\text{BB}}} M_{i,l} M^{BB}_{i,j,l} \ , \tag{11}$$

with $M^{BB}_{i,j,l}$ the non-linear beam–beam map of beam i, bunch j at location l defined by

$$M^{BB}_{i,j,l} : x'_{i,j} \mapsto x'_{i,j} + \Delta x'_{\text{coh}}(x_{i,j} - x_{S(i,j,l)}) \ . \tag{12}$$

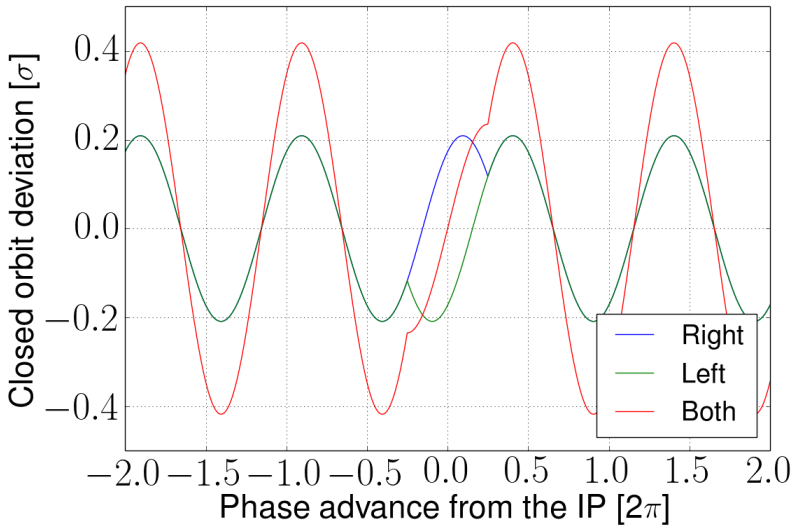


Fig. 2: Illustration of the orbit of three PACMAN bunches in an interaction region with a crossing angle around the interaction point. One experiences a long-range beam–beam interaction only on the left-hand side of the interaction point (green), the other only on the right-hand side (blue), and the last on both sides (red). While the common part of the orbit effect can be corrected using standard orbit correctors, a spread of the closed orbit remains.

The knowledge of the collision scheme was introduced with the function $S(i, j, l)$ giving the indices identifying the bunch of the other beam colliding with bunch j of beam i at the location l. In the absence of beam–beam interaction at the given location, the beam–beam map becomes the identity. Thus, the problem can be reformulated as a closed orbit problem:

$$\bar{x}_{i,j,k+1} = M'_{i,j} \cdot \bar{x}_{i,j,k} \ , \tag{13}$$

where the fixed point gives the closed orbit of the different bunches of each beam, see Ref. [6].

This effect was critical at the Large Electron–Position Collider (LEP), when operated with bunch trains. The orbit variations due to long-range beam–beam interactions resulted in a luminosity loss due to bunches colliding with an offset at the interaction point, see Ref. [7]. It is important to note that since different bunches have different orbits, any correction scheme needs to be fast enough to act differently on individual bunches. In the LHC, the orbit effect at the interaction point is much smaller than the beam size, and the effect on the luminosity is therefore negligible. A harmless displacement of the luminous region could be observed by the experiments (Fig. 3). While small, the orbit variations due to beam–beam interactions play an important role when performing Van Der Meer scans for luminosity calibration, see Ref. [8].

3.2 Dynamic β effect

The formalism developed in previous section to evaluate the bunch by bunch orbits can be extended, by including higher orders in the beam–beam map, in order to allow for the computation of the modification of the machine optics due to multiple head-on and long-range beam–beam interactions. Similarly, one finds that beam–beam interactions cause individual bunches to have different optics functions. These effects may have an important impact on the performance of a collider since bunches with different tunes or different chromaticities may behave very differently in terms of lifetime, for example. In the case of the LHC, mitigation techniques had to be implemented, see Ref. [10]. In some cases, the modification of the optics may rather be controlled in order to reduce the β function at the interaction point and therefore achieve a higher luminosity, see Ref. [11].

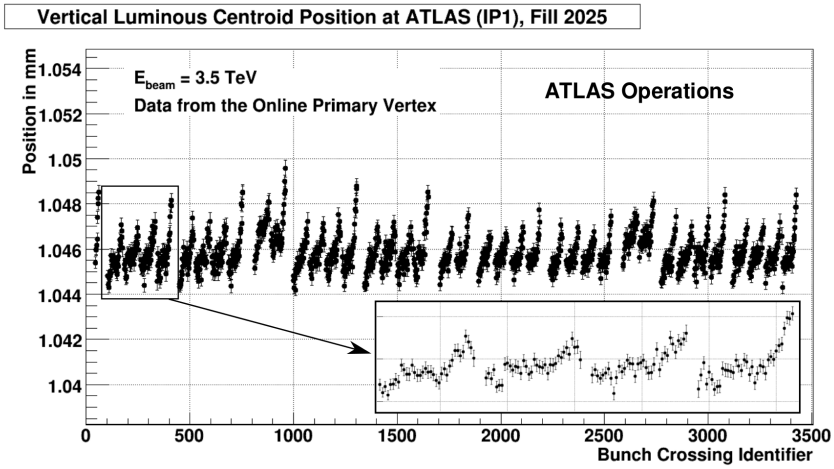


Fig. 3: Bunch-by-bunch displacement of the centroid of the luminous region measured by ATLAS during dedicated experiments in the LHC, see Ref. [9]. The correlation with the position of the bunch in the train, and consequently with the number of long-range beam–beam interactions, is clear.

While complex configurations of beam–beam interactions requires the use of a computer, one can derive equations to describe the dynamics in simpler, yet interesting, configurations. Let us consider a single head-on beam–beam interaction and approximate it to first order. The modification of the optics is then identical to the one due to a quadrupolar error Ref. [12]:

$$\left(\frac{\beta_0^*}{\beta^*}\right)^2 = 1 - \beta_0^* k_{\rm BB} \cot(2\pi Q_0) - \frac{1}{4}\beta_0^{*2} k_{\rm BB}^2 \,, \tag{14}$$

with β_0^* and β^* the unperturbed and perturbed β functions at the interaction point, and $k_{\rm BB}$ the quadrupolar strength of the head-on beam–beam interaction given by Ref. [2]

$$k_{\rm BB} = \frac{r_0 N}{\epsilon_n \beta^*} \,, \tag{15}$$

with ϵ_n the normalized transverse emittance. Assuming that both beams behave identically, we have

$$\left(\frac{\beta_0^*}{\beta^*}\right)^2 = 1 - 2ab\frac{\beta_0^*}{\beta^*} - \left(a\frac{\beta_0^*}{\beta^*}\right)^2 \,, \tag{16}$$

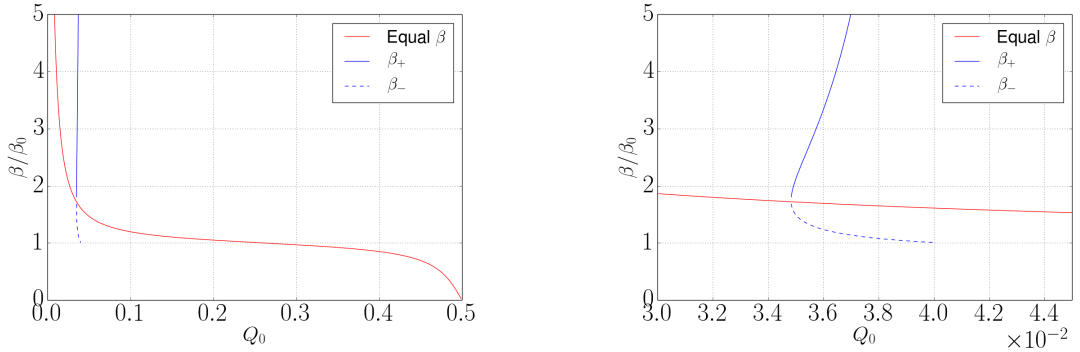
where we have introduced $a \equiv \dfrac{r_0 N}{2\epsilon_n}$ and $b \equiv \cot(2\pi Q_0)$ for convenience. The solution is then

$$\frac{\beta_0^*}{\beta^*} = -\frac{ab \pm \sqrt{1 + a^2(1+b^2)}}{1+a^2} \,. \tag{17}$$

Relaxing the assumption that the two beams have to behave identically, Eq. (14) becomes

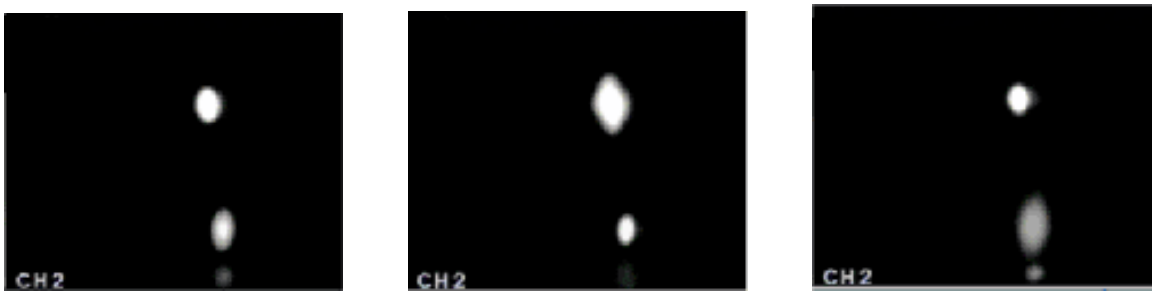
$$\begin{cases} \left(\dfrac{\beta_0^*}{\beta_+^*}\right)^2 &= 1 - 2ab\dfrac{\beta_0^*}{\beta_-^*} - \left(a\dfrac{\beta_0^*}{\beta_-^*}\right)^2 \\ \left(\dfrac{\beta_0^*}{\beta_-^*}\right)^2 &= 1 - 2ab\dfrac{\beta_0^*}{\beta_+^*} - \left(a\dfrac{\beta_0^*}{\beta_+^*}\right)^2 \,, \end{cases} \tag{18}$$

where we have introduced β_+^* and β_-^* the β function of the two beams at the interaction point. The solutions given by Eq. (17) with $\beta_+^* = \beta_-^* = \beta^*$, remains a solution of the system, yet it admits two other solutions with $\beta_+^* \neq \beta_-^*$:



(a) **(b)** zoom

Fig. 4: Dynamic β effect for a beam–beam parameter of 0.02



(a) Equal **(b)** Blown-up positron beam **(c)** Blown-up electron beam

Fig. 5: Flip–flop effect observed at VEPP-2000, see Ref. [13]. With small beam–beam parameters the two beams stay symmetric (left-hand side), but for $\xi \gtrapprox 0.1$ two asymmetric configurations are observed (right-hand side).

$$\frac{\beta_0^*}{\beta_\pm^*} = \frac{ab(1-a^4) \pm \sqrt{(a^2-1)^2(a^2+1)(1+a^4(1+b^2)-a^2(2+3b^2))}}{(a^2-1)^2(a^2+1)} \,. \tag{19}$$

The two solutions (Fig. 4) correspond to one of the two beams being blown-up, whereas the other is squeezed. Since the two beams are identical, both solutions are perfectly equivalent, yet the system may spontaneously chose one of the states, and possibly switch between the two under small perturbations, resulting in the so-called flip–flop effect (Fig. 5), see Refs. [13–15].

4 Coherent beam–beam modes

Assuming that a proper description of the new equilibrium condition including beam–beam interactions was found, we study the oscillation of the two beams around that equilibrium. In the weak–strong regime, the two beams oscillate independently of each other, but in strong–strong configurations the two beams may oscillate coherently. The strong–strong condition is, however, not sufficient to ensure coherent oscillation of the two beams, several mechanisms of decoherence exists. This section introduces the different models used to describe the coherent modes of oscillation and their decoherence.

4.1 Models

4.1.1 The rigid-bunch model

A first step in the description of the coherent beam–beam modes consists in solving the equations of motion of the two beams self-consistently, allowing for their respective transverse positions and momentum

to vary, while all other parameters are kept constant. In other words, the beam distributions are fixed to the equilibrium one, usually assumed to be Gaussian. The two beams are then modelled by their average position and momentum (x_i, x_i'), with respect to their closed orbits. We follow an approach similar to the one developed in Section 3 to derive the one-turn map including self-consistently the beam–beam interactions. Since we are interested in small amplitude modes of oscillations, we keep only the first-order terms, the maps therefore becoming matrices. The normal-mode analysis of the one-turn matrix, including the beam–beam interactions self-consistently, will provide the coherent mode of oscillation. Let us start with the one-turn matrix of each beam:

$$\begin{pmatrix} x_{i,k+1} \\ x'_{i,k+1} \end{pmatrix} = \begin{pmatrix} \cos(2\pi Q) & \beta \sin(2\pi Q) \\ -\dfrac{1}{\beta}\sin(2\pi Q) & \cos(2\pi Q) \end{pmatrix} \cdot \begin{pmatrix} x_{i,k} \\ x'_{i,k} \end{pmatrix} . \tag{20}$$

For the two identical beams, we have (see Eq. (8))

$$\begin{pmatrix} x_{1,k+1} \\ x'_{1,k+1} \\ x_{2,k+1} \\ x'_{2,k+1} \end{pmatrix} = \begin{pmatrix} \cos(2\pi Q) & \beta \sin(2\pi Q) & 0 & 0 \\ -\dfrac{1}{\beta}\sin(2\pi Q) & \cos(2\pi Q) & 0 & 0 \\ 0 & 0 & \cos(2\pi Q) & \beta \sin(2\pi Q) \\ 0 & 0 & -\dfrac{1}{\beta}\sin(2\pi Q) & \cos(2\pi Q) \end{pmatrix} \cdot \begin{pmatrix} x_{1,k} \\ x'_{1,k} \\ x_{2,k} \\ x'_{2,k} \end{pmatrix} \equiv M_0 \cdot \begin{pmatrix} x_{1,k} \\ x'_{1,k} \\ x_{2,k} \\ x'_{2,k} \end{pmatrix} , \tag{21}$$

where we have defined M_0 as the two-beam one-turn matrix. The matrix for a beam–beam interaction may be derived by linearizing Eq. (3) around (x_0, y_0), the closed orbit difference between the two beams at the interaction point:

$$\Delta x'_{\text{coh}}(x, y) \approx \Delta x'_{\text{coh}}(x_0, y_0) + \frac{\partial \Delta x'_{\text{coh}}}{\partial x}(x_0, y_0)\Delta x \, , \tag{22}$$

with

$$\frac{\partial \Delta x'_{\text{coh}}}{\partial x}(x, y) = -\frac{2Nr_0}{\gamma_r}\left[\left(\frac{1}{r^2} - \frac{x^2}{r^4}\right)\left(1 - \mathrm{e}^{-\dfrac{-r^2}{4\sigma^2}}\right) + \frac{x^2}{2r^2\sigma^2}\mathrm{e}^{-\dfrac{-r^2}{4\sigma^2}}\right] . \tag{23}$$

Defining $k_0 \equiv \dfrac{\partial \Delta x'_{\text{coh}}}{\partial x}(x_0, y_0)$, one can then write the coupling matrix between the two beams, due to the beam–beam interaction:

$$M_{\text{BB}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -k_0 & 1 & k_0 & 0 \\ 0 & 0 & 1 & 0 \\ k_0 & 0 & -k_0 & 1 \end{pmatrix} . \tag{24}$$

Thus we can write the one-turn matrix of the two beams including a single beam–beam interaction:

$$\begin{pmatrix} x_{1,k+1} \\ x'_{1,k+1} \\ x_{2,k+1} \\ x'_{2,k+1} \end{pmatrix} = M_{\text{BB}} \cdot M_0 \cdot \begin{pmatrix} x_{1,k} \\ x'_{1,k} \\ x_{2,k} \\ x'_{2,k} \end{pmatrix} . \tag{25}$$

The normal-mode analysis reveals two frequencies each corresponding to two degenerated modes. The first mode corresponds to in-phase oscillation of the two beams (σ-mode), and its coherent tune is the unperturbed machine tune $Q_\sigma = Q$. The second mode of oscillation corresponds to out-of-phase oscillation of the two beams (π-mode); we have

$$\cos(2\pi Q_\pi) = \cos(2\pi Q) - \beta^* k_0 \sin(2\pi Q) \, . \tag{26}$$

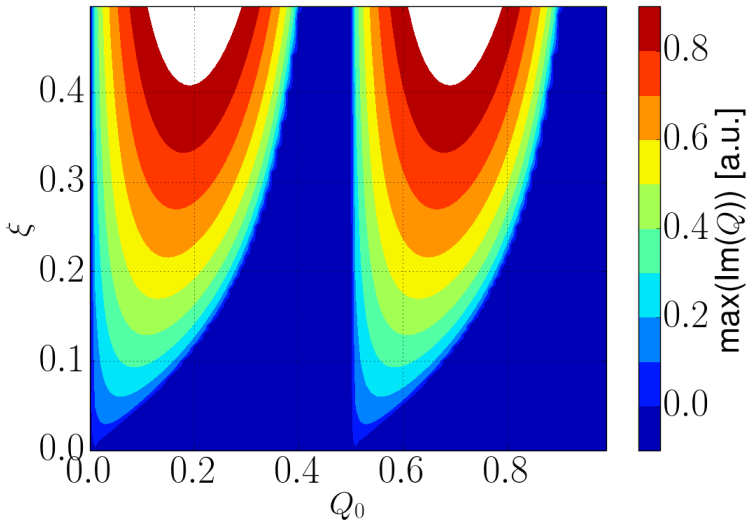


Fig. 6: Largest imaginary part of the eigenvalues of Eq. (25), defining the stable area in terms of unperturbed tune Q_0 and beam–beam parameter ξ.

The stability of the beam–beam modes is given by the imaginary part of the eigenvalues of the matrix given by Eq. (25), which is reported in Fig. 6. From a comparison with Eq. (26), we see that the stability boundary is given by the resonance condition $2Q_\pi = n$. Since we have limited our description of the lattice and of the beam–beam interactions to first order, only the lowest-order resonance are visible. In principle higher-order resonances could also drive the coherent beam–beam modes, see Ref. [16]. It is therefore important to make sure there exist damping mechanisms for these modes; the description of these mechanisms requires more powerful models.

4.1.2 Vlasov perturbation theory

While the rigid-bunch model provides a reasonable description of the coherent modes, the non-linearity of the beam–beam force is neglected. In order to fully assess the effect of the non-linearities on the coherent motion of the two beams, we start from the Liouville equation for the two beams' distributions $\Psi_{1,2}$ with their respective Hamiltonians H_1 and H_2:

$$\begin{cases} \dfrac{\partial \Psi_1}{\partial t} &= \{H_1, \Psi_1\} \\ \\ \dfrac{\partial \Psi_2}{\partial t} &= \{H_2, \Psi_2\} \ , \end{cases} \tag{27}$$

with $\{\cdot, \cdot\}$ denoting the Poisson brackets. The equations are coupled, since the Hamiltonian for a given beam depends on the electromagnetic field generated by the other beam and consequently on its distribution. While mathematically more involved, the treatment of this system of equations is similar to the approach taken in the rigid-bunch approximation. Having derived a Hamiltonian describing both the lattice and the beam–beam interactions, the system can be linearized around the equilibrium distribution $\Psi_{i,0}$ by introducing a first-order perturbation of the distribution $\Psi_i = \Psi_{i,0} + \Psi_{i,1}$. The modes of oscillation and their frequencies are then obtained through the normal-mode analysis of the corresponding linear operator, see Refs. [17, 18]. While the rigid-bunch model only considers fixed particle distributions, the perturbation $\Psi_{i,1}$ is properly matched to the non-linear system. As a result, the frequency of the coherent modes of oscillation is modified. Away from resonances, Eq. (26) obtained within the

rigid-bunch model can be approximated by

$$Q_\pi = Q_0 - \xi \,, \tag{28}$$

while within Vlasov perturbation theory, the frequency shift of the π-mode with respect to the unperturbed tune is increased by the so-called Yokoya factor Y:

$$Q_\pi = Q_0 - Y\xi \,. \tag{29}$$

The values of the Yokoya factor depends on the geometry of the beam–beam interaction, e.g. for round beams ($\sigma_x = \sigma_y$) we have $Y \approx 1.21$ and $Y \approx 1.33$ for flat beams ($\sigma_x \gg \sigma_y$), see Ref. [17]. This model has been shown to provide the most accurate value of the frequency of the coherent beam–beam mode in several colliders Refs. [19–22].This frequency shift has an important impact on the stability of the coherent beam–beam modes. We shall discuss further this aspect in Section 4.2. Let us first consider the limitations of the present model. We approximate the perturbation to first order, i.e neglecting high-order terms of the beam–beam interactions, which has a small impact on the prediction of the modes frequencies. These higher-order terms might, however, impact significantly on the stability of the coherent modes, e.g. through Landau damping. Also, one may note that an accurate description of the interplay between the effect of beam–beam interactions and other mechanisms such as lattice-induced non-linearities or the beam-coupling impedance becomes very cumbersome in this formalism.

4.1.3 Multiparticle tracking simulations

The use of computer tracking simulations allows for an accurate description of the coherent dynamics of the two beams in arbitrarily complex configurations of beam–beam interactions. The interplay with other mechanisms is also conveniently described using a modular description of the different effect as in Ref. [23]. Also, strong assumptions can be relaxed with respect to most theoretical approaches, the cost being in computational resources. Similarly to the method described in Ref. [23], we model the two beams by a set of charged macro-particles. The effect of beam–beam interactions is introduced by evaluating the electromagnetic fields generated by the distribution of macro-particles and apply the corresponding variation of the macro-particles' momentum of the other beam. In cases where the equilibrium particle distribution is close to Gaussian, the fields can be computed based on the numerical evaluation of the moments of the other beam's particle distribution, using the analytical formula for the beam–beam kick (see Eq. (1)). This so-called soft-Gaussian approximation has the advantages of being computationally cheap and introduces a minimal amount of numerical noise in the simulation. In high-energy lepton colliders, the equilibrium distribution is far from Gaussian, due to the effect of synchrotron radiation and the non-linearities of the beam–beam interactions. The use of the soft-Gaussian approximation is therefore excluded in these cases, see Ref. [24]. Numerical solvers for the Poisson equation should be used instead, resulting in fully self-consistent simulations of the beam–beam interactions. Let us illustrate the impact of the different approximations by simulating the simple configuration analysed in previous section, i.e. two bunches colliding in a single interaction point. The Fourier transform of the turn-by-turn position of one of the beam is shown in Fig. 7 for two simulations using either the soft-Gaussian approximation or a fully self-consistent calculation using the HFMM algorithm, see Ref. [25]. One observes that the soft-Gaussian method provides a better estimate of the π-mode frequency with respect to the rigid-bunch model; it is, however, not accurate enough to provide the same frequency as provided by the Vlasov perturbation theory.

Coherent beam-beam modes were observed in most colliders with frequencies in remarkable agreement with the predictions of both Vlasov perturbation theory and tracking simulations. An example of such observation in the LHC is shown in Fig. 8.

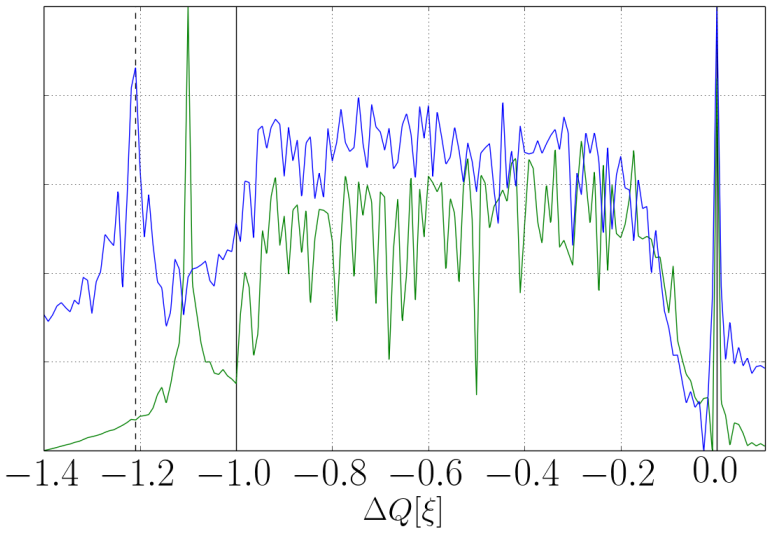


Fig. 7: Comparison of the spectrum obtained with tracking simulation of two bunches colliding in a single interaction point with different models for the beam–beam interactions. The solid black lines show the predictions of the rigid-bunch model, while the dashed black line corresponds to the Vlasov perturbation theory. The green and blue lines correspond to tracking simulations using respectively the soft-Gaussian approximation and a fully self-consistent model.

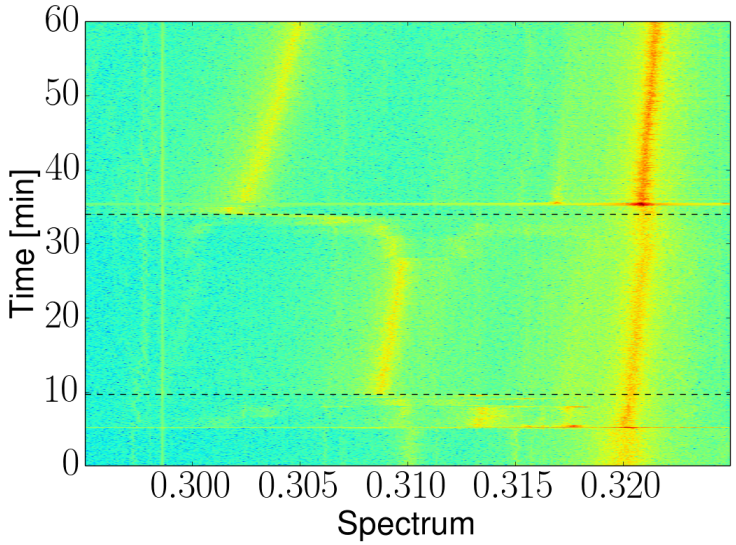
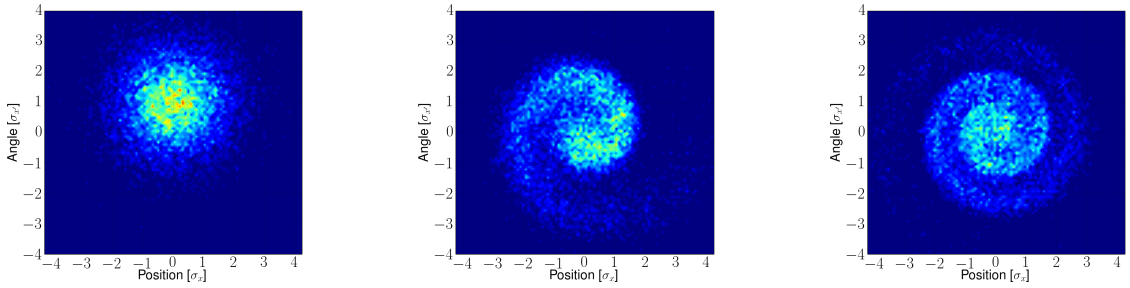


Fig. 8: Measured spectrogram during a dedicated experiment at the LHC with a single bunch per beam. The first dashed line indicates when the beams were brought into collision in one interaction point, revealing a second line shifted down with respect to the unperturbed tune $Q \approx 0.32$ corresponding to the beam–beam π-mode. The second dashed line indicates when the second interaction point is put into collision, pushing the frequency of the beam–beam π-mode further down.



(a) Turn 1 **(b)** Turn 400 **(c)** Turn 800

Fig. 9: Evolution of the transverse phase-space density of a beam under the influence of a beam–beam interaction in the weak strong regime. The initial coherent kick equal to the beam divergence σ'_x decoheres rapidly due to the spread in the oscillation frequency of the individual particles.

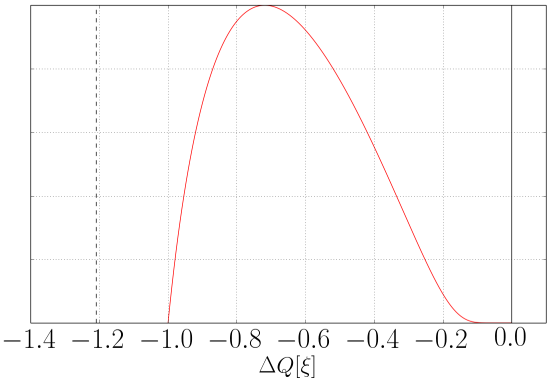


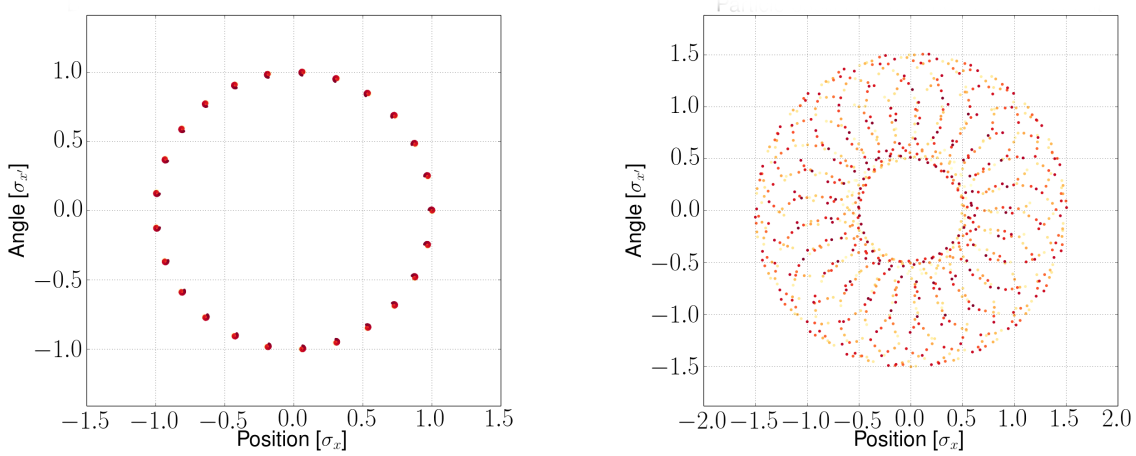
Fig. 10: Spectrum of the single-particle frequencies (red), compared with the frequencies of the coherent modes of oscillation (solid black: σ-mode, dashed black: π-mode) for two symmetric round beams colliding head-on in a single interaction point.

4.2 Decoherence and Landau damping

The presence of external sources of noise acting on the beam, e.g. due to the ripple of dipole fields, leads to emittance growth through filamentation, see Ref. [26]. This effect is illustrated in Fig. 9. In the weak–strong regime, the effect of beam–beam interactions can be treated similarly to other lattice non-linearities, see Ref. [27]. In particular, the motions of the single particles are regular, but the frequency spread results in a damping of the coherent motion at the cost of an increase of the beam emittance. In the presence of beam–beam interactions in the strong–strong regime, the situation is very different. Although the model described in Section 4.1.2 allows for a consistent treatment of decoherence and Landau damping, see Ref. [18], here we rather use a simpler model in order to illustrate the mechanism. The main difference with respect to the filamentation mechanism in the weak–strong regime is the presence of a gap between the coherent-mode frequencies and the single-particle frequencies. This can be understood by considering the equation of motion of a single particle, in the presence of an external force dependent on the single-particle position x_i and on the position of the bunch centroid $\langle x \rangle$, $F(x_i, \langle x \rangle)$:

$$\ddot{x}_i + \omega_0 x_i = \frac{F(x_i - \langle x \rangle)}{m} \,, \tag{30}$$

where $\omega_0 = 2\pi Q$ is the unperturbed tune. In the weak–strong regime, the position of the bunch centroid is fixed, let us chose $\langle x \rangle = 0$ meaning head-on collision of the two beams. Following Ref. [2], we can



(a) Centroid motion

(b) Single-particle motion

Fig. 11: Evolution of a single particle and of the beam centroid position in a fully self-consistent simulation of two symmetric beams colliding head-on. In this configuration, the coherent modes are outside of the incoherent spectrum (Fig. 10), and the absence of decoherence is clearly visible, since the amplitude of oscillation of the centroid position remains constant over time. The single-particle motion is a composition of the incoherent motion and the driven oscillation at the frequency of the beam–beam mode.

write

$$\ddot{x}_i + \omega_i x_i = 0 \,, \tag{31}$$

where ω_i is the perturbed frequency of oscillation of the particle i. Since the perturbed frequency depends on the amplitude of oscillation, the different oscillation amplitudes in the beam results in a frequency distribution, which is represented in Fig. 10 for the particular case of two Gaussian beams colliding head-on. Let us now assume that the two beams oscillate coherently with an amplitude A at a frequency $\Omega = 2\pi Q$. The average position of the beams becomes

$$\langle x_i \rangle = A \cos(\Omega t) \,. \tag{32}$$

The external force becomes time dependent; following the same procedure, Eq. (30) becomes

$$\ddot{x}_i + \omega_0 x_i = \frac{F(x_i - A\cos(\Omega t))}{m} \,. \tag{33}$$

Approximating the external force to first order, we have

$$\ddot{x}_i + \omega x_i = \frac{1}{m}\frac{\partial F(0)}{\partial x} A \cos(\Omega t) \,, \tag{34}$$

which is the equation of a driven oscillator, and the solution for the single-particle motion with $\omega \neq \Omega$ is given by

$$x_i(t) = A_i \cos(\omega_i t + \phi_i) + A_{\mathrm{coh},i} \cos(\Omega t). \tag{35}$$

While A_i and ϕ_i are determined by the initial conditions of each single particles, A_{coh} is given by

$$A_{\mathrm{coh},i} = \frac{1}{m}\frac{\dfrac{\partial F(0)}{\partial x} A}{\omega_i^2 - \Omega^2} \,. \tag{36}$$

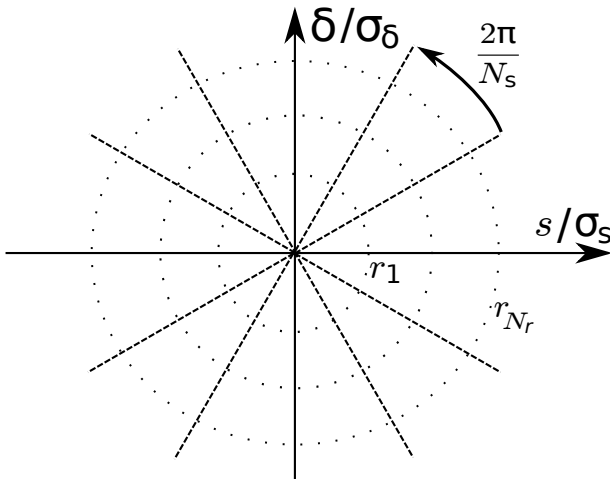


Fig. 12: Discretization of the longitudinal phase space into N_s slices and N_r rings

The bunch centroid position is obtained by averaging Eq. (35) over the particles. Since the initial distribution of phases is assumed uniform, the first term averages out, while the second term remain, we have

$$\langle x_i \rangle(t) = \langle A_{\mathrm{coh},i} \rangle \cos(\Omega t) \; . \tag{37}$$

The exact value of the single-particle amplitudes is not easily obtained within this model, since the excitation amplitude (see Eq. (32)) and the average over all particles (see Eq. (37)) need to be treated consistently. Yet this model shows an important feature of the coherent excitation: its oscillatory component persists, or in other words there is no mechanism of decoherence in these conditions. This effect is illustrated with fully self-consistent macro-particle simulations in Fig. 11. The single-particle motion is well described by the combination of its oscillation at its frequency ω_i and the forced oscillation at frequency of the coherent mode Ω. The oscillation of the bunch centroid around the closed orbit is, however, unperturbed by the presence of the frequency spread as opposed to the result obtained in the weak–strong regime (Fig. 9).

The presence of a resonant condition between the coherent motion and single particles $\omega_i = \Omega$ leads to divergence of the amplitude of oscillation of the single particles (see Eq. (36)), which suggests that there is an energy transfer between the coherent force and the resonant single particles and therefore indicates a decoherence mechanisms. In other words, the necessary condition for damping of the coherent modes is the presence of an overlap between the single-particle frequency spread and the coherent-mode frequency.

While this analysis illustrates the importance of the strong–strong treatment in the understanding of decoherence, more powerful models based on Vlasov equation are needed to properly describe both decoherence and Landau damping in such conditions, see Ref. [18, 28].

4.3 Beam coupling impedance

The impact of beam–beam interactions on the stability of impedance-driven modes is twofold. In the weak–strong regime, the frequency spread resulting from the non-linearities of the beam–beam forces have an impact on the Landau damping of the head-tail modes, which can be quantified using their dispersion relation, see Refs. [2, 28]. If the two beams oscillate coherently, the modes of oscillation are different with respect to the single-beam modes, and therefore they do not satisfy the same dispersion relation. A model describing both the effect of the beam coupling impedance and of the beam–beam interactions is necessary to assess the stability of the coherent beam–beam modes under the influence of the beam coupling impedance. The circulant matrix model, see Refs. [29–31], offers a convenient way

to describe the transverse oscillation of the beams, including the effect of the transverse wake fields and beam–beam interactions. This model is an extension of the rigid-bunch model, allowing for different part of the longitudinal phase space to oscillate independently. The longitudinal phase space is discretized in polar coordinates into so-called slices and rings as illustrated in Fig. 12. The transverse motion of each discrete element can be treated as in the rigid-bunch model, except that all the combinations of beam–beam interactions between the elements need to be considered. Equation 22 becomes

$$\Delta x_i = k_0 \left(\frac{\sum_{j=0}^{N_s N_r} Q_j x_j}{\sum_{j=0}^{N_s N_r} Q_j} - x_i \right) . \tag{38}$$

As an example, let us use two slices and a single ring, and start from Eq. (25):

$$\begin{pmatrix} x_{1,1,k+1} \\ x'_{1,1,k+1} \\ x_{1,2,k+1} \\ x'_{1,2,k+1} \\ x_{2,1,k+1} \\ x'_{2,1,k+1} \\ x_{2,2,k+1} \\ x'_{2,2,k+1} \end{pmatrix} = M_{\text{BB}} \cdot M_0 \cdot \begin{pmatrix} x_{1,1,k} \\ x'_{1,1,k} \\ x_{1,2,k} \\ x'_{1,2,k} \\ x_{2,1,k} \\ x'_{2,1,k} \\ x_{2,2,k} \\ x'_{2,2,k} \end{pmatrix} , \tag{39}$$

where $x_{i,j,k}$ refer to the position of slice j from beam i at turn k. The lattice matrix M_0 can easily be extended, since all slices go through the same lattice, and the beam–beam coupling matrix becomes

$$M_{\text{BB}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -k_0 & 1 & 0 & 0 & k_0/2 & 0 & k_0/2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -k_0 & 1 & k_0/2 & 0 & k_0/2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ k_0/2 & 0 & k_0/2 & 0 & -k_0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ k_0/2 & 0 & k_0/2 & 0 & 0 & 0 & -k_0 & 1 \end{pmatrix} . \tag{40}$$

Such a matrix can be build in a systematic way for an arbitrary number of slices an rings, and, for a complex configuration of beam–beam interactions, following the same approach as in Section 3.1, see Ref. [32]. In order to introduce the effect of the wake, we need to take a closer look at the discretization of the longitudinal phase space. In particular, the longitudinal position of the discrete elements needs to be defined. The definition of the discretization is somewhat arbitrary, but it is convenient to split the phase space such that the charge contained in each element is identical, as was implicitly assumed when deriving Eq. (40). For a Gaussian distribution of particles, the slices are uniformly distributed $\theta_i = 2\pi i / N_s$ and the radius of the rings set such that

$$\mathrm{e}^{-r_{j+1}} - \mathrm{e}^{-r_j} = \frac{1}{N_r} , \tag{41}$$

where $r_j = \sqrt{(s_j/\sigma_s)^2 + (\delta_j/\sigma_\delta)^2}$ is the radius of the jth ring in the normalized longitudinal phase space, i.e. σ_s and σ_δ are the bunch length and relative momentum spread. Therefore one obtains the longitudinal position $s_{i,j}$ and moment deviations $\delta_{i,j}$ of the ith slice and jth ring:

$$\begin{cases} s_{i,j} &= r_j \sigma_s \cos\theta_i \\ \delta_{i,j} &= r_j \sigma_\delta \sin\theta_i \, . \end{cases} \tag{42}$$

Thus we can write the interaction between the discrete elements of the distribution through the beam coupling by using the integrated dipolar and quadrupolar wake functions $W_{\text{dip}}(\Delta s)$ and $W_{\text{quad}}(\Delta s)$, see Ref. [33]:

$$\Delta x_i' = \sum_{j=0}^{N_s N_r} W_{\text{dip}}(s_j - s_i)x_j + W_{\text{quad}}(s_j - s_i)x_i \ . \tag{43}$$

This can be written in a matrix form, in our two-slice model and assuming that the two beams experience identical impedances, we have

$$M_{\text{Z}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ W_{\text{quad}}(s_1 - s_0) & 1 & W_{\text{dip}}(s_1 - s_0) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ W_{\text{dip}}(s_0 - s_1) & 0 & W_{\text{quad}}(s_0 - s_1) & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & W_{\text{quad}}(s_1 - s_0) & 1 & W_{\text{dip}}(s_1 - s_0) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & W_{\text{dip}}(s_0 - s_1) & 0 & W_{\text{quad}}(s_0 - s_1) & 1 \end{pmatrix} , \tag{44}$$

such that the equation of motion becomes

$$\begin{pmatrix} x_{1,1,k+1} \\ x'_{1,1,k+1} \\ x_{1,2,k+1} \\ x'_{1,2,k+1} \\ x_{2,1,k+1} \\ x'_{2,1,k+1} \\ x_{2,2,k+1} \\ x'_{2,2,k+1} \end{pmatrix} = M_{\text{Z}} \cdot M_{\text{BB}} \cdot M_0 \cdot \begin{pmatrix} x_{1,1,k} \\ x'_{1,1,k} \\ x_{1,2,k} \\ x'_{1,2,k} \\ x_{2,1,k} \\ x'_{2,1,k} \\ x_{2,2,k} \\ x'_{2,2,k} \end{pmatrix} . \tag{45}$$

We have written the transverse one-turn matrix for the longitudinal distribution, yet the longitudinal motion has been put aside. Thanks to the choice of decomposition of the longitudinal phase space, the longitudinal motion can be introduced rather simply, as it consists of a rotation of the slices. The longitudinal one-turn matrix is given by the circulant matrix

$$S_r = P_{N_{\text{s}}}^{N_{\text{s}} Q_s} \ , \tag{46}$$

where Q_s is the synchrotron tune and $P_{N_{\text{s}}}$ is a permutation matrix,

$$P_{N_{\text{s}}} = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ 1 & & & 0 & 1 \end{pmatrix} . \tag{47}$$

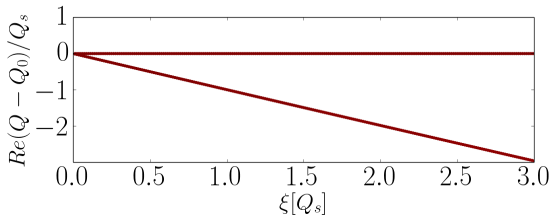
Since the rotation is identical for all rings and for both beams, the matrix in the same basis can be built using the outer product with identity matrix:

$$M_s = \mathbb{I}_2 \otimes \mathbb{I}_{N_{\text{r}}} \otimes S_r \otimes \mathbb{I}_2 \ . \tag{48}$$

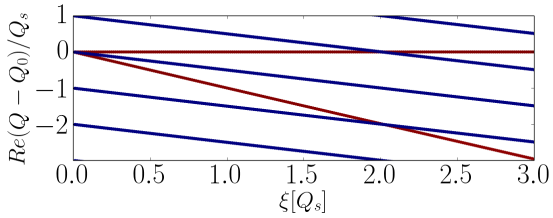
The full one-turn matrix, including the synchro-betatron motion, the beam coupling and beam–beam interactions is then given by

$$M = M_{\text{Z}} \cdot M_{\text{BB}} \cdot M_s \cdot M_0 \ , \tag{49}$$

and its stability can be studied through normal-mode analysis. Let us discuss a simple configuration of two identical bunches colliding at a single interaction point, assuming that the lattice and the impedance



(a) 1 ring, 1 slice



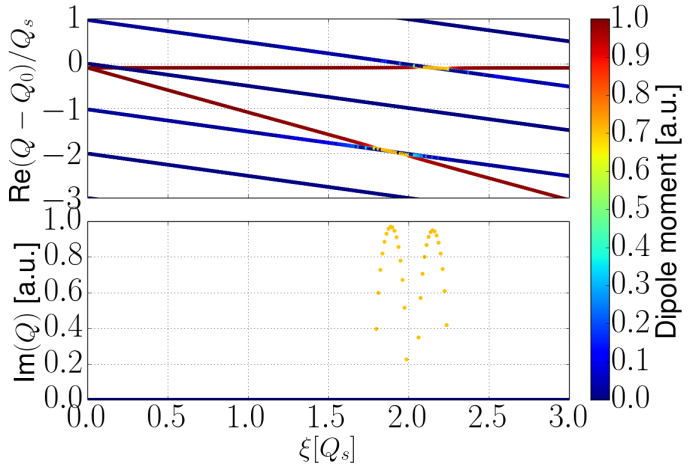
(b) 1 ring, 10 slices

Fig. 13: Eigenfrequencies of the coherent mode of oscillation of two round symmetric beams colliding head-on in on single interaction point for different beam–beam parameter. The points are colour coded according to their dipole moment: the σ and π modes are purely dipolar (red), while the synchrotron sidebands have no dipolar moment (blue). The eigenvalues are all real in the absence of other mechanisms.

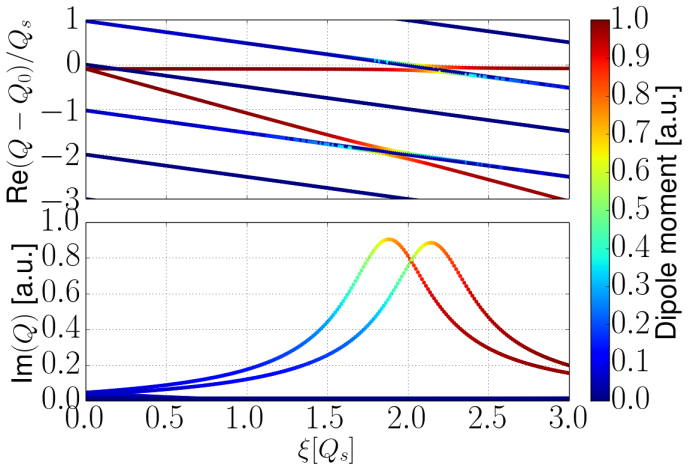
experienced by both beams are identical. Figure 13(a) shows the frequency of the two normal modes obtained with a single slice and a single ring. As expected, we obtain the same solution as the rigid-bunch model, where the σ-mode frequency stays unperturbed, while the π-mode frequency is shifted by $-\xi$. Figure 13(b) shows the same result with a single ring and 10 slices, which allows us to see the frequency of azimuthal modes, appearing as sidebands of the betatron tune. Their frequencies are shifted by $-\xi/2$ due to the beam–beam interaction. This difference between the behaviour of the sidebands can be understood by looking at Eq. (38), where we observe that the sum over the positions of the slices is actually the dipolar moment of the oscillation. Since only the azimuthal mode 0 has a dipolar component, the other modes are only affected by the beam–beam interaction in an incoherent way. In other words, the frequency of the modes are shifted, but the corresponding sidebands of the two beams do not oscillate coherently. In the presence of wake fields the situation is different, since the perturbed azimuthal modes may also have a dipolar component. Figure 14(a) illustrates the impact of a resistive wall impedance on the frequency of the normal mode in the same configuration. The perturbed modes have indeed acquired a dipolar moment, which allows them to interact through the beam–beam interaction. This effect manifests strongly as a mode-coupling instability where the frequency of the π-mode reaches the one of the azimuthal mode -1 and where the frequency of the azimuthal mode 1 reaches the one of the σ-mode. Such an instability mechanism was observed in the LHC, see Ref. [31]. While the operation of the machine is not directly limited by these instabilities, passive or active mitigation techniques need to be put in place, e.g. adjusting the chromaticity or using a transverse feedback.

5 Complex configurations of beam–beam interactions

The complexity of the coherent beam–beam modes increases rapidly when considering realistic machines. In particular, the presence of multiple beam–beam interaction, both head-on or long-range, with different phase advances between the interactions and possibly asymmetries between the beams, create a variety of modes with different frequencies. To illustrate this complexity, the example of the nominal LHC is shown in Fig. 15. The computational power required to accurately describe such a complex system becomes important, even using the simplest models, see Ref. [34]. Let us illustrate the importance of such mechanisms with a simple example.



(a) $Q' = 0$



(b) $Q' = 2$

Fig. 14: Eigenfrequencies of the coherent mode of oscillation of two round symmetric beams colliding head-on in on single interaction point for different beam–beam parameter in the presence of a resistive-wall type of impedance. In the absence of chromaticity, a coupling instability appears when the frequencies of the coherent beam–beam modes reaches those of the synchrotron sidebands. In the presence of chromaticity, the coherent interaction between the two beams has an impact on the stability of head-tail modes, at any beam–beam parameters.

5.1 Tune split

Considering two beams with different unperturbed tunes Q_1 and Q_2 colliding in a single interaction point, we find the frequencies of the coherent mode of oscillation by diagonalizing the matrix of Eq. (25), with the one-turn matrix for the two beams:

$$M_0 = \begin{pmatrix} \cos(2\pi Q_1) & \beta\sin(2\pi Q_1) & 0 & 0 \\ -\frac{1}{\beta}\sin(2\pi Q_1) & \cos(2\pi Q_1) & 0 & 0 \\ 0 & 0 & \cos(2\pi Q_2) & \beta\sin(2\pi Q_2) \\ 0 & 0 & -\frac{1}{\beta}\sin(2\pi Q_2) & \cos(2\pi Q_2) \end{pmatrix} . \tag{50}$$

Figure 16 shows the behaviour of the frequencies when varying the tune split $\Delta Q = |Q_1 - Q_2|$. With $\Delta Q = 0$, we have the configuration studied previously, and the modes are outside the incoherent spectrum. However, for tune splits larger than the beam–beam parameter, the beams are decoupled and the

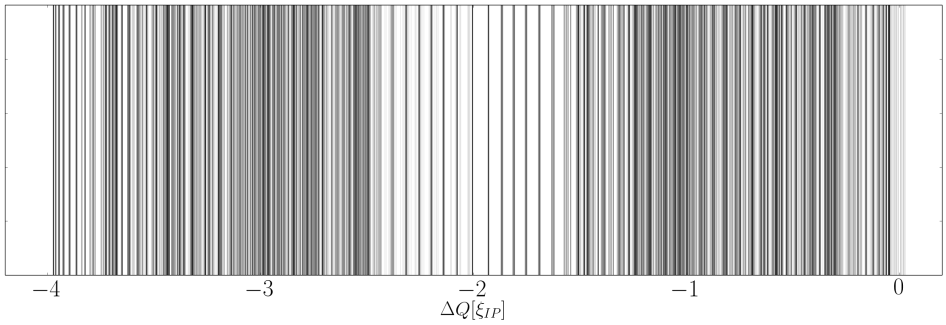


Fig. 15: The LHC coherent beam–beam mode pattern. Since each LHC beam is composed of 2808 bunches each experiencing a different pattern of head-on and long-range interactions, there are 2808^2 modes with different frequencies. The frequency is given in terms of the head-on tune shift per head-on interaction ξ_{IP}. Since there are up to four head-on interactions per bunch, the coherent-mode frequencies are spread between the unperturbed tune Q and $Q - 4\xi_{IP}$.

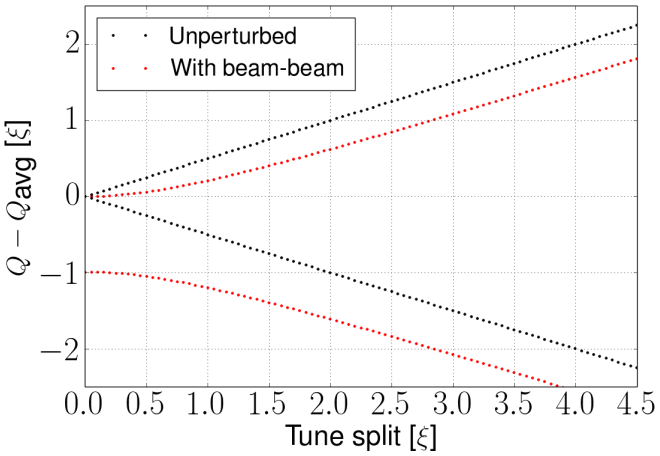


Fig. 16: Frequency of the coherent beam–beam modes for two beams with different tunes

frequency shift tends to half of the beam–beam parameter, i.e. the modes are brought back into the incoherent spectrum. In fact, since the two beams are decoupled because of their different frequencies, their behaviour becomes identical to the one derived for a weak–strong configuration.

References

[1] G. Franchetti, Space charge in circular machines, these proceedings.

[2] T. Pieloni, Beam–beam effects in hadron colliders, these proceedings.

[3] M. Bassetti and G. A. Erskine, Closed expression for the electrical field of a two-dimensional Gaussian charge, CERN-ISR-TH-80-06 (1980).

[4] K. Hirata, *Nucl. Instrum. Methods Phys. Res. A* **269** (1988) 7. http://dx.doi.org/10.1016/0168-9002(88)90856-X

[5] D. Rice, in *Handbook of Accelerator Physics and Engineering*, Eds. A. Chao and M. Tigner (World Scientific Publishing, Singapore, 1999).

[6] H. Grote, Self-consistent orbits with beam–beam effect in the LHC, CERN-LHC-Project-Report-404 (2000).

[7] E. Keil, Truly self-consistent treatment of the side effects with bunch trains, CERN-SL-95-75-AP (1995).

[8] W. Kozaneki, Impact of beam–beam effects on precision luminosity determination at the LHC, Proc. of the ICFA Mini-workshop on Beam–Beam Effects in Hadron Colliders, Geneva, 2013, Eds. W. Herr and G. Papotti (CERN, Geneva, Switzerland, 2014), p. 227.

[9] M. Schaumann and R. Alemany, in Proc. of the ICFA Mini-Workshop On Beam–Beam Effects in Hadron Colliders, 18–22 March 2013, Geneva, Switzerland, edited by W. Herr and G. Papotti, CERN–2014–004 (CERN, Geneva, 2014), pp. 231–235. http://dx.doi.org/10.5170/CERN-2014-004.231.

[10] W. Herr, Features and implications of different LHC crossing schemes, CERN-LHC-Project-Report-628 (2003).

[11] D. Brandt *et al.*, Is LEP beam–beam limited at its highest energy?, Proc. of the 1999 Particle Accelerator Conference, New York, USA, Vol. 5 (1999), p. 3005.

[12] O. Brüning, in Proc. of the CAS–CERN Accelerator School–Intermediate Accelerator Physics, Zeuthen, Germany, 15–26 September 2003, edited by D. Brandt, CERN–2006–002 (CERN, Geneva, 2006), pp. 129–182, http://dx.doi.org/10.5170/CERN-2006-002.129.

[13] D. Shwartz, in Proc. of the ICFA Mini-Workshop On Beam–Beam Effects in Hadron Colliders, 18–22 March 2013, Geneva, Switzerland, edited by W. Herr and G. Papotti, CERN–2014–004 (CERN, Geneva, 2014), pp. 43–49. http://dx.doi.org/10.5170/CERN-2014-004.43.

[14] M. H. R. Donald and J. M. Paterson, *IEEE Trans. Nucl. Sci.* **26** (1979) 3580. http://dx.doi.org/10.1109/TNS.1979.4330106

[15] J. L. Tennyson, *Flipflop Modes in Symmetric and Asymmetric Colliding Beam Storage Rings*, (Lawrence Berkeley, 1989).

[16] C. Alex, *Coherent Beam–Beam Effects*, SSCL-346 (Superconducting Super Collider Laboratory, Dallas, 1991).

[17] K. Yokoya and H. Koiso, *Part. Accel.* **27** (1990) 181.

[18] Y. Alexahin, *Nucl. Instrum. Methods Phys. Res.* A **480** (2002) 253. http://dx.doi.org/10.1016/S0168-9002(01)01219-0

[19] A. Piwinski, *IEEE Trans. Nucl. Sci.* **26** (1979) 4267. http://dx.doi.org/10.1109/TNS.1979.4330764

[20] H. Koiso *et al.*, *Part. Accel.* **27** (1990) 83.

[21] W. Fischer *et al.*, Observation of strong–strong and other beam–beam effects in rhic, Proc. of the 2003 Particle Accelerator Conference, Portland, USA, Vol. 1, (2003), p.135.

[22] X. Buffat *et al.*, in Proc. of the ICFA Mini-Workshop On Beam–Beam Effects in Hadron Colliders, 18–22 March 2013, Geneva, Switzerland, edited by W. Herr and G. Papotti, CERN–2014–004 (CERN, Geneva, 2014), pp. 227–230. http://dx.doi.org/10.5170/CERN-2014-004.227.

[23] K. Li, Numerical methods, these proceedings.

[24] K. Ohmi *et al.*, *Phys. Rev. Lett.* **92** (2004) 214801.

[25] W. Herr, M. Zorzano and F. Jones, *Phys. Rev. ST Accel. Beams* **4** (2001) 054402.

[26] D. Möhl, in Proc. of the CAS–CERN Accelerator School–Intermediate Accelerator Physics, Zeuthen, Germany, 15–26 September 2003, edited by D. Brandt, CERN–2006–002 (CERN, Geneva, 2006), pp. 245–270, http://dx.doi.org/10.5170/CERN-2006-002.245.

[27] V. A. Lebedev, *AIP Conf. Proc.* **326** (1995) 396. http://dx.doi.org/10.1063/1.47298

[28] V. Kornilov, Passive mitigation, these proceedings.

[29] V. Danilov and E. Perevedentsev, *Nucl. Instrum. Methods Phys. Res.* A **391** (1997) 77. http://dx.doi.org/10.1016/S0168-9002(97)00363-X

[30] E. A. Perevedentsev and A. A. Valishev, *Phys. Rev. ST Accel. Beams* **4** (2001) 024403. http://dx.doi.org/10.1103/PhysRevSTAB.4.024403

[31] S. White *et al. Phys. Rev. ST Accel. Beams* **17** (2014) 041002. http://dx.doi.org/10.1103/PhysRevSTAB.17.041002

[32] X. Buffat, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, 2015.

[33] A. Chao, Beam instabilities in circular machines, in *these proceedings*, 2016.

[34] T. Pieloni, PhD thesis, École Polytechnique Fédérale de Lausanne, 2008.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Electron Clouds

G. Rumolo and G. Iadarola
CERN, Geneva, Switzerland

Abstract

The term 'electron cloud' refers to an accumulation of electrons inside the vacuum chamber of a particle accelerator, which is sufficiently strong to produce undesired effects on the accelerator operation, e.g., by causing beam loss, emittance growth, increase in the vacuum pressure, or unacceptable heat load on cold surfaces. Electrons in the beam chamber can primarily be generated by a number of processes, e.g., ionization of the residual gas. Their number, however, can exponentially increase via a beam-induced multipacting mechanism, which relies on acceleration of electrons in the field of the particle beam and efficient secondary emission from their impact on the chamber wall. Several machines running with high-intensity positively charged beams, made of trains of closely spaced bunches, suffer severe effects from electron clouds, and in some cases their performance is even limited by it. Techniques of electron cloud suppression or mitigation exist; the most popular ones are based on the reduction of the secondary electron yield of the chamber inner surfaces. This can be achieved passively through the so-called process of machine scrubbing, or actively by coating the inner pipe walls with appropriate low secondary electron yield materials.

Keywords

Collective effects; two-stream interactions; heat load; mitigation; scrubbing; coherent instabilities.

1 Introduction

In a particle accelerator, free electrons in the beam chambers can be generated by different mechanisms, such as ionization of the residual gas or photoemission from the chamber's wall due to the synchrotron radiation emitted by the beam. When these electrons are accelerated in the electromagnetic field of the beam and reach the chamber walls, they may generate secondary electrons, in a number depending on the impact energy and on the secondary electron yield of the surface. When the accelerator is operated with closely spaced intense bunches of positively charged particles, and assuming that the secondary electron yield of the chamber's inner walls is larger than one, this mechanism can drive an avalanche multiplication process of the electrons—known as multipacting—resulting in the formation of a so-called electron cloud in the chamber. Even accelerators operated with trains of negatively charged particle bunches can be affected by electron cloud formation, if the distance between subsequent bunches is such that the electrons emitted at the wall surface have enough time to move far into the vacuum chamber before they receive a strong repulsive kick back to the wall by the next coming bunch. The presence of a relatively large electron density in the beam pipe, as well as of a strong electron flux to the chamber's wall, can limit the achievable performance of the accelerator through different detrimental effects, such as transverse beam instabilities, transverse emittance growth, particle losses, energy loss, vacuum degradation, and heating of the chamber's surface. Electron cloud effects have been observed in several accelerators around the world, much more commonly in those operated with positively charged particles (e.g., positrons, protons, heavy ions), and are presently a major performance limitation for high-energy colliders, such as the Relativistic Heavy Ion Collider in the USA [1], the KEKB electron positron collider in Japan [2], the DAΦNE electron positron collider in Italy [3], and, more recently, the CERN LHC [4–7].

https://doi.org/10.23730/CYRSP-2017-003.411

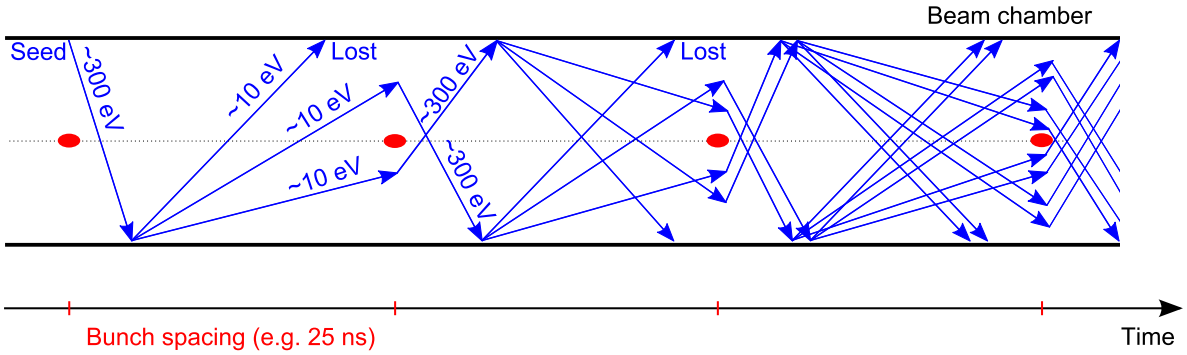


Fig. 1: Formation of an electron cloud in a particle accelerator (a similar sketch can be found in Ref. [8])

A qualitative sketch of the electron cloud build-up at a section of an accelerator operated with bunches of positively charged particles is given in Fig. 1 (see also Ref. [8]). The circulating beam particles can produce electrons through different mechanisms, e.g., ionization of the residual gas in the beam chamber or photoemission from the chamber's wall due to the synchrotron radiation emitted by the beam. These are called 'primary' or 'seed electrons'. Seed electrons are attracted by the passing particle bunch and can be accelerated to energies up to several hundred electronvolts. When an electron with this energy impacts the wall, 'secondary electrons' are likely to be emitted (the probability and efficiency of secondary emission for a certain energy of the incident electron are properties of the surface and also depend on its history). The secondaries have energies up to only few tens of electronvolts and, if they impact the wall with these energies, they are either absorbed or elastically reflected but cannot produce any secondary electrons. Conversely, if they survive until the passage of the following bunch they can in turn be accelerated, projected onto the wall, and produce secondary electrons.

This can trigger an avalanche multiplication effect, which builds up the electron cloud during the passage of an entire bunch train. Although this picture is instructive and represents a possible process leading to an avalanche creation of electrons inside the vacuum chamber, it is important to highlight that this is not a unique resonance condition occurring for a narrow range of the beam parameters and chamber radius. In reality, several different conditions for multipacting can be found, e.g., based on the acceleration of the electrons through multiple bunches or hitting a cyclotron resonance with an external magnetic field.

Section 2 will introduce the different phenomena involved in the formation of an electron cloud; Section 3 will then describe the main features of the electron cloud build-up mechanism. Finally, Sections 4 and 5 will review how the presence of an electron cloud in the beam chambers can affect the performances of a particle accelerator and what countermeasures can be taken.

2 Electron cloud build-up

2.1 Primary electrons

The circulating beam particles can produce electrons through different mechanisms, in particular:

– **ionization of the residual gas in the beam chamber:** this mechanism is always present and is responsible for a production rate of electron–ion pairs proportional to the vacuum pressure and to the ionization cross-section for the colliding particles' species and their energies. These electrons are therefore generated within the volume swept by the beam and must be initialized within the beam cross-section in numerical simulations. Usually, the ionization of the residual gas occurs through a scattering process; however, the rate of electrons produced can become much higher if the electric field of the beam is above the threshold for direct field ionization, i.e., for very intense and bright beams, as might be required for future lepton colliders;

- **photoemission from the chamber's wall:** this mechanism applies to high-energy particle beams (usually for leptons, but this applies also to protons in the LHC), which emit a significant amount of synchrotron radiation with frequencies above the work function of the metal of which the inner wall of the beam chamber is made. The photons of the synchrotron radiation will therefore produce electrons ('photoelectrons') with a certain production efficiency ('photoemission yield'). The photoelectrons are generated at the chamber wall surface, usually with a large fraction concentrated within the small cone of direct incidence of the synchrotron radiation against the chamber wall and the rest distributed around the remaining part of the wall, according to distributions depending on the chamber geometry;
- **beam particle loss at the inner wall of the beam chamber:** this mechanism depends on the loss rate and the electron emission rate for the grazing incidence of high-energy particles. However, in standard operation, losses are usually very low by design in particle accelerators, except in collimators or aperture restrictions, where they are controlled, so the previous two mechanisms are generally dominant as sources of primary electrons.

The electrons produced through these mechanisms are called 'primary' or 'seed' electrons, which are usually produced in sufficiently small amount that they would not affect the accelerator or the beam if their number was not exponentially amplified by the mechanism described in the next subsection. Only the photoelectrons, which can be produced in large amounts, owing to the large number of photons generated by high-energy particles circulating in a ring, have the potential to accumulate to the point of affecting the beam, even in the absence of any amplification mechanism. In any case, primary electrons interact electromagnetically with the passing particle bunch and can be accelerated to energies up to several hundred electronvolts.

2.2 Secondary electron emission

The secondary electron emission process can be described through the secondary electron yield of the surface, which is defined as the ratio between the electron current impinging the wall and the corresponding emitted electron current, and is a function of the energy of the impacting electrons:

$$\delta(E) = \frac{I_{\text{emit}}}{I_{\text{imp}}(E)} \,. \tag{1}$$

A typical secondary electron yield curve is presented in Fig. 2. Following the approach presented in Refs. [9–13], this quantity can in turn be decomposed into two main components:

$$\delta(E) = \delta_{\text{elas}}(E) + \delta_{\text{true}}(E) \,, \tag{2}$$

where $\delta_{\text{elas}}(E)$ and $\delta_{\text{true}}(E)$ represent the electrons that are elastically reflected by the surface and the so-called 'true secondaries', respectively. The dependence of the two components on the energy of the impacting electrons is shown by the green and red curve in Fig. 2. The close-up displayed in the right-hand plot of Fig. 2 shows that elastic reflection of electrons impinging the wall plays a significant role only for very low energies (typically below a few tens of electronvolts) and causes electrons in this range of energies to have a much higher probability of survival on impact. This mechanism is not at all negligible in terms of electron cloud build-up, as it obviously helps to increase the speed of electron accumulation.

We will call δ_{max} the maximum of the secondary electron yield curve, which occurs for a certain energy of the incident electron E_{max}. Both these parameters are strongly dependent on surface material, roughness, and history and play a key role in the electron cloud build-up, as will be discussed in Section 2.3. In the following, δ_{max} will be often referred to simply as the 'secondary electron yield parameter'. A typical energy distribution of the true secondary electrons is shown in Fig. 3. True secondary electrons are emitted with a cosine angular distribution with respect to the direction normal to the surface and their energy spectrum is well fitted by a 'log–normal' distribution.

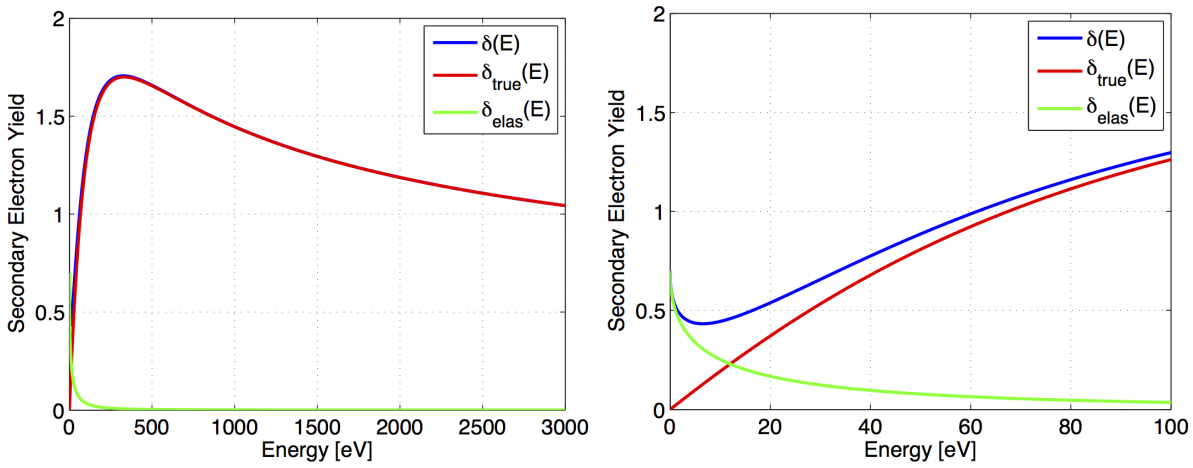


Fig. 2: Left-hand side: secondary electron yield curve for $\delta_{\max} = 1.7$; elastic component $\delta_{\text{elas}}(E)$, 'true secondary' component $\delta_{\text{true}}(E)$, and total $\delta(E)$. Right-hand side: close-up of low-energy region.

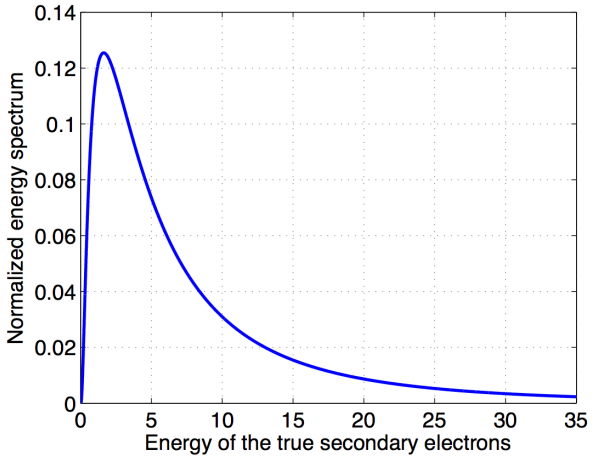


Fig. 3: Energy distribution of true secondary electrons

2.3 Electron cloud build-up mechanism

Let us consider a train of uniformly spaced bunches passing at a certain section of an accelerator, which does not contain any electron before the passage of the first bunch. Let n_0 be the number of primary electrons generated by a single-bunch passage and n_i the number of electrons in the chamber at the instant t_i right before the passage of the ith bunch. We can define $\delta_{\text{eff},\,i}$ such that

$$n_{i+1} = \delta_{\text{eff},\,i}\, n_i + n_0\,, \tag{3}$$

where $\delta_{\text{eff},\,i}\, n_i$ is the number of electrons generated by the interaction of the electron cloud with the chamber's wall (such a quantity can also be negative, when the wall acts as a net electron absorber).

The quantity $\delta_{\text{eff},\,i}$ can be directly related to the secondary electron yield of the chamber's surface $\delta(E)$ and to the energy spectrum of the impacting electrons, since we can write

$$n_{i+1} = n_i + \int_0^{\infty}\int_{t_i}^{t_{i+1}} \Phi(E,t)\,(\delta(E)-1)\,\mathrm{d}t\,\mathrm{d}E + n_0\,, \tag{4}$$

where

$$\Phi(E,t) = \oint n(\vec{r},E,t)\vec{v}\cdot\mathrm{d}S \tag{5}$$

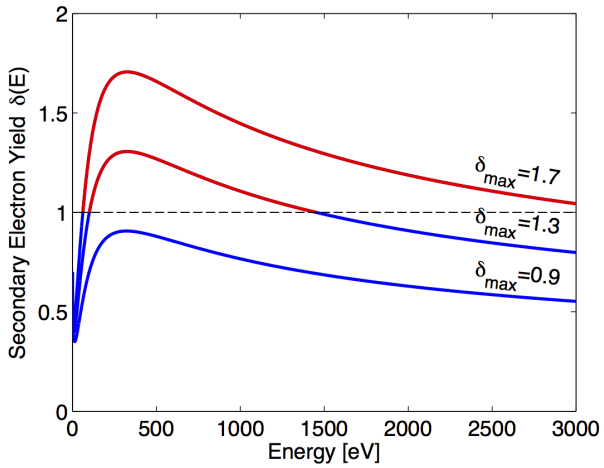


Fig. 4: Secondary electron yield curve for different values of $\delta_{\max}$. Values for which the material behaves as electron absorber or emitter are plotted in blue and red, respectively.

is the instantaneous energy spectrum of the electrons impinging the wall. If we define the normalized energy spectrum for the the ith bunch passage as

$$\phi_i(E) = \frac{1}{n_i} \int_{t_i}^{t_{i+1}} \Phi(E,t) \mathrm{d}t \,, \tag{6}$$

we can rewrite Eq. (4) as

$$n_{i+1} = n_i \left(1 + \int_0^\infty \phi_i(E) \left(\delta(E) - 1 \right) \mathrm{d}E \right) + n_0 \tag{7}$$

and, comparing this with Eq. (3), we obtain

$$\delta_{\mathrm{eff},\,i} = 1 + \int_0^\infty \phi_i(E) \left(\delta(E) - 1 \right) \mathrm{d}E \,. \tag{8}$$

The meaning of this equation is quite intuitive: the secondary electron yield curve can be divided into two regions, one in which $\delta(E) < 1$ and the wall acts as an electron absorber, and the other in which $\delta(E) > 1$ and the wall acts as an electron emitter. The two regions are shown in blue and red, respectively, in Fig. 4, for different values of $\delta_{\max}$. Looking at Eq. (8), we observe that, if the electron flux $\phi_i(E)$ mainly overlaps the $\delta(E) < 1$ region, the integral is negative, $\delta_{\mathrm{eff},\,i} < 1$, and the chamber's wall behaves like a net absorber. Conversely, if $\phi_i(E)$ mainly overlaps the $\delta(E) > 1$ region, the integral is positive, $\delta_{\mathrm{eff},\,i} > 1$, and the chamber's wall behaves like a net emitter.

If the electrons do not influence each other's trajectories, which means that the Coulomb forces between them are negligible, then we can assume that $\phi_i(E)$ does not depend on the bunch index:

$$\phi_i(E) = \phi(E) \,, \tag{9}$$

hence the same holds for $\delta_{\mathrm{eff},\,i}$:

$$\delta_{\mathrm{eff},\,i} = \delta_{\mathrm{eff}} \,. \tag{10}$$

In these conditions, by recursively applying Eq. (3) we find:

$$n_i = n_0 \sum_{k=1}^{i} \delta_{\mathrm{eff}}^k = n_0 \frac{1 - \delta_{\mathrm{eff}}^i}{1 - \delta_{\mathrm{eff}}} \,. \tag{11}$$

From this expression, we can recognize two different regimes. When $\delta_{\text{eff}} < 1$, we observe that, for sufficiently large i, n_i tends to the constant value:

$$n_i \simeq \frac{n_0}{1-\delta_{\text{eff}}}\,, \tag{12}$$

which is essentially an equilibrium condition between primary electron production and electron absorption at the chamber's wall. We will therefore call this condition the *seed accumulation regime*.

When $\delta_{\text{eff}} > 1$, we observe an exponential growth of the number of electrons in the chamber, i.e., for sufficiently large i:

$$n_i \simeq n_0 \frac{\delta_{\text{eff}}^i}{\delta_{\text{eff}}-1}\,, \tag{13}$$

which is indeed an avalanche multiplication of the electrons driven by the secondary emission. We will therefore call this condition the *multipacting regime*.

In the seed accumulation regime, a significant amount of electrons can be accumulated only if the primary electron production mechanisms are very strong, as can be the case for synchrotron radiation in a high-energy lepton machine, while, typically for hadron accelerators, sizeable electron cloud effects are only observed in the multipacting regime.

Equation (13) seems to suggest that the number of electrons can increase indefinitely. In fact, other mechanisms intervene to limit the number of electrons. To explore the validity of this simple model, we used the PyECLOUD code [14] to simulate the electron cloud build-up in the very simple case of a cylindrical chamber (radius 22 mm, i.e., the horizontal size of the LHC arc beams screen) without any externally applied magnetic field, with nominal LHC bunch parameters, and a uniform train of 25 ns spaced bunches. For the analysis of these results, it is also useful to introduce a few other quantities, namely the total electron flux on the chamber's wall,

$$F_i = \int_0^\infty \int_{t_i}^{t_{i+1}} \Phi(E,t)\,\mathrm{d}t\,\mathrm{d}E\,, \tag{14}$$

and the fractions of the impacting electrons that lie in the region of the secondary electron yield curve, where the wall acts as an electron absorber or electron emitter, respectively:

$$F_{\text{absor},\,i} = \int_{\{E:\,\delta(E)<1\}} \int_{t_i}^{t_{i+1}} \Phi(E,t)\,\mathrm{d}t\,\mathrm{d}E\,, \tag{15a}$$

$$F_{\text{emit},\,i} = \int_{\{E:\,\delta(E)>1\}} \int_{t_i}^{t_{i+1}} \Phi(E,t)\,\mathrm{d}t\,\mathrm{d}E\,. \tag{15b}$$

The simulation results for the case $\delta_{\text{max}} = 1.1$ are shown in Fig. 5. In particular, the blue curve in the top plot shows the evolution of n_i. In the middle plot we compare two ways of estimating the $\delta_{\text{eff},\,i}$ coefficients from the simulation results, i.e., using both the recursive formula, Eq. (3), and simulated bunch-by-bunch electron spectra to evaluate the integral in Eq. (8). The two estimates are very consistent, showing that the angular dependence of the secondary electron yield, which is included in the simulation but not in the estimate of Eq. (8), is, in this case, negligible. The bottom plot shows, finally, the evolution of the coefficients $F_{\text{absor},\,i}$ and $F_{\text{emit},\,i}$, as defined in Eq. (15). The energy spectrum as defined in Eq. (6) is shown in the bottom plot of Fig. 6, while the top plot shows the secondary electron yield curve using the same x-axis range. Figure 6 shows that the energy spectrum $\phi_i(E)$ is the same all along the simulation (i.e., the condition of Eq. (9) is fulfilled) and lies mainly in the energy region for which the wall behaves as an electron absorber.

For the case of $\delta_{\text{max}} = 1.75$, the results are presented in Figs. 7 and 8. In these plots we can recognize two different stages, one going from the first passage up to around the 45th, and the second

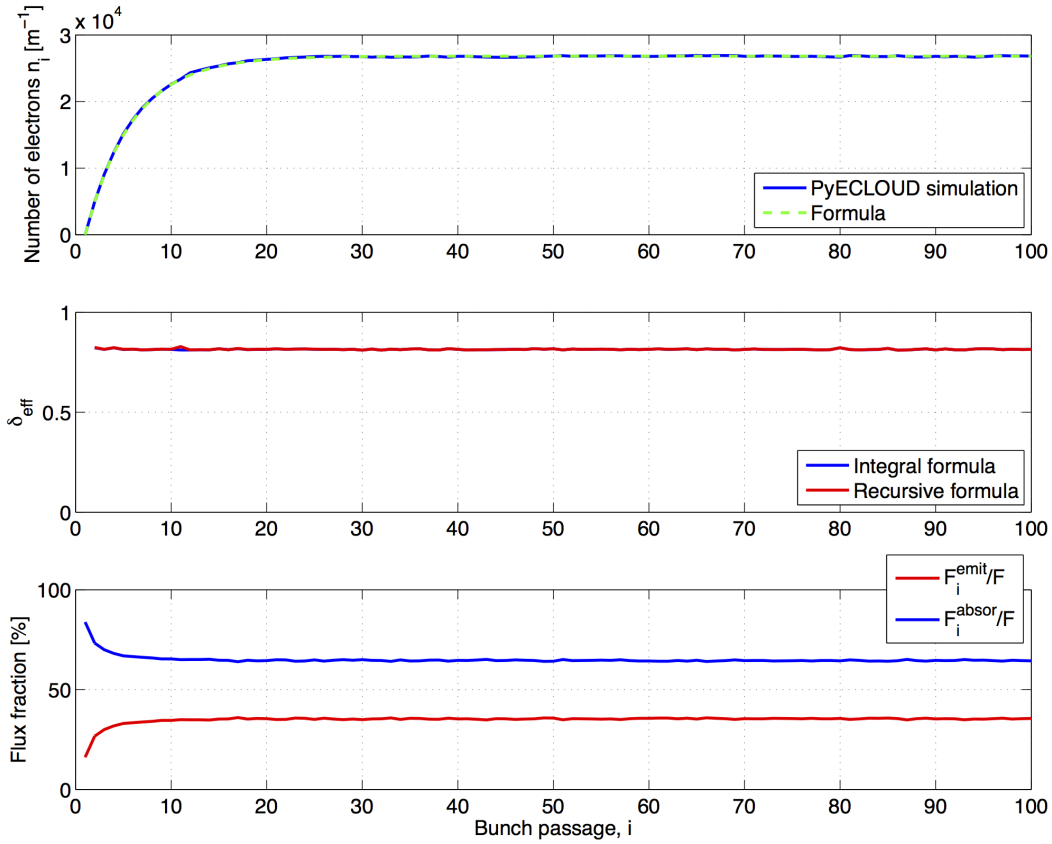


Fig. 5: Simulated electron cloud build-up for cylindrical tube of radius 22 mm, with nominal LHC bunch parameters, a uniform train of 25 ns spaced bunches, and $\delta_{\max} = 1.1$. Top: number of electrons before each bunch passage (blue, directly from simulation; dashed green, estimated from Eq. (11)). Centre: δ_{eff} (both from the integral formula, Eq. (8), and the recursive formula, Eq. (3)). Bottom: fractions of the electron energy spectrum falling in the absorber or emitter regions of the secondary electron yield curve.

from that point onward. In the first stage, the condition of Eq. (9) is verified and $\delta_{\text{eff},\,i}$ is greater than one, which means that we are in the multipacting regime. Indeed, the energy spectrum $\phi_i(E)$ lies mainly in the energy region where the wall behaves like a net electron emitter, as confirmed by Figs. 7 (bottom) and 8. In this case Eq. (11) predicts an exponential increase in the number of electrons, which is exactly what is observed Fig. 7 (top).

Later on, we observe that the evolution of n_i deviates from the expected exponential increase and finally 'saturates' to a constant value, which is larger than the equilibrium value reached in the seed accumulation regime by several orders of magnitude (compare Figs. 5 and 7). By looking at Fig. 8, we observe that during this transition the condition of Eq. (9) is no longer fulfilled, since one can notice a strong increase in the number of electrons hitting the wall with extremely low energy (< 10 eV). We also observe that the electron flux becomes dominated by the fraction lying in the net absorber region (see Fig. 7, bottom) and that the effect of this change in the electron spectrum is that the parameter $\delta_{\text{eff},\,i}$ reduces to one (see Fig. 7, middle).

The reason for this change can be understood by considering the fact that most true secondary electrons are emitted with energies of the order of a few electronvolts (see Fig. 3), and therefore, if they impact on the wall before being accelerated by a bunch passage, they have a high chance of being absorbed (the wall acts as net absorber for these energies, see Fig. 8, top). Figure 9 shows how the electron density and the electrostatic potential evolve during the build-up (all plots correspond to snapshots taken immediately before the passage of the corresponding bunch).

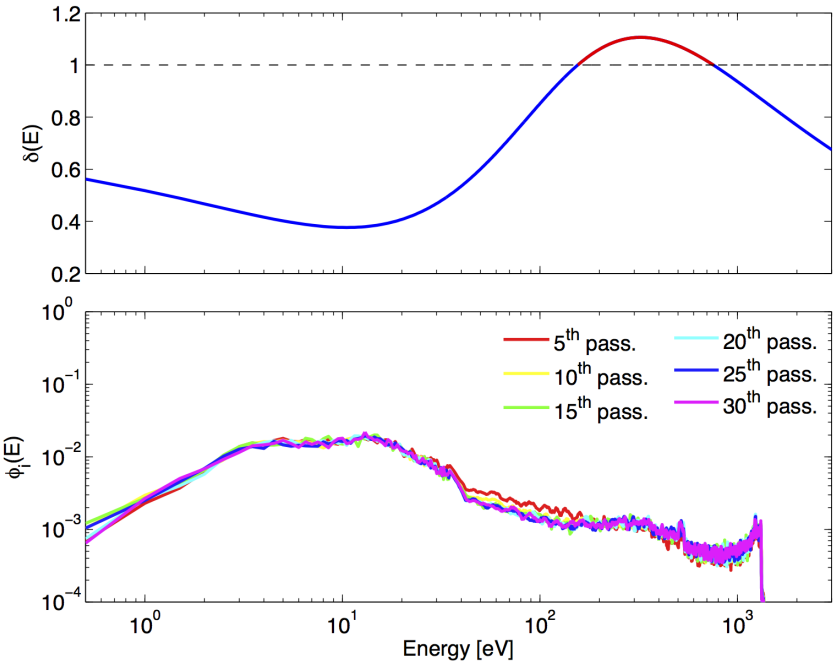


Fig. 6: Top: secondary electron yield curve. Bottom: energy spectrum $\phi_i(E)$ for different bunch passages

During the first stages, the electron density is quite modest, and, as a result, the electrostatic potential in the chamber (with respect to the wall) is less then 1 eV. In these conditions, most of the true secondaries are practically free to move in the chamber. Owing to their initial velocity, they drift towards the centre and have a high chance of avoiding impact on the wall before the next bunch passage.

As the electron density in the chamber increases, so also does the electrostatic potential, which means that the forces due to 'space charge' effects within the electron cloud itself become increasingly stronger. Around the 45th bunch passage, the true secondaries emitted by the wall see a potential barrier comparable to their kinetic energy and therefore tend to be confined in a region close to the chamber's wall. As a consequence, the electron density assumes an annular shape (see Fig. 9) and the probability that low-energy electrons reach the wall before the next passage strongly increases.

This causes the change in the energy spectrum observed in Fig. 8 towards an equilibrium condition such that:

$$\int_0^\infty \phi_i(E)\,(\delta(E)-1)\,\mathrm{d}E = 0\,. \tag{16}$$

Here, electron emission and absorption at the wall perfectly balance one another and therefore $\delta_{\mathrm{eff},\,i} = 1$ (see Eq. (8)).

Figure 10 shows how the maximum number of electrons in the chamber and δ_{eff} in the first stage of the build-up simulation (before space charge effects become significant) depend on the secondary electron yield parameter δ_{max}. The value of δ_{max} for which $\delta_{\mathrm{eff}} = 1$ is called the *multipacting threshold* and separates the seed accumulation and the multipacting regimes. The multipacting threshold can be easily recognized also by the number of electrons in the the beam chamber (see Fig. 10, top), since around this point an increase of several orders of magnitude is observed with respect to the seed accumulation regime. This kind of dependence is also observed for many other quantities related to the electron cloud in the chamber, e.g., the electron flux onto the wall, the electron density at the beam position, and the energy deposition onto the wall. Typically, if δ_{max} is below the multipacting threshold and therefore no avalanche multiplication occurs, the electron cloud tends to be harmless for the machine performance, unless very strong seeding mechanisms are present, as previously discussed.

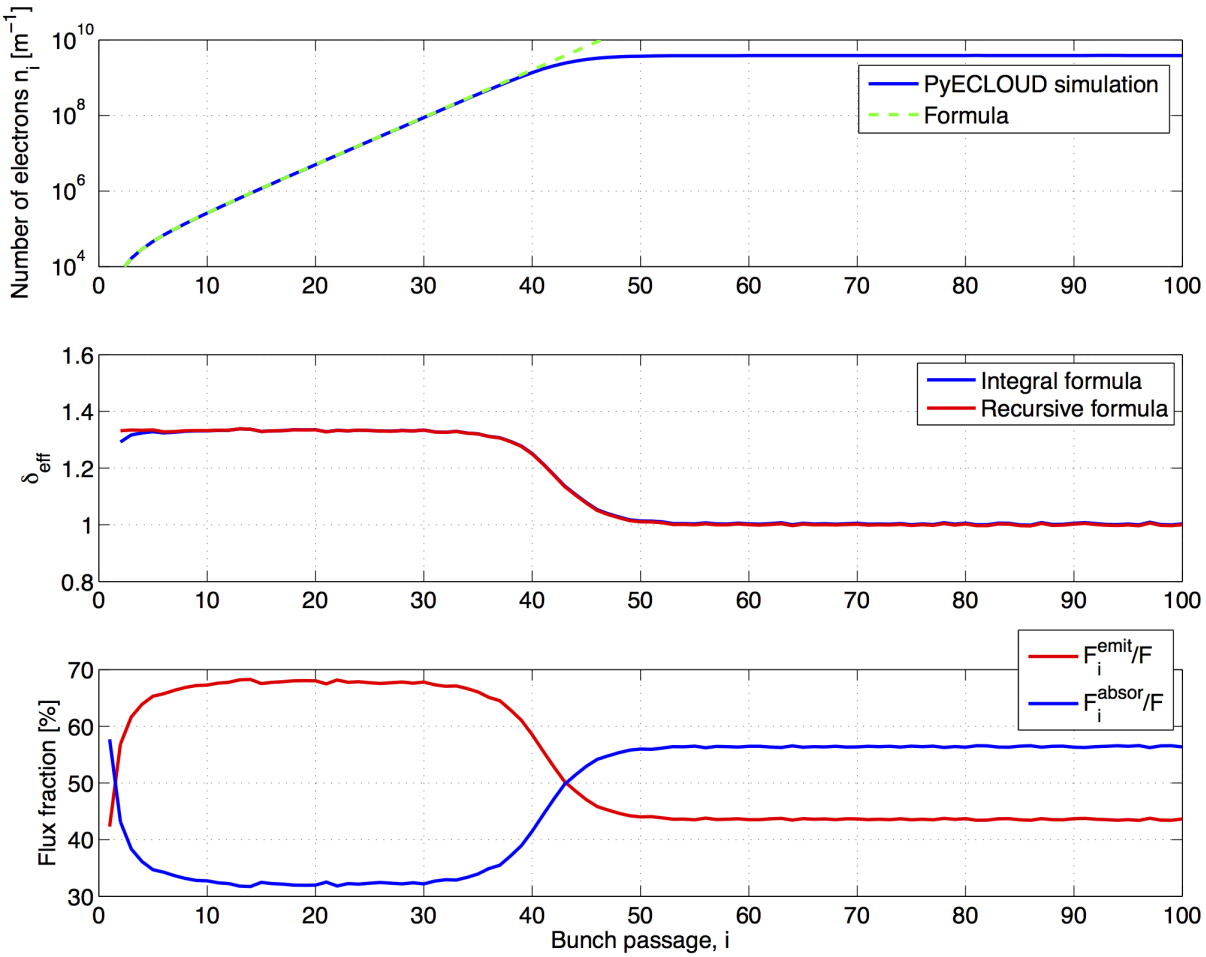


Fig. 7: Simulated electron cloud build-up for $\delta_{\max} = 1.75$. Top: number of electrons before each bunch passage (blue, directly from simulation; dashed green, estimated from Eq. (11)). Centre: δ_{eff} (both from the integral formula, Eq. (8), and the recursive formula, Eq. (3)). Bottom: fractions of the electron energy spectrum falling in the absorber or emitter regions of the secondary electron yield curve.

2.4 Effect of externally applied magnetic fields

The features of the electron cloud build-up are strongly influenced by externally applied magnetic fields, like those present in bending and focusing magnets of a particle accelerator.

It is simple to prove [15] that a non-relativistic electron moving in a uniform magnetic field of magnitude B (as for example in the case of a bending magnet) follows a helicoidal trajectory around the field lines. In a typical electron cloud build-up, the total kinetic energy of an electron is typically not larger than 2 keV (see, for example, the energy spectra in Figs. 6 and 8), which implies that the cyclotron radius never exceeds a few millimetres. This means that the electrons are practically constrained to move around the field lines. Electrons trapped by different field lines will receive different kicks from the passing bunches according to their horizontal positions, giving rise to different efficiencies for the multipacting process. If the energies to which the electrons are accelerated to the wall span from a value above $E_{\max}$ at the centre of the chamber and a value below $E_{\max}$ at the sides of the chamber, there will be two horizontal positions (symmetrical with respect to the chamber axis), for which the production of secondary electrons is maximally efficient. This generates the characteristic two-stripe pattern of the electron density in the chamber, like the one shown in Fig. 11 (left-hand side).

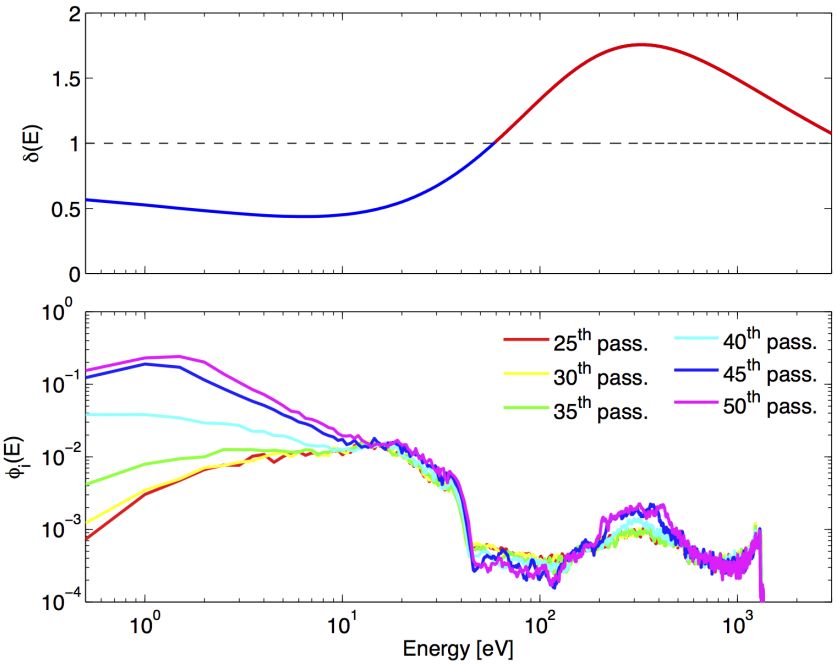


Fig. 8: Top: secondary electron yield curve. Bottom: energy spectrum $\phi_i(E)$ for different bunch passages

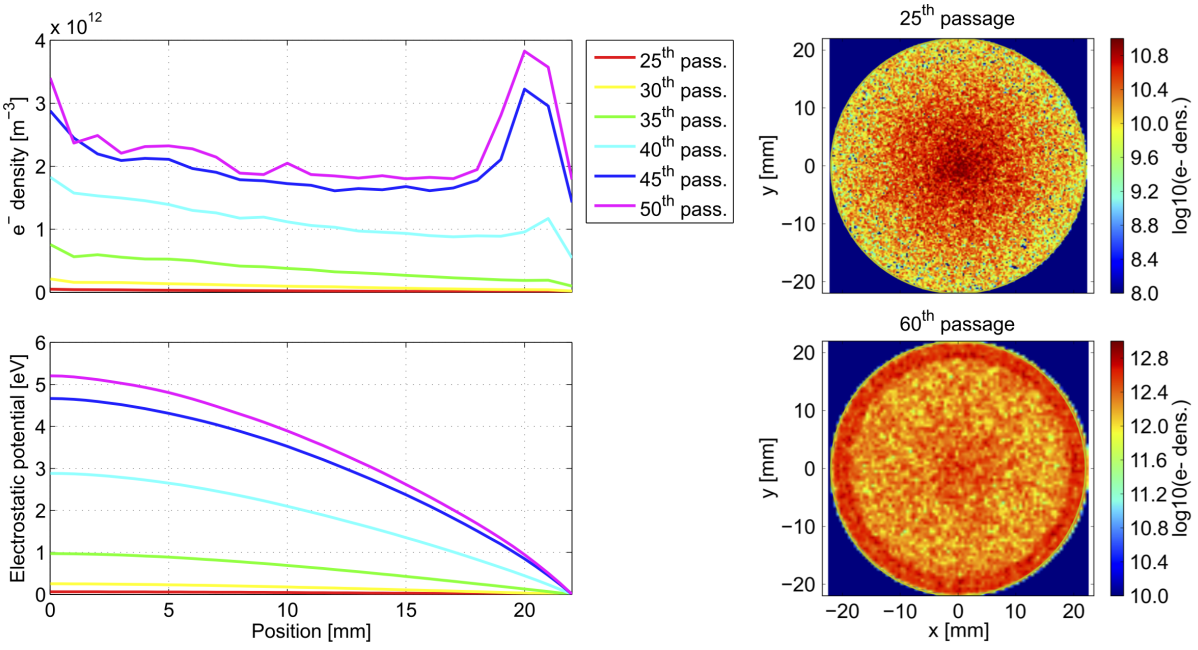


Fig. 9: Simulated electron cloud build-up for $\delta_{\max} = 1.75$. Left-hand side: electron density and electrostatic potential as a function of the distance from the centre and of the bunch passage. Right-hand side: snapshots of the electron density in the chamber, one taken immediately before a bunch passage during the exponential increase (top) and one taken immediately before a bunch passage during the saturation phase (bottom).

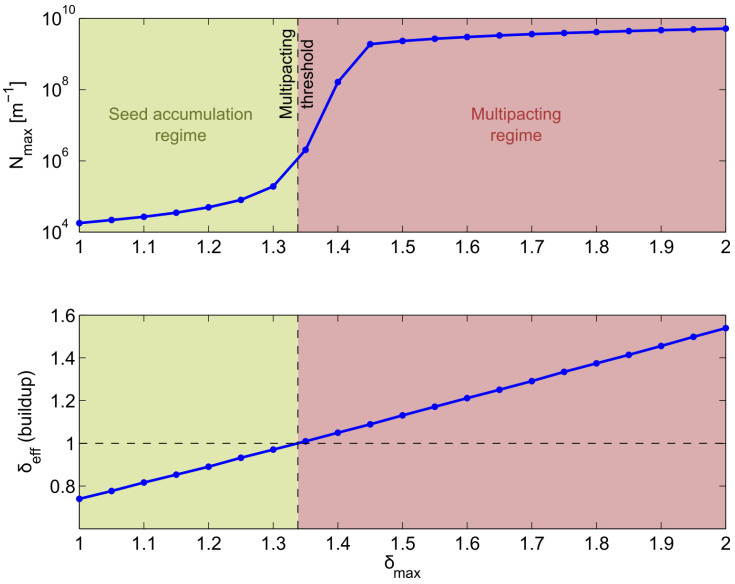


Fig. 10: Maximum number of electrons in the chamber and δ_{eff} coefficient as a function of the secondary electron yield parameter.

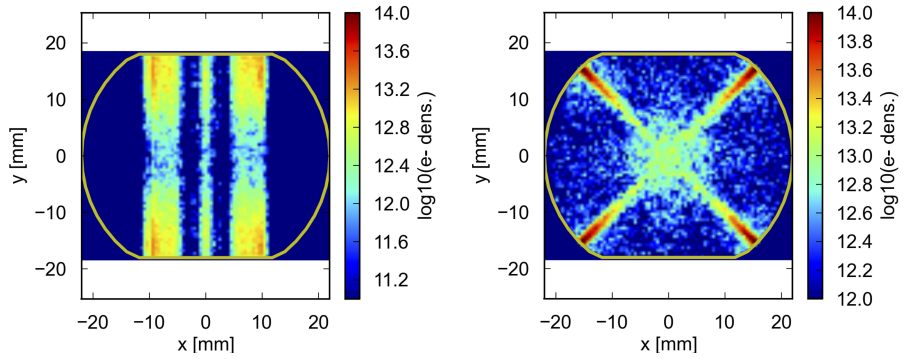


Fig. 11: Snapshots of the electron cloud density in a dipole (left-hand side) and a quadrupole (right-hand side) magnet of the LHC.

Similar effects are also observed in quadrupole magnets. For example, with the chamber and beam parameters of LHC, the electron density shows an X-like shape as shown in Fig. 11 (b). In the case of quadrupoles, the presence of a magnetic field gradient can also trigger electron-trapping mechanisms, which can make the electron cloud build-up even more severe by helping electron survival between the passage of successive bunches [16].

3 Impact of electron cloud effects on the accelerator's performances

The presence of an electron cloud in the beam chamber of a particle accelerator can limit its achievable performance through different effects, which will be briefly reviewed next. In general, the effects of the electron cloud in a particle accelerator can be classified as:

- **global:** the electron cloud is present in a large fraction of the machine and can significantly influence the beam dynamics;
- **local:** the electron cloud is only generated in certain machine elements (owing to their geometry or wall properties). Its impact on the beam dynamics is usually negligible, but it can nevertheless be responsible for local (detrimental) phenomena.

3.1 Impact on beam dynamics: coherent and incoherent effects

When the electron cloud covers a significant fraction of a machine, the integrated effect of its electric forces on the particle beam affects the collective beam motion, leading to a coherent tune shift, as well as to the onset of different types of transverse coherent instability above a certain electron density threshold. When a particle beam suffers a transverse instability (horizontal or vertical), the beam signal seen on a beam position monitor increases exponentially and the unstable motion eventually leads to emittance blow-up or fast beam loss. Electron cloud effects can, obviously, only appear in a machine operating with long trains of bunches, because, as was explained in the previous sections, the electron cloud only builds up and reaches saturation after several bunch passages. Despite this, both coupled-bunch and single-bunch phenomena (typically affecting only the last bunches in a long train) have been observed in running machines and were studied in the past, showing that the electron cloud can indeed be the source of unconventional wake fields, which affect the beam dynamics in a similar fashion as an impedance source. A multibunch dipole-mode instability was observed at the KEK Photon Factory on positron beam operation, and was subsequently explained as an effect of the variation in the electron cloud centroid position induced by an offset bunch, which can feed into the motion of subsequent bunches in an unstable loop [17]. More studies on this subject were carried out for the Beijing Electron Positron Collider and for the PEP-II B factory. Coupled-bunch instabilities in the horizontal plane were also observed at CERN, first at the Super Proton Synchrotron with LHC-type beams [18], and then at the Proton Synchrotron in the last phase of production of the LHC-type beams [19]. They were also recorded at the LHC during the first injections of trains of 48 bunches of a 25 ns spaced beam [20]. Fortunately, owing to their coupled-bunch nature, these instabilities can usually be damped by a transverse feedback system without posing excessively stringent requirements on its bandwidth.

With a similar mechanism, however, an electron cloud inside the beam pipe can also be the origin of a short-range wake field for a bunch that goes through it, giving rise to head–tail coupling and single-bunch instabilities. Since this mechanism relies on the pre-existence of an electron cloud when the bunch arrives, it can obviously only affect the bunches at the tail of a long train, such that the electron cloud has formed with the passage of the preceding bunches. Assuming that the bunch goes into the electron cloud with its head slightly offset, a global net force will act on the electrons around the head centroid position and consequently an accumulation of electrons will take place in that region. The newly reconfigured electron distribution will thus kick the following bunch particles towards the higher-density region. The motion of the head will therefore be transmitted and potentially amplified at the tail of the bunch. The tail deflection will then increase over successive turns and will eventually transfer back to the head, thanks to the longitudinal mixing given by the synchrotron motion (after a few synchrotron periods). This head–tail coupling mechanism naturally follows the oscillation of the electrons in the bunch potential and therefore the oscillation frequency of the associated wake can be roughly expressed as

$$\omega_{ey(x)} = \sqrt{\frac{N r_{\mathrm{e}} c^2}{2\sigma_{y(x)}\sigma_z(\sigma_x + \sigma_y)}}\,. \tag{17}$$

In this equation, N is the number of positively charged particles in the bunch, $\sigma_{x,y,z}$ its r.m.s. sizes in the three directions, and r_{e} represents the classical electron radius. The frequency given by Eq. (17) can quickly reach the gigahertz range and above, especially for high-intensity or brightness, high-energy beams made of trains of short bunches. Owing to the important high-frequency content, the conventional transverse feedback systems are usually incapable of controlling this type of electron cloud induced instability. One has to rely instead on altering the head–tail phasing through high-chromaticity settings, or on Landau damping [21] using octupole magnets. Solutions of this type, however, typically come at the expense of transverse emittance preservation and beam lifetime [22,23]. The single-bunch instability due to the electron cloud, observed in the Super Proton Synchrotron and in the LHC, happens mainly in

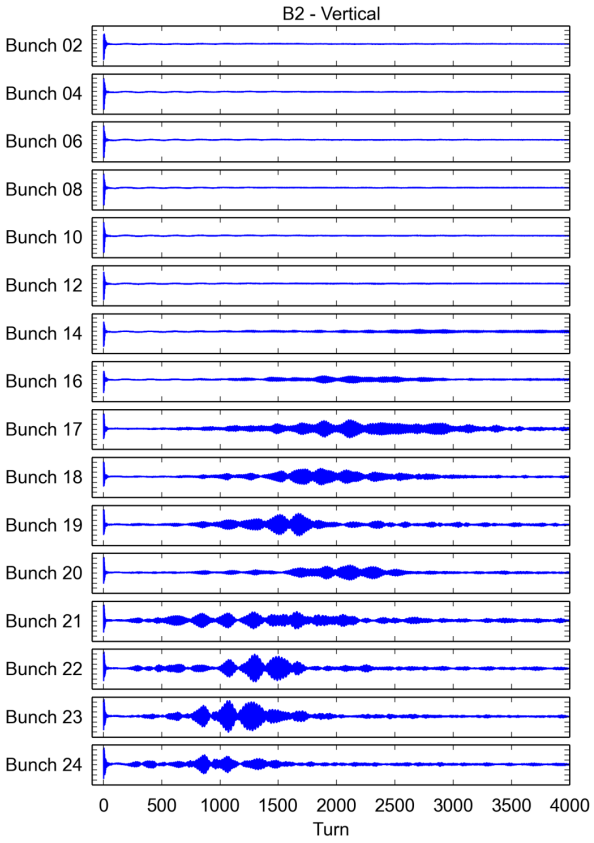


Fig. 12: Vertical position of selected bunches in a train of bunches over the first 4000 turns after injection. A transverse instability developing on the trailing bunches is clearly visible.

the vertical plane, if a large fraction of the driving electron cloud is concentrated in the dipoles, but it can equally affect the horizontal plane, if the driving electron cloud comes from drift sections or quadrupoles. In a machine like the LHC, which has 66% of its circumference covered by dipoles, the single-bunch instability will mainly affect the vertical plane, at least as long as the secondary electron yield of the beam screen of the dipoles is sufficiently high. The bunch-by-bunch position signal from the first 4000 turns after injection, as acquired from a beam position monitor for a train of 24 bunches, is given in Fig. 12 (every second bunch). It is clear that, while the first 12 bunches are stable, an unstable signal begins to appear after bunch 14 and the rise time of the instability tends to become shorter while moving to the tail of the train.

It must be noted here that this simple picture of the electron cloud single-bunch instability only applies for zero chromaticity. With non-zero chromaticity, the situation becomes more complex and the presence of an electron cloud can favour the onset of 'classical' head–tail instabilities in either plane, which can be damped with a classical transverse feedback system if the mode number is low enough to be handled by the system.

Even when the beam remains transversely stable, either because the integrated electron cloud density is low enough or thanks to stabilizing mechanisms (chromaticity, Landau damping, transverse feedback), its interaction with the electron cloud can drive incoherent effects, which slowly degrade the beam quality. These effects are usually caused by the fact that, even if the beam as a whole does not respond coherently to the electron cloud excitation, the single particles are still detuned by the additional transverse force coming from the electron cloud (usually focusing and strongly non-linear, but in general dependent on the detailed electron distribution) and their tunes can be individually pushed onto

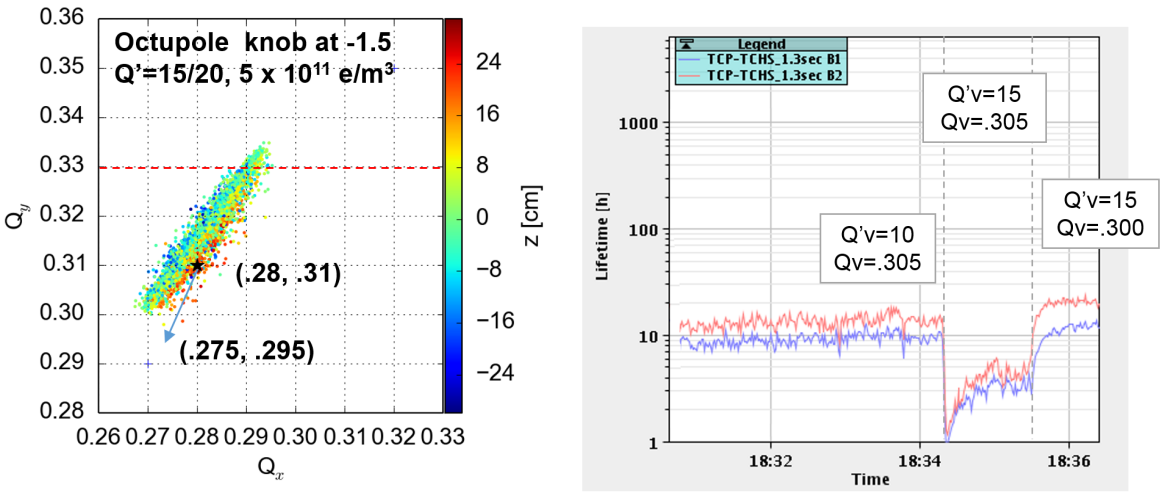


Fig. 13: Left-hand side: tune footprint at 450 GeV, as obtained from PyECLOUD-PyHEADTAIL simulations. Right-hand side: beam lifetime measured with 25 ns beams in the LHC for different settings of vertical tune and chromaticity.

resonance lines with consequent growth of the particles' amplitudes. Therefore, the resulting tune spread from the electron cloud can be the origin of such phenomena as slow emittance blow-up or slow particle losses, which are particularly worrying in storage rings and colliders, where the aim is to store the beam in the ring for a very long time (several hours), while preserving beam quality, as much as possible, and limiting all types of unwanted beam loss. An example of an incoherent effect in the LHC is illustrated in Fig. 13. The left-hand plot shows the calculated tune spread of a single bunch in the LHC at injection energy (450 GeV), assuming the operational settings for chromaticity (15 units in the horizontal plane and 20 units in the vertical plane) and octupole currents (20 A), and in addition an electron cloud density of 5×10^{11} m^{-3} all around the machine dipoles. The nominal working point is (0.28, 0.31). The effect of the octupoles on the tune footprint is quite negligible compared with the one imprinted by chromaticity, while the electron cloud makes the tune spread asymmetrical around the nominal tunes by pushing the footprint towards higher tune values in the vertical plane. The visible consequence of the electron cloud contribution to the footprint is that some particles come to cross the third-order resonance. In this configuration, important losses are expected in the LHC, mainly affecting the bunches at the tails of the long injected trains. The right plot shows an experiment in the machine, in which a strong degradation of the beam lifetime was observed when increasing the vertical chromaticity from 10 to 15 units, which could be recovered (and even slightly improved) by moving the vertical tune down by 0.005.

3.2 Other effects

The presence of an electron cloud in a particle accelerator can be also revealed by the following observables.

– **Vacuum degradation:** The electron flux on the chamber's wall stimulates the desorption of gas molecules from the surface (dependent on the desorption yield of the wall for electron impingement), which results in an increased residual gas density in the beam chamber, and therefore in a pressure increase. This is obviously a local effect that can take place only in specific parts of the machine, which are prone to electron cloud formation as a result of their geometry or surface properties, or in extended machine sectors, if the beam chamber geometry and secondary electron yield are such as to support widespread electron cloud build-up. The pressure rise is associated with several deleterious effects, such as larger equipment irradiation, worse background in the

experimental areas, increased probability of breakdown in high-voltage devices like kickers or electrostatic septa, and impact on the beam lifetime [24].

- **Beam energy loss and heat load:** The electrons accelerated in the beam field subtract energy from the beam and also deposit a large fraction of this energy on the chamber's wall when they hit it and produce secondary electrons (usually it takes one electron with an energy of a few hundred electronvolts to produce one or more electrons with energies of a few electronvolts). Therefore, two different observables can be associated with this process:
  - first, if the amount of integrated electron cloud on the beam path is sufficiently large in the accelerator, then the total energy loss of a bunch due to the electron cloud over one turn can become significant. Since the lost energy must then be restored by the RF system, this can result in a measurable contribution to the stable (or synchronous) phase shift of the bunch in its RF bucket (beside the contributions coming from beam loading and the longitudinal impedance). This is a global measurement and provides information on the total amount of electron cloud present in the machine. Usually, this is assumed to be detectable if it is at least few tenths of a degree;
  - second, the energy deposited by the electrons on the chamber's wall heats it up and the additional heat load could be measured (either as an increase of the chamber wall's temperature or as an increased power required from the cooling system to keep the chamber at a desired temperature). This effect is only local and, while it is typically negligible in room temperature accelerator components, it can become a serious issue in devices operating at cryogenic temperatures, like the superconducting magnets of the LHC. Here, the heat load induced by the electron cloud can even reach the cooling capacity limit of the cryogenic system [25].
- **Impact on beam diagnostics:** The presence of an undesired electron flux at the frequency of the bunch spacing can be a source of spurious signals, and therefore malfunctions, on beam diagnostic devices, such as pick-ups (beam position monitors) and beam gas ionization chambers [26].

All these effects have been observed at the LHC and in its injector chain.

4 Mitigation strategies

In some cases, the accumulation of primary electrons alone inside the beam chamber of an accelerator can be the source of detrimental effects, even in the absence of beam-induced multipacting. Since this mainly happens for photoelectron production in bending or wiggler chambers of very-high-energy beams, an obvious mitigation technique would be to reduce the photoelectron production rate by either using surfaces with naturally low photoelectron yield or by guiding the photons into a region where the produced photoelectrons can then do no harm. When the electron cloud formation is mainly caused by secondary emission, it is necessary to find methods to reduce the effective secondary electron yield of the inner chamber walls in order to suppress or at least reduce the electron cloud build-up, and thereby limit its adverse effects. In either case, other viable options for mitigation could be:

- to alter the electron dynamics to avoid large fluxes of high-energy electrons towards the chamber walls. This can be achieved with electric fields (e.g., clearing electrodes) or magnetic fields (e.g., solenoids);
- not to suppress the electron cloud, merely alleviate its effects on the beam or on the devices that could be affected.

For machines like the KEKB photon factory, the LHC, and future circular colliders, the primary production of photoelectrons would be so high that the electron cloud could reach saturation within a few bunch passages, even without multipacting, if no countermeasures were taken. The solutions implemented in current machines are an antechamber (KEKB) or, for dipole fields, a sawtooth pattern impressed on the chamber wall (LHC). Weak solenoids of the order of 50 G are another possibility, which

was also successfully implemented at KEKB. The solenoids do not really reduce the photoemission, but they quickly loop the emitted photoelectrons back to the wall, thus mitigating the subsequent beam–electron interaction. For future circular colliders, different schemes are under study, based on photon absorbers or stoppers to intercept the photons at controlled locations and limit the associated photoelectron production, or on a novel design of the vacuum chamber with lateral slits shaped to trap the photons and subsequently shield the photoelectrons from the beam field. Electrons generated by beam loss at a collimator can be controlled by solenoids or clearing electrodes. For example, the SNS project has installed solenoids along the collimator straights.

When the electron cloud build-up is dominated by the process of secondary emission, the surface of the inner wall of the vacuum chamber needs to be treated such that its effective secondary electron yield is minimized, and ideally reduced to a value below 1. This can be achieved by either coating (i.e., changing the chemical properties of the exposed surface) or machining (i.e., changing the geometrical properties of the exposed surface). A well-established method to reduce multipacting in RF couplers is coating with TiN, a material that has been proven to condition to very low values of secondary electron yield. The thickness of the coating is of the order of a micrometre, which should not alter the resistive impedance seen by the beam.

A more favourable getter material made from TiZrV (a non-evaporable getter) was developed at CERN and has the advantage of pumping while having low secondary electron yield. This non-evaporable getter is characterized by greater structural stability than TiN, as well as a low activation temperature. The warm sections of the LHC, especially those around the experimental areas (about 20% of the circumference), have been coated with this material [27]. The non-evaporable getter coating has already been widely used at several synchrotron light sources around the world (both in insertion devices and for general coating of the vacuum chambers) to guarantee ultrahigh vacuum and improve the beam lifetime while reducing the probability of exciting fast beam ion instabilities.

In the last 10 years, impressive work has been done at CERN to develop a new type of coating with amorphous carbon, which does not require activation, has an intrinsically very low secondary electron yield and does not degrade with time [28]. This coating has been widely tested at the CERN Super Proton Synchrotron, where the suppression of the electron cloud was successfully proven in dedicated strip monitors, as well as in a few main bending units. In particular, an amorphous-carbon coated liner has remained installed in a strip monitor since 2007 and no electron cloud signal was ever measured in it, even after long technical stops and extensive machine venting. This confirms that the amorphous-carbon coating can preserve its low secondary electron yield even after being long exposed to air.

More recently, another type of procedure based on laser engineering of the surface (applicable, for example, to copper, stainless steel, and aluminium) has been proposed to reduce the effective secondary electron yield. The laser treatment, which imprints a surface topography made of organized microstructures, has the advantage of relatively easy application and possible retrofitting in existing machines [29]. In parallel with the research on coating and laser treating, several authors have also proposed a method to suppress multipacting by machining the wall surface to produce macroscopic grooves on it. These grooves essentially act as electron traps, as the electrons emitted by the surface are very likely to be re-absorbed quickly before they can be accelerated in the beam field. Much optimization work has been done to define the shape and the size of the grooves so as to obtain the best electron cloud suppression [30].

Finally, another way to reduce the secondary electron yield of the inner surface of an accelerator vacuum chamber is to rely on its conditioning with time, thanks to electron bombardment during beam operation with electron cloud. This technique is called ‘scrubbing’ and is based on the experimental observation that the secondary electron yield of a surface exposed to a continuous flux of high-energy electrons decreases with the deposited electron dose. While this decrease is usually very fast at the beginning for large values of the secondary electron yield, since scrubbing means physically removing the external layers of molecules and oxides present on the surface of the bare metal, it then tends to slow

Fig. 14: Evolution of (top) beam intensity, (middle) average heat load in the arc magnets, and (bottom) heat load normalized to the beam intensity during the intensity ramp-up, with 25 ns beams

down exponentially and eventually requires enormous doses to make tiny steps in the region of secondary electron yields below 1.3–1.4.

Scrubbing has been widely used at CERN for the Super Proton Synchrotron and LHC, both of which have reached nominal operation with 25 ns beams thanks to extended scrubbing runs. In the case of the LHC, an important part of the scrubbing process has been carried out not only through dedicated scrubbing periods, necessary nevertheless to prepare the machine to operate with 25 ns beams, but also through physics stores with 25 ns beams. Assuming the heat load in the cold arcs to be a measure of the amount of electron cloud present in the machine, Fig. 14 displays the evolution of this quantity over two months during the 2015 run, when increasing numbers of bunches were injected into the LHC and brought to collision at 6.5 TeV. The scrubbing of the surface of the beam screen in the arcs is visible as the decrease of the heat load normalized to the total beam current, which has taken place with a time constant of weeks and has led to about half the value over the full two months' period. For completeness, it must be highlighted here that the decrease of the normalized heat load is not fully ascribable to surface scrubbing, but was also aided by relaxing the filling pattern into the LHC (moving from trains of 72 bunches to trains of 36 bunches) and increasing the bunch length at top energy (as shown in the plot of 'filling pattern' and the 'target beam length' strips). These additional electron cloud mitigating measures were necessary to increase the number of bunches injected into LHC, while keeping the produced heat load in the arcs within the cooling capacity of the cryogenic system.

Multipacting may also be suppressed by solenoids, though one should pay attention to the possibility of cyclotron resonances (i.e., conditions for which the cyclotron period of the electrons in the solenoid field is a multiple of the bunch spacing). Electric clearing fields are also an efficient cure in simulations. They were already used to cure electron–proton instabilities for the coasting proton beams in the CERN ISR during the early 1970s. At the SNS, operating with long proton bunches, all beam position monitors can be biased with a clearing voltage of 1 kV. To be effective for the multipacting experienced by short bunches with close spacing, the clearing electrodes must be mounted all around

the ring at close distances and voltages of the order of 1 kV are probably required. The impedance introduced by many such devices could be an obvious showstopper. However, a continuous long wire on an insulating support would not necessarily exhibit a prohibitively large impedance. Other options for a practical implementation of electric clearing fields might involve using 'stealth' electrodes, or splitting the beam pipe into top and bottom halves, isolated from each other and held at different potentials. Biasing the two jaws of a collimator against each other is a similar idea.

Proper tailoring of the bunch filling patterns (bunch spacing, bunch trains, and bunch charges) is yet another way of achieving an acceptable electron density. The application of this technique can be two-fold. On the one hand, the arrangement of the bunches in a train can be such as to minimize the electron cloud build-up and allow an electron cloud free operation of the accelerator. In particular, a larger bunch spacing can help, or gaps within trains can reduce the electron cloud density and reset the cloud, at least partly. However, this usually comes at the expense of the total amount of beam intensity that can be accelerated in the machine, as the full potential of the available free buckets is not exploited. Examples here include mini-trains in the PEP-II, the actual bunch spacings chosen for the PEP-II and KEKB operation, which are two or three times the design spacing, the 150–75–50 ns beam configuration used until 2012 in the LHC, and finally the so-called 8b + 4e configuration for 25 ns proposed for the LHC, in which long trains of 25 ns spaced bunches are replaced with trains exhibiting gaps of four empty buckets for every eight full buckets [7]. On the other hand, special beam configurations can be put in place, with the goal of increasing as much as possible the electron cloud formation and accelerate the scrubbing process. A typical example of this approach is the use of 'doublet' beams in the Super Proton Synchrotron and LHC, i.e., beams made of 5 ns spaced bunch pairs separated from each other by 25 ns. These beams were expected, and proved, to produce a large electron cloud in both the Super Proton Synchrotron and LHC, offering the potential of possibly scrubbing the wall surfaces to secondary electron yields below the electron cloud build-up thresholds for the nominal 25 ns beams [7].

Acknowledgements

We would like to thank G. Arduini, H. Bartosik, K. Li, L. Mether, A. Romano, M. Schenk, F. Zimmermann and many others for their invaluable help in electron cloud studies at the CERN accelerator complex over the years.

References

[1] W. Fischer *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **11** (2008) 041002. https://doi.org/10.1103/PhysRevSTAB.11.041002

[2] H. Fukuma, in Proc. ECLOUD12: Joint INFN-CERN-EuCARD-AccNet Workshop on Electron-Cloud Effects, La Biodola, Isola d'Elba, Italy, edited by R. Cimino, G. Rumolo, F. Zimmermann, CERN-2013-002 (CERN, Geneva, 2013), pp. 27-30. https://doi.org/10.5170/CERN-2013-002.27

[3] M. Zobov *et al.*, in Proc. ECLOUD12: Joint INFN-CERN-EuCARD-AccNet Workshop on Electron-Cloud Effects, La Biodola, Isola d'Elba, Italy, edited by R. Cimino, G. Rumolo, F. Zimmermann, CERN-2013-002 (CERN, Geneva, 2013), pp. 259-265. https://doi.org/10.5170/CERN-2013-002.259

[4] G. Rumolo *et al.*, Electron cloud observation in the LHC, Proc. IPAC11 Int. Particle Accelerator Conf., San Sebastian, Spain, edited by Petit-Jean-Genaz et l., 2011, p. 2682.

[5] G. Rumolo *et al.*, Electron cloud effects in the LHC in 2011, Proc. Evian 2011 LHC Beam Operation Workshop, Evian, France, edited by B. Goddard and S. Dubourg, CERN-ATS-2012-083, p.165, 2011.

[6] G. Iadarola *et al.*, Electron cloud effects in 2012 in the LHC, Proc. Evian 2012 LHC Beam Operation Workshop, Evian, France, edited by B. Goddard and S. Dubourg, CERN-ATS-2013-045, p. 119, 2012.

[7] G. Iadarola *et al.*, Performance limitations from electron cloud in 2015, Proc. Evian 2015 LHC Beam Operation Workshop, Evian, France, edited by S. Dubourg and B. Goddard, CERN-ATS-2015-376, p. 101, 2015.

[8] G. Rumolo *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **4** (2001) 012801. https://doi.org/10.1103/PhysRevSTAB.4.012801

[9] B. Henrist *et al.*, in Mini Workshop on Electron Cloud Simulations for Proton and Positron Beams, 15–18 April 2002, CERN, Geneva, Switzerland, edited by G. Rumolo and F. Zimmermann, CERN-2002-001 (CERN, Geneva, 2002), pp. 75-78. https://doi.org/10.5170/CERN-2002-001.75

[10] V. Baglin *et al.*, A summary of main experimental results concerning the secondary electron emission of copper. LHC Project Report 472 (2002).

[11] R. Cimino *et al.*, *Phys. Rev. Lett.* **93** (2004) 014801. https://doi.org/10.1103/PhysRevLett.93.014801

[12] R. Cimino and I.R. Collins. *Appl. Surf. Sci.* **235** (2004) 231. https://doi.org/10.1016/j.apsusc.2004.05.270

[13] A. Kuzucan *et al.*, Secondary electron yield on cryogenic surfaces as a function of physisorbed gases, Proc. IPAC11 Int. Particle Accelerator Conf., San Sebastian, Spain, edited by Petit-Jean-Genaz et al., p. 1575, 2011.

[14] G. Iadarola and G. Rumolo, in Proc. ECLOUD12: Joint INFN-CERN-EuCARD-AccNet Workshop on Electron-Cloud Effects, La Biodola, Isola d'Elba, Italy, edited by R. Cimino, G. Rumolo, F. Zimmermann, CERN-2013-002 (CERN, Geneva, 2013), pp. 189-194. https://doi.org/10.5170/CERN-2013-002.189

[15] D. Halliday *et al.*, *Fundamentals of Physics* (John Wiley, New York, 2010).

[16] R.J. Macek *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **11** (2008) 010101. https://doi.org/10.1103/PhysRevSTAB.11.010101

[17] F. Zimmermann, Electron cloud studies for the low energy ring of KEKB, CERN-SL-Note-2000-004, 2000.

[18] G. Arduini *et al.*, The LHC proton beam in the CERN SPS: an update, 20th IEEE Particle Accelerator Conf., Portland, OR, USA, 2003. https://doi.org/10.1109/pac.2003.1288647

[19] R. Cappi *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **5** (2002) 094401. https://doi.org/10.1103/PhysRevSTAB.5.094401

[20] H. Bartosik *et al.*, in Proc. ECLOUD12: Joint INFN-CERN-EuCARD-AccNet Workshop on Electron-Cloud Effects, La Biodola, Isola d'Elba, Italy, edited by R. Cimino, G. Rumolo, F. Zimmermann, CERN-2013-002 (CERN, Geneva, 2013), pp. 211-217. https://doi.org/10.5170/CERN-2013-002.211

[21] A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).

[22] K. Li and G. Rumolo, Review of beam instabilities in the presence of electron clouds in the LHC, Proc. IPAC11 Int. Particle Accelerator Conf., San Sebastian, Spain, edited by Petit-Jean-Genaz et al., p. 760, 2011.

[23] K. Li *et al.*, LHC injection instabilities, Presentation at 2-day internal review of LHC performance limitations during run I, 25–26 Sept. 2013, CERN, 2013.

[24] G. Bregliozzi *et al.*, Vacuum pressure observations during 2011 proton run, Proc. Evian 2011 LHC Beam Operation Workshop, Evian, France, edited by B. Goddard and S. Dubourg, CERN-ATS-2013-083, p. 177, 2011.

[25] L. Tavian, Performance limitations of the LHC cryogenics: 2012 review and 2015 outlook, Proc. Evian 2012 LHC Beam Operation Workshop, Evian, France, edited by B. Goddard and S. Dubourg, CERN-ATS-2013-045, p. 129, 2012.

[26] W. Höfle, Observations of the electron cloud effect on pick-up signals in the SPS. 10th Workshop on LEP-SPS Performance, Chamonix, France, edited by P. Le Roux, J. Poole and T. Truchet, CERN-SL-2000-007-DI, p. 102, 2000.

[27] O. Bruning *et al.*, *LHC Design Report* (CERN, Geneva, 2004).

[28] C. Yin Vallgren *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **14** (2011) 071001. https://doi.org/10.1103/PhysRevSTAB.14.071001

[29] R. Valizadeh *et al.*, *Appl. Phys. Lett.* **105** (2014) 231605. https://doi.org/10.1063/1.4902993

[30] G. Stupakov and M. Pivi, in Proc. 31st Advanced ICFA Beam Dynamics Workshop on Electron-Cloud Effects, 19–23 April 2004, Napa, CA, USA, CERN-2005-001 (CERN, Geneva, 2005), pp. 139-141. https://doi.org/10.5170/CERN-2005-001.139

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam–Beam Effects in Linear Colliders

D. Schulte
CERN, Geneva, Switzerland

Abstract
Linear colliders are promising candidates for future high-energy lepton colliders. Two options are studied in global collaborations, the International Linear Collider (ILC) and the Compact Linear Collider (CLIC). Beam–beam effects are an important driver of linear collider design choices and they affect the performance of physics experiments. The lecture introduces the main physics relevant for the beam collision and the impact on the parameter choice.

Keywords
Linear Collider; beam-beam; pinch effect; beamstrahlung; ILC; CLIC.

1 Linear colliders

To date only one linear collider has been operated for physics, the Stanford linear collider (SLC) (no definitive reference is available for the SLC, but one can use the report [1] and the references therein as a starting point). It has been operated at the Z-resonance. Two future linear collider projects are under consideration, the International Linear Collider (ILC [2–6]) and the Compact Linear Collider (CLIC [7,8]). The ILC aims at a centre-of-mass energy of $500\,\mathrm{GeV}$, potentially starting at $250\,\mathrm{GeV}$. The CLIC is foreseen to be implemented in three stages ranging from a centre-of-mass energy of $380\,\mathrm{GeV}$ up to $3\,\mathrm{TeV}$.

In circular electron–positron colliders synchrotron radiation becomes a severe issue as the beam energy increases, since the radiation power increases with the fourth power of the energy. This can be avoided in linear colliders. A generic example of a linear collider is shown in Fig. 1. The beams are produced in an electron and a positron source, respectively. They are slightly accelerated and transported to a damping ring. Here their emittance is reduced to very small values, especially in the vertical plane. Then the beams are transported through the ring to main linac system (RTML). During the transport they are slightly more accelerated and compressed longitudinally. In the main linac they are accelerated to full energy. In the beam delivery system (BDS) the beams are then focused to very small sizes at the collision point. Then they are disposed of in beam dumps.

The main challenges of a linear collider are first to achieve the beam energy in the main linac. This requires very high gradients for the acceleration. The second challenge is to achieve the high luminosity in a single pass. This requires very dense beams at the collision point. Both ILC and CLIC will deliver short pulses of bunches that collide with longer intervals between pulses.

The ILC is based on the use of superconducting cavities to accelerate the beams. These allow the use of long beam pulses. To provide the accelerating field, the cavity needs to be filled with energy. This energy is lost very slowly in the walls of the cavity and hence one can afford a long pulse. In contrast, the CLIC is based on high-gradient normal-conducting accelerating structures. These require very short pulses, since the energy in the accelerating structures is lost rapidly in the copper walls. To achieve sufficient efficiency it is therefore necessary to use very short pulses and to increase the beam current in the pulse as much as possible. This requires short distances between the bunches. The advantage of the normal-conducting accelerating structures is that they allow us to use higher accelerating fields than superconducting cavities (a factor of about three between CLIC and ILC). To achieve multi-TeV energies at practical machine length and cost thus requires the use of normal-conducting technology. The main beam parameters for linear colliders are given in Table 1.

https://doi.org/10.23730/CYRSP-2017-003.431

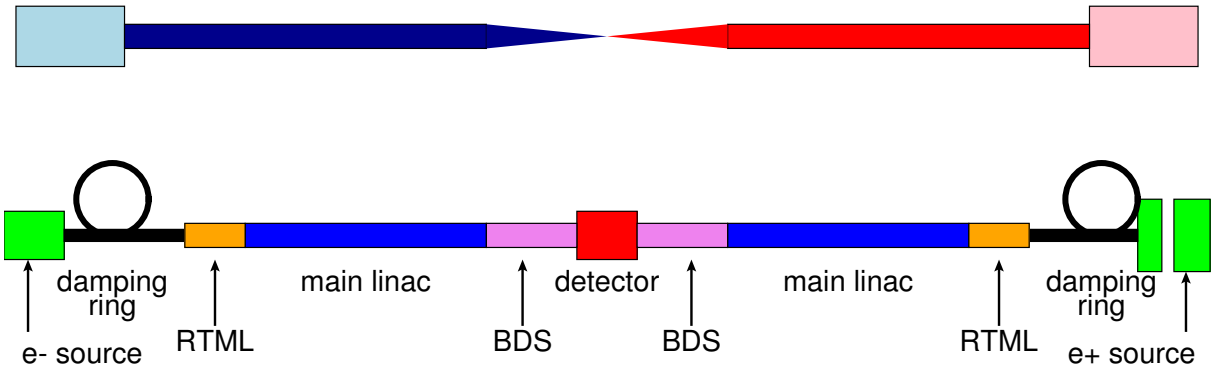


Fig. 1: Generic layout of a linear collider. The beams are illustrated at the top. They are produced with relatively large emittances, which are then reduced in the damping rings. From there to the interaction point the emittances increase somewhat before the beams are focused at the interaction point.

Table 1: The main parameters of SLC, ILC and CLIC. More parameter sets exist for ILC and CLIC at different energies. The CLIC set at 3 TeV includes the higher order optics and radiation effects in the beam delivery system.

Parameter	Symbol [unit]	SLC	ILC	CLIC	CLIC
Centre-of-mass energy	$E_{\rm cm}$ [GeV]	92	500	380	3000
Geometric luminosity	$\mathcal{L}_{\rm geom}$ [$10^{34}\,{\rm cm}^{-2}\,{\rm s}^{-1}$]	0.00015	0.75	0.8	4.3
Total luminosity	$\mathcal{L}$ [$10^{34}\,{\rm cm}^{-2}\,{\rm s}^{-1}$]	0.0003	1.8	1.5	6
Luminosity in peak	$\mathcal{L}_{0.01}$ [$10^{34}\,{\rm cm}^{-2}\,{\rm s}^{-1}$]	0.0003	1	0.9	2
Gradient	G [MV m^{-1}]	20	31.5	72	100
Particles per bunch	N [10^9]	37	20	5.2	3.72
Bunch length	σ_z [μm]	1000	300	70	44
Collision beam size	$\sigma_{x,y}$ [nm nm^{-1}]	1700/600	474/5.9	149/2.9	40/1
Emittance	$\epsilon_{x,y}$ [μm nm^{-1}]	3/3000	10/35	0.95/30	0.66/20
Beta function	$\beta_{x,y}$ [mm mm^{-1}]	100/10	11/0.48	8.2/0.1	6/0.07
Bunches per pulse	$n_{\rm b}$	1	1312	352	312
Distance between bunches	Δz [ns]	—	554	0.5	0.5
Repetition rate	$f_{\rm r}$ [Hz]	120	5	50	50
Horizontal disruption	D_x	0.6	0.3	0.24	0.2
Vertical disruption	D_y	1.7	24.3	12.5	7.6
Photons per beam particle	n_γ	—	1.9	1.5	2.1
Average photon energy	$\langle E_\gamma/E_0\rangle$ [%]	—	2.4	4.5	13
Coherent pairs	$N_{\rm coh}$ [10^8]	—	—	—	6.8
Their energy	$E_{\rm coh}$ [10^8 TeV]	—	—	—	2.1
Incoherent pairs	$N_{\rm incoh}$ [10^3]	—	196	58	300
Their energy	$E_{\rm incoh}$ [TeV]	—	484	187	2.3×10^4

2 Luminosity and parameter drivers

The luminosity target for linear colliders is around $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ following the requests of the experiments. In Table 1, one can note that ILC and CLIC use flat beams to achieve this ambitious luminosity target and that the vertical beam size is only of the order of a nanometre. In the following we will discuss the reason for this.

The luminosity $\mathcal{L}$ in a linear collider is given by the following formula:

$$\mathcal{L} = H_\mathrm{D} \frac{N^2}{4\pi\sigma_x\sigma_y} n_\mathrm{b} f_\mathrm{r}. \tag{1}$$

Here, N is the number of particles per bunch, $\sigma_{x,y}$ are the horizontal and vertical beam sizes at the collision point, n_b is the number of bunches per train, f_r is the rate of trains per second and H_D is a factor that contains the impact of beam–beam forces and other relevant effects. The factor H_D is typically in the order of 1.5–2. It is useful to rewrite the formula in the following form:

$$\mathcal{L} \propto H_\mathrm{D} \frac{N}{\sigma_x\sigma_y} N n_\mathrm{b} f_\mathrm{r}. \tag{2}$$

The term $N n_\mathrm{b} f_\mathrm{r}$ represents the beam current. Its upper limit arises from the power consumption of the collider and the efficiency to turn this power into beam power. It is therefore important to maximize the luminosity per beam current, i.e. the factor $N/(\sigma_x\sigma_y)$. However, a lower limit to the beam size arises from the beam–beam effects.

At collision, the beams are so dense that they generate very strong electromagnetic fields. In an electron–positron collider they focus each other; Fig. 2 illustrates this. In circular colliders this deflection is quite small and can be understood as a thin-lens kick. In a linear collider the beams are so dense that the particles move strongly during the collision. This so-called pinch effect reduces the effective beam size and leads to an increase in luminosity. It has been experimentally verified in the SLC [1], where under some conditions the luminosity was more than doubled. But this motion also leads to the emission of radiation, the so-called beamstrahlung, which alters the energy of the colliding beam particles. As a consequence the luminosity is not only provided at full beam energy but also at lower energies: a luminosity spectrum forms; see Fig. 2. This places some constraint on the beam parameters.

One can rewrite the luminosity formula further:

$$\mathcal{L} \propto H_\mathrm{D} \frac{N}{\sigma_x} N n_\mathrm{b} f_\mathrm{r} \frac{1}{\sigma_y}. \tag{3}$$

In the following we will see that the term N/σ_x is related to the beamstrahlung emitted. Its upper limit is defined by the requirements of the physics experiments. The term $1/\sigma_y$ finally is limited by the ability to achieve and preserve a small beam emittance and to squeeze the beam to a very small size.

3 Basic beam–beam effects

In the following we discuss the fundamentals of the pinch effect and the beamstrahlung. We will discuss head-on collisions even if the beams collide at an angle. The angle can however be neglected because in the proposed designs it is compensated by crab crossing.

3.1 Pinch effect

3.1.1 Disruption

The focusing effect of the colliding electron and positron bunches can be described by the so-called disruption parameter. First, let us assume that the effect is so weak that the particles receive a transverse kick during the passage of the oncoming bunch but that they do not change their positions. If a particle

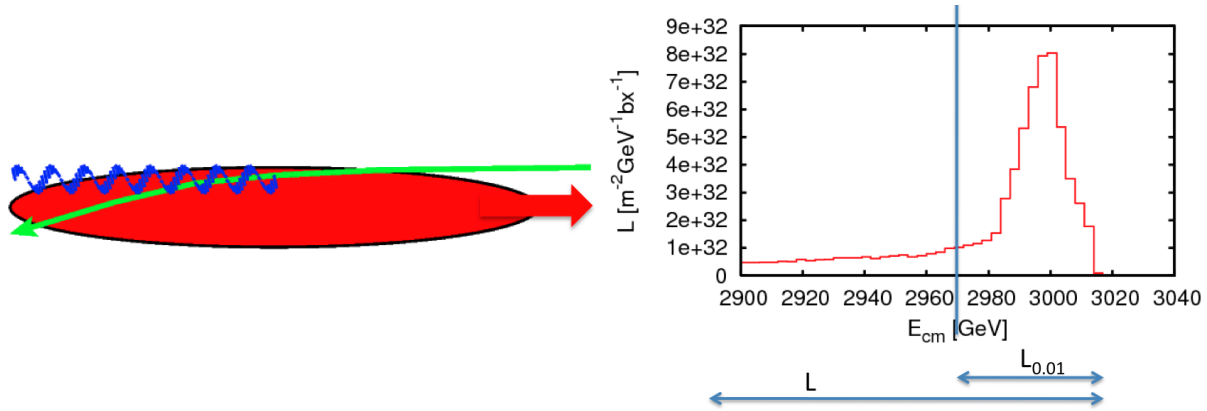


Fig. 2: Left: an electron is deflected by the fields of the oncoming bunch and emits a beamstrahlung photon. Right: the luminosity spectrum in CLIC at 3 TeV. The part of the luminosity above 99% of the nominal centre-of-mass energy is called $\mathcal{L}_{0.01}$. Values are given as integrated luminosity per energy band and bunch crossing (bx).

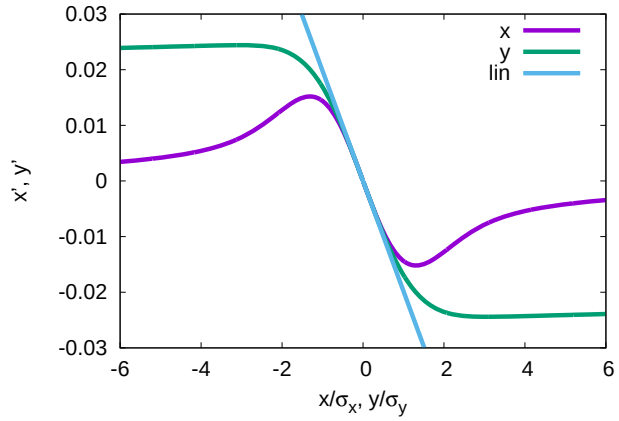


Fig. 3: The deflection of particles due to the fields of the oncoming bunch

is close to the centre of the beam, the deflection will depend linearly on the offset, while it will grow less than linearly and even become smaller at larger offset. See Fig. 3 for an example of a very flat beam. One can easily calculate the deflection of a particle in the linear regime:

$$\left.\frac{\mathrm{d}x}{\mathrm{d}z}\right|_{\text{final}} = \frac{2Nr_\mathrm{e}}{\gamma\sigma_x(\sigma_x+\sigma_y)}x, \quad \left.\frac{\mathrm{d}y}{\mathrm{d}z}\right|_{\text{final}} = \frac{2Nr_\mathrm{e}}{\gamma\sigma_y(\sigma_x+\sigma_y)}y.$$

Here, $r_\mathrm{e} \approx 2.8$ fm is the classical electron radius and γ is the relativistic factor of the particles. In each plane, the core of the beam is thus focused into one point with a distance $f_{x,y}$ after the collision point. We define the disruption parameters $D_{x,y} = \sigma_z/f_{x,y}$. Hence,

$$D_x = \frac{2Nr_\mathrm{e}\sigma_z}{\gamma\sigma_x(\sigma_x+\sigma_y)}. \quad D_y = \frac{2Nr_\mathrm{e}\sigma_z}{\gamma\sigma_y(\sigma_x+\sigma_y)}.$$

For $D_{x,y} \ll 1$, the particles will not move significantly during the collision, so our model is correct. For $D_{x,y}$, the motion during the collision will be large, so we need to take it into account. In linear colliders one typically finds that the horizontal disruption is small and the vertical disruption is large; see Table 1.

3.1.2 Simulation codes

For large disruption analytical models are difficult to develop and computer codes are used to simulate the effect. The two most widely used codes are GUINEA-PIG [10] and CAIN [9]. These codes represent

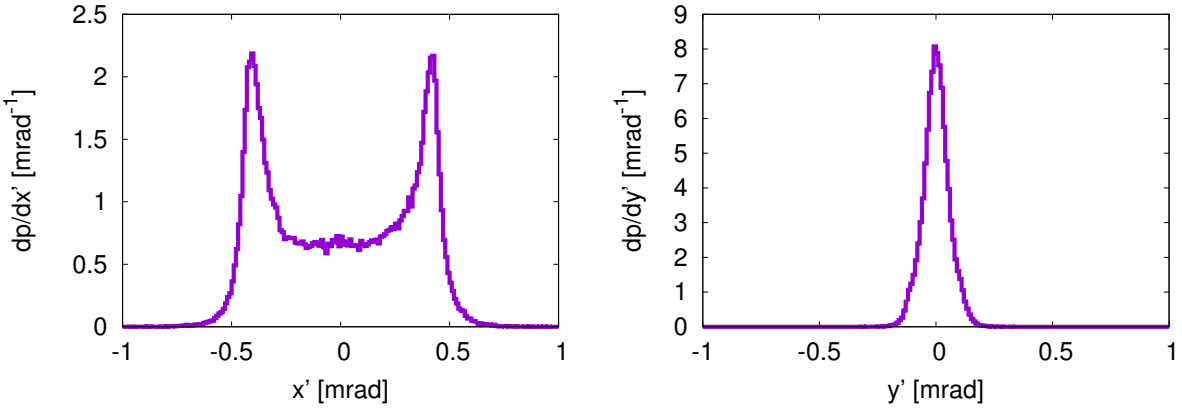


Fig. 4: The angular distribution of the particles of the spent beam

the beam by a number of macroparticles. Since the fields of the particles are transverse one can divide the two beams into slices. At any time, only the two slices of the beams which are at the same z-position can interact with each other. A particle is affected only by the fields of the other beam. After this interaction they are moved forward to interact with the next slice. The interaction within a slice can be treated as a two-dimensional problem of line charges, which can be solved using particles-in-cell or clouds-in-cell methods. The slices are divided transversely into cells by a grid. In the first step the charges are distributed over this grid. In the second step the fields at the grid points are computed. Using clouds, i.e. extended particles, instead of point-like particles helps to suppress high-frequency components that are an artefact of the finite resolution of the grid. In the third step the forces on each particle are calculated and the particles are moved forward. The grid is also used to calculate the luminosity. The particles of the two beams that are in the same cell collide with each other.

The codes also include the beamstrahlung effect, which is described in more detail below, and the generation of the different electron–positron background processes. GUINEA-PIG also includes the generation of hadronic background and the generation of muons; both processes will not be discussed here.

3.1.3 Typical results

Table 1 shows the luminosities for ILC and CLIC. The geometric value is calculated using the simplified formula

$$\mathcal{L}_{\mathrm{geometric}} = \frac{N^2}{4\pi\sigma_x\sigma_y} n_{\mathrm{b}} f_{\mathrm{r}}. \tag{4}$$

The total luminosity is based on a full simulation of all beam–beam effects. The ratio of the two values is H_{D} and one can easily see that it is within the range of 1.5–2 as mentioned above.

The angular distribution of the spent beam is shown in Fig. 4 for the case of CLIC. In the detector design one has to take care to avoid that any of the beam particles can be lost inside the detector. However, the maximum angle of the particles is quite small; typically it is less than 1 mrad.

3.2 Beamstrahlung

When a particle is forced on a curved trajectory by the other beam, it will emit radiation in a similar fashion as in a bending magnet. This radiation is called beamstrahlung. For typical parameters each particle emits one to a few photons each of which carries some percentage of the particle's energy. It is therefore important to be aware of the stochastic nature of the beamstrahlung. Due to beamstrahlung the particles lose energy during the collision and can therefore collide with less than the initial energy. This

leads to the formation of a luminosity spectrum. This obviously impacts the performance of the physics experiments. How important the effect is depends on the physics analysis that is being carried out.

The beamstrahlung can be described by its critical energy $\hbar\omega_\mathrm{c}$, which can be calculated as

$$\hbar\omega_\mathrm{c} = \frac{3}{2}\frac{\hbar\gamma^3 c}{\rho}. \tag{5}$$

Here, ρ is the bending radius of the particle trajectory. Conventionally one uses the so-called beamstrahlung parameter Υ, which is defined as

$$\Upsilon = \frac{2}{3}\frac{\hbar\omega_\mathrm{c}}{E}. \tag{6}$$

The beamstrahlung spectrum is described by the Sokolov–Ternov spectrum

$$\frac{\mathrm{d}\,\dot{w}}{\mathrm{d}\,\omega} = \frac{\alpha}{\sqrt{3}\pi\gamma^2}\left[\int_x^\infty \mathrm{K}_{5/3}(x')\mathrm{d}x' + \frac{\hbar\omega}{E}\frac{\hbar\omega}{E-\hbar\omega}\mathrm{K}_{2/3}(x)\right]. \tag{7}$$

Here, $x = \frac{\omega}{\omega_\mathrm{c}}\frac{E}{E-\hbar\omega}$ and $\mathrm{K}_{5/3}$ and $\mathrm{K}_{2/3}$ are the modified Bessel functions. In the limit $\Upsilon \ll 1$ the power of the photon radiation of a particle is proportional to Υ^2:

$$P = \frac{e^2}{6\pi\epsilon_0}\frac{c}{\rho^2}\gamma^4 = \frac{2}{3}\frac{r_\mathrm{e}c}{\lambda_\mathrm{c}^2}mc^2\Upsilon^2 \tag{8}$$

with $\lambda_\mathrm{c} = \hbar/(mc)$. The radiation spectrum of particles at different values of Υ is shown in Fig. 5. The average beamstrahlung parameter can be estimated as

$$\langle\Upsilon\rangle = \frac{5}{6}\frac{Nr_\mathrm{e}}{\alpha\sigma_z(\sigma_x+\sigma_y)}. \tag{9}$$

Here, α is the fine structure constant. Conventionally one often omits the brackets. The maximum beamstrahlung parameter is about

$$\Upsilon_\mathrm{max} \approx \frac{12}{5}\Upsilon. \tag{10}$$

For $\Upsilon \ll 1$, the spectrum corresponds to synchrotron radiation and one speaks of the classical regime. For $\Upsilon \gg 1$, the radiation is partially suppressed since the critical energy is above the beam energy. This is the so-called quantum regime. Only CLIC at 3 TeV will operate in this regime.

As one can see in Table 1, all projects have a value of n_γ in the range of 1.5–2. The typical angular distribution of the photons is small, similar to the one of the beam particles after collision; see Fig. 6 for an example. Hence, the beamstrahlung does not generate direct background in the detector.

4 Choice of beam parameters

In this section, the choice of beam parameters at the collision point is described based on the beam–beam physics introduced above. We will focus on the case of classical beamstrahlung. For simplicity it is assumed that the bunch charge and length as well as the transverse emittances are already determined, e.g. by the main linac and other systems of the collider. This leaves the choice of beta functions and the longitudinal position of the waist. In reality the choice of parameters is more complex.

4.1 Choice of flat beams

In the classical regime, the number of photons emitted per beam particle n_γ depends on the bunch charge and transverse dimensions:

$$n_\gamma \propto \Upsilon\frac{\sigma_z}{\gamma} \propto \frac{N}{\sigma_x+\sigma_y}. \tag{11}$$

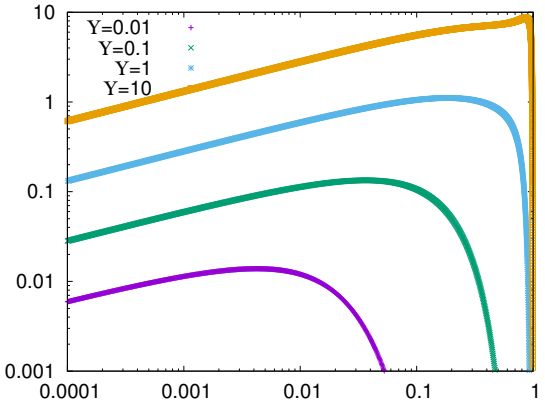


Fig. 5: The beamstrahlung power spectrum

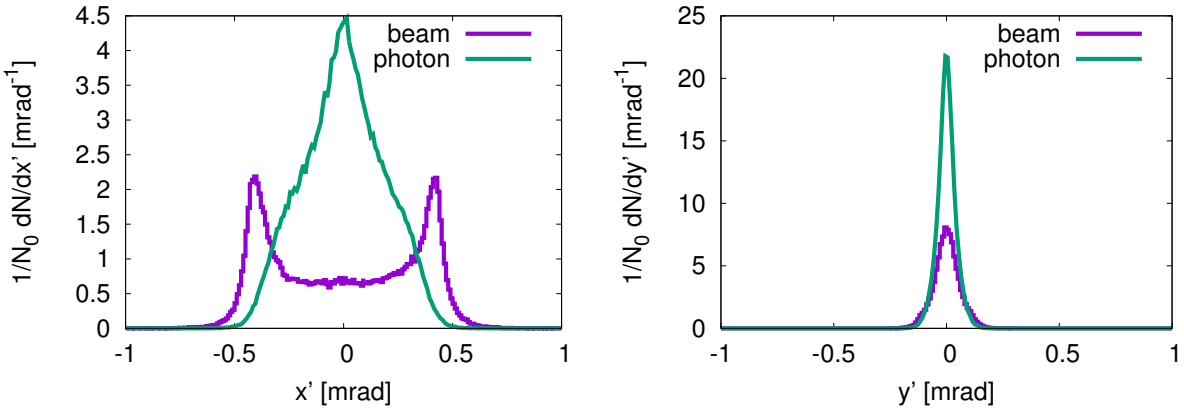


Fig. 6: The angular distribution of the spent beam including beamstrahlung

Similarly, the average energy E_γ of each photon is proportional to

$$E_\gamma \propto \Upsilon \frac{1}{\gamma} \propto \frac{N}{(\sigma_x + \sigma_y)\sigma_z}. \tag{12}$$

Typically, the emission of a single photon reduces the energy of the particle enough to not produce luminosity in the interesting energy range. Hence, the relevant parameter is the number of photons emitted rather than their energy.

For a given N, in order to reduce the beamstrahlung one thus aims to increase the sum of the transverse beam sizes $\sigma_x + \sigma_y$. At the same time the luminosity is proportional to $1/(\sigma_x\sigma_y)$, so one aims to minimize the product of the two beam sizes. Both goals can be simultaneously achieved by using a flat beam $\sigma_x \gg \sigma_y$; the horizontal beam size is chosen to be larger than the vertical, since the damping rings naturally deliver a horizontal emittance that is larger than the vertical. For $\sigma_x \gg \sigma_y$, we can approximate $\sigma_x \approx \sigma_x + \sigma_y$. Hence, the term N/σ_x in Eq. (3) is proportional to the number of beamstrahlung photons.

4.2 Luminosity spectrum and choice of horizontal beam size

A lower limit of horizontal beam size and beta function arises from the beamstrahlung to limit the degradation of the luminosity spectrum. It is important to note that the luminosity spectrum is also affected by another process. If particles collide, they can emit a photon just before the collision as a part of the physics process; this is called initial state radiation. This emission is a radiative correction to the physics process. In contrast to beamstrahlung, it therefore happens only to colliding particles that undergo some

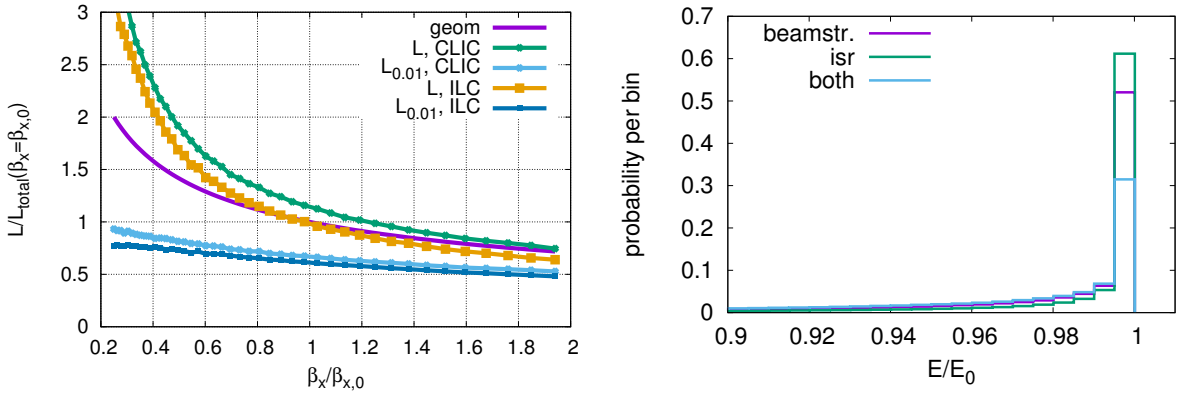


Fig. 7: Left: the total luminosity $\mathcal{L}$ and the luminosity above 99% of the nominal energy $\mathcal{L}_{0.01}$ as a function of the horizontal beta function. The beta function is normalized to the chosen value $\beta_{x,0}$. The curve labelled 'geom' neglects the beam–beam forces; in this case no beamstrahlung is emitted and the two luminosities are identical. Right: the luminosity spectrum in CLIC at 380 GeV. For comparison the spectra are also given taking into account only initial state radiation (isr) and only beamstrahlung.

physics process. However, it will degrade the luminosity spectrum in a similar fashion to beamstrahlung; the typical centre-of-mass energy spectrum of the colliding electrons and positrons is shown on the right-hand side of Fig. 7. Usually the experiments require that the degradation of the luminosity spectrum due to beamstrahlung is similar to the degradation due to initial state radiation. As a measure, one uses the ratio of the luminosity $L_{0.01}$, i.e. the part above 99% of the nominal centre-of-mass energy, and the total luminosity. In case of CLIC at 380 GeV a ratio of 60% has been targeted.

The total and peak luminosities are shown in Fig. 7 for CLIC and ILC as a function of the horizontal beta function. One can observe that the total luminosity increases strongly for smaller beta functions. It increases even faster than the geometric luminosity. This is a result of the fact that a smaller horizontal beam size increases the disruption and therefore leads to an increase in the pinch enhancement factor H_{D}. However, the peak luminosity only increases slightly for smaller beam sizes. Hence, the ratio of peak to total luminosity decreases rapidly for small beta functions, which yields a lower limit. It should be noted that additional lower limits for the horizontal beta function exist, e.g. from the ability to design the beam delivery system. With the chosen value, CLIC indeed reaches a ratio of 60%, as one can see on the right-hand side of Fig. 7.

In case of CLIC at 3 TeV, the requirement on the spectrum quality is somewhat relaxed (30%), since the tail of the luminosity spectrum also contributes to the creation of interesting physics events. An important example is the double-Higgs production, which allows us to measure the Higgs self-coupling. Basically the whole luminosity spectrum contributes to this production process at high energies, which increases the importance of the total luminosity with respect to the peak luminosity.

4.3 Choice of vertical beta function

The beam size at the collision point decreases as the beta function decreases proportionally to $\sqrt{\beta_y}$. However, the beam size just around the collision point increases faster as the beta function decreases. At any point s around the collision point $s = 0$, the beam size is proportional to

$$\sigma_y(s) \propto \sqrt{\beta(0) + \frac{s^2}{\beta(0)}}. \tag{13}$$

This hourglass effect limits the luminosity that one can gain by reducing the beta function; it is illustrated in Fig. 8. In the figure the luminosity is also shown as a function of the vertical beta function for the case

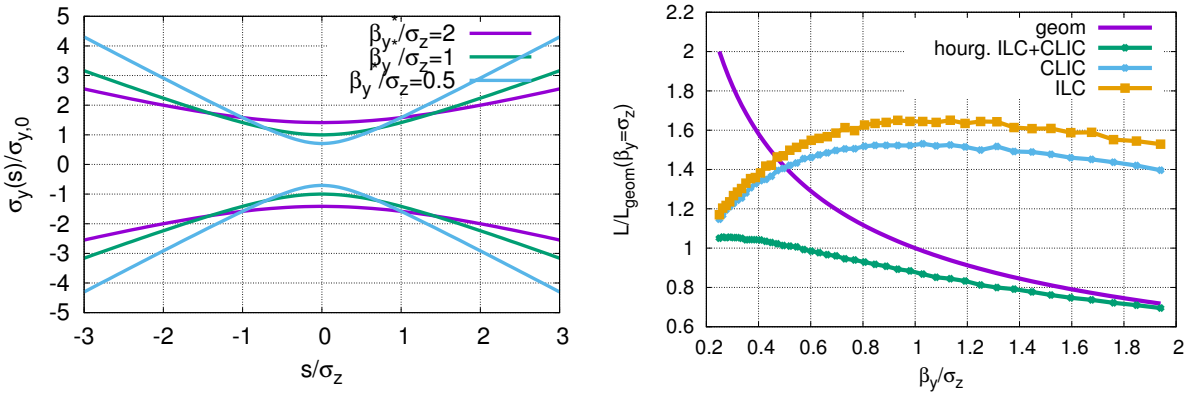


Fig. 8: Left: the beam size around the collision point for different beta functions. Right: the dependence of the luminosity on the vertical beta function. The curve labelled 'geom' indicates the dependence neglecting the hourglass effect.

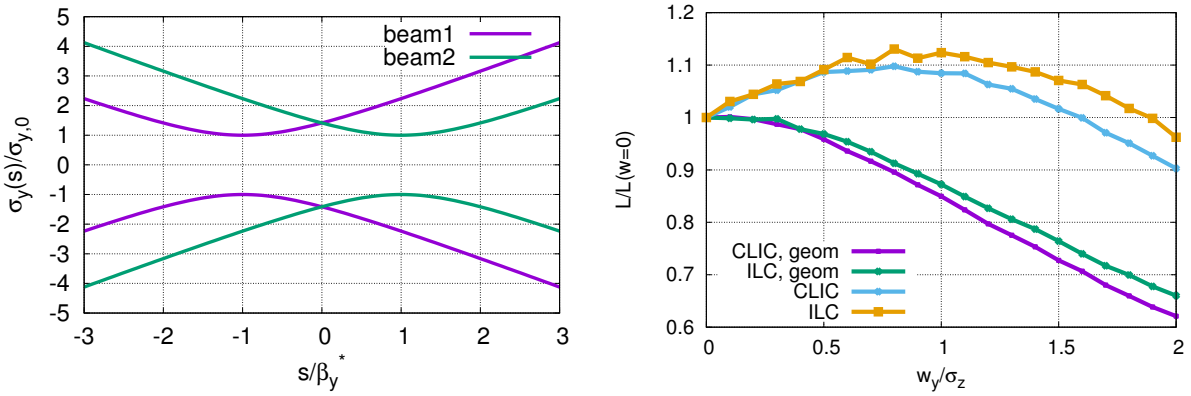


Fig. 9: Left: the beam size around the collision point for a waist before the collision point. Beam 1 comes from the left, beam 2 from the right. Right: the dependence of the luminosity on the vertical waist position. For the curve labelled 'geom' the beam–beam forces are neglected.

without beam–beam forces. As can be seen, the optimum beta function is about a quarter of the bunch length and the optimum is quite flat.

The dependence of the luminosity on the beta function is strongly modified by the beam–beam forces; see Fig. 8. For larger beta functions the pinch effect is more efficient in increasing the luminosity, moving the optimum choice to roughly $\beta_y \approx \sigma_z$. Obviously, the value depends on the strength of the beam–beam forces. For both projects a vertical beta function somewhat larger than the optimum has been chosen, $\beta_y = 1.6\ \sigma_z$ in ILC and $\beta_y = 1.4\ \sigma_z$ in CLIC. These choices lead to a very minor loss of luminosity. The larger beta function helps to ease the design of the beam delivery system.

4.4 Waist shift

If the beam–beam forces are weak the maximum luminosity is obtained if the waists of the two beams are placed at the collision point. A distance W_y of the waists to the collision point leads to a reduced luminosity. Figure 9 shows the relative luminosity as a function of the waist position for ILC and CLIC if beam–beam forces are neglected. The dependence on W_y/σ_z differs slightly, since the ratio β_y/σ_z differs. If the beam–beam forces are taken into account the situation changes significantly. The maximum luminosity is obtained if the waists are positioned before the collision point, $W_y > 0$. For the optimum

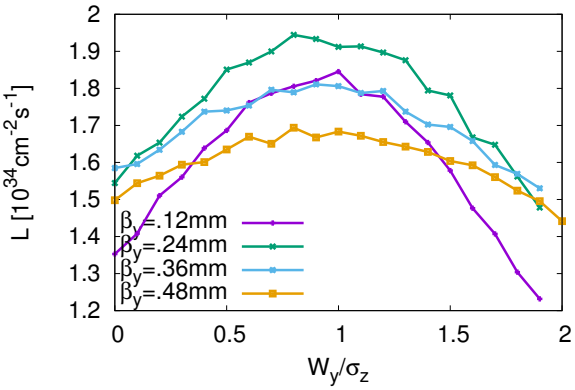


Fig. 10: The luminosity in ILC as a function of the waist shift for different beta functions

$W_y \approx \sigma_z$, the luminosity is increased by 10% instead of the expected reduction of 10%.

In principle a full optimization of the vertical beta function and waist position can be performed. Figure 10 shows the results for ILC. The highest luminosity would be achieved using $\beta_y \approx 0.24$ mm. It is interesting to note that using $\beta_y \approx 0.48$ mm yields more luminosity than $\beta_y \approx 0.12$ mm if the waists are at the collision point. But if the waist position is adjusted for maximum luminosity, the smaller beta function obtains more luminosity.

4.5 Note on crossing angle

The beams in a linear collider cross with an angle in the horizontal plane, $\theta_\mathrm{c} = 14$ mrad in ILC and $\theta_\mathrm{c} = 20$ mrad in CLIC. This angle would reduce the luminosity to

$$\mathcal{L} = H_\mathrm{D} \frac{N^2}{4\pi\sigma_x\sigma_y} n_\mathrm{b} f_\mathrm{r} \frac{1}{\sqrt{1+\left(\frac{\sigma_z}{\sigma_x}\tan\frac{\theta_\mathrm{c}}{c}\right)^2}}. \tag{14}$$

The reduction is about an order of magnitude. Therefore, crab cavities are used to rotate the bunches such that they collide head-on, even if their trajectories cross at an angle. One can thus ignore the crossing angle and assume that the collisions are head-on.

5 Imperfect collisions

The strong beam–beam forces change the dependence of the luminosity on the beam–beam offset. They can also lead to strong luminosity loss due to correlations in the longitudinal and transverse profiles.

5.1 Beam–beam offsets

If the beams are separated in the collision point, the luminosity is reduced. For rigid beams and neglecting the hourglass effect, the luminosity ratio can be described as a function of the offset Δy as

$$\frac{\mathcal{L}}{\mathcal{L}_0} = \exp\left(-\frac{\Delta y^2}{4\sigma_y^2}\right). \tag{15}$$

The beam–beam forces strongly modify this behaviour. Figure 11 shows the luminosity ratio as a function of offset for different disruption parameters. For weak disruption the luminosity decreases more slowly with offset than for rigid beams. For larger disruption very small offsets can lead to a large loss of luminosity. This is due to the fact that the collision becomes unstable, the kink instability, which is

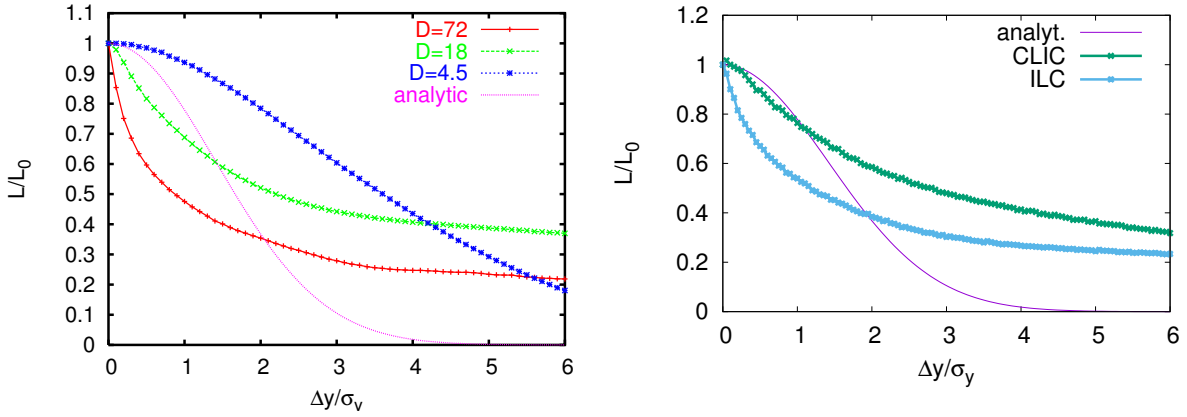


Fig. 11: Left: the luminosity ratio as a function of the offset for different disruption parameters. Right: the luminosity ratio as a function of vertical beam offset for ILC and CLIC.

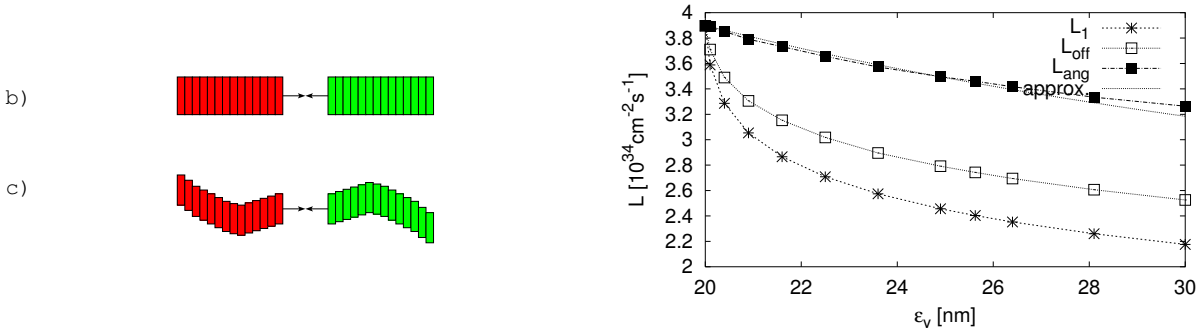


Fig. 12: Left: sketch of the emittance growth in the main linac due to wakefields and dispersion and the possible translations into beam shapes at the interaction point. Right: luminosity as a function of the emittance in the case of TESLA.

a typical two-stream instability. For large offsets, the beam–beam forces maintain more luminosity than for rigid beams. This is because the beams attract each other.

In order to avoid luminosity reduction a control of the beam–beam offset is required at the level of a fraction of a nanometre. The motion of the ground and vibrations of technical components make this a challenging task. Two main methods are used to address this challenge. First, in case of CLIC, the beam-guiding magnets are stabilized with active feedback systems that sense the motion of the magnet and correct it using movers. Second, in both ILC and CLIC, the beam–beam offset is measured and corrected with a beam-based feedback system. In case of ILC this feedback system can correct from one bunch to the next within the pulse; in CLIC it has a latency of a few bunch crossings and mainly acts from one pulse to the next. This feedback can easily detect an offset even of a fraction of a nanometre, since the resulting deflection of the beams is in the order of tens of μrad. A few metres downstream of the collision point such an angle has translated into an offset that can be easily measured with a beam position monitor.

5.2 The banana effect

Up to now we considered that the beam particles are independently distributed in the different dimensions and that there are no correlations. In practice one has to take correlations into account, i.e. to model the beam including the effects before the interaction point. This is particularly important if the vertical disruption is larger than about 15.

A good example of the importance of the integration of different subsystems is the so-called banana effect [11]. This effect has initially been studied for an older linear collider design, TESLA, but is also important for ILC. In early studies the emittance growth in the main linac was studied independently from the beam–beam simulations. However, wakefield effects in the main linac introduced correlated offsets (i.e. the mean vertical particle position depends on the longitudinal position within the bunch $y(z)$), see c) on the left-hand side of Fig. 12. The projected emittance has then been used as input for the beam–beam simulations as in b). Instead simulations with full correlation should have been used as in c).

Systematic studies showed that the luminosity drops much faster with increasing beam emittance if the correlations are used in simulations than anticipated from the projected emittances [13]. This is shown on the right-hand side of Fig. 12. The horizontal scale indicates the emittance at the interaction point. The emittance consists of an uncorrelated part of $20\,\mathrm{nm}$ and an additional contribution from the wakefields in the main linac. The luminosity is shown on the vertical axis. The approximation assumes that the luminosity scales with $1/\sigma_y$ and that the correlations are not relevant. The curve L_1 shows the luminosity if both beams are centred in position and angle. Evidently the luminosity is very strongly affected by the emittance growth due to the wakefields. This is because the collision is unstable as a result of the high disruption.

The luminosity can be recovered if a full luminosity optimization is performed at the interaction point by varying the beam–beam offset (L_{off}) and by varying offset and angle (L_{ang}). This procedure however requires that the luminosity is measured online. Hence, it takes much more time than a simple beam position monitor-based feedback. In ILC it is foreseen to perform such an optimization during each bunch train. For a smaller disruption, as in CLIC, the banana effect is negligible. Hence, a luminosity optimization scan during the train is not required.

6 Beam–beam background and its impact on the detector design

The beam–beam effects lead to the generation of background for the physics experiments. This includes the production of electrons, positrons, muons and hadrons. In this lecture we will only discuss the production of electron–positron pairs, which have an important impact on the detector design.

6.1 Coherent pair creation

If the fields of the beams are very strong they can rip up the beamstrahlung photons, forming an electron–positron pair. This is called coherent pair creation, since it is caused by the beamstrahlung photons interacting with the coherent field of the beams, not by interacting with individual photons [14]. For $\Upsilon \ll 1$, almost no coherent pair creation exists. For larger values it becomes substantial. Both GUINEA-PIG and CAIN allow us to simulate the effect.

For the lower energy linear colliders coherent pair creation does not play an important role. For CLIC at $3\ \mathrm{TeV}$ the creation of pairs from beamstrahlung is significant. About 6.8×10^8 pairs are produced per bunch crossing. The total charge of the pairs is hence about 20% of the beam charge. The spectrum of the pairs is shown in Fig. 13; the average particle energy is about $300\ \mathrm{GeV}$. The pair particles are produced at very small angles. However, a produced electron can move in the direction of the electron beam or in the direction of the positron beam. In the first case it is focused by the oncoming positron beam and in the second case it is defocused by the oncoming electron beam. It can then be deflected to larger angles. The equivalent is true for the produced positrons. It is therefore important that the detector provides a large enough acceptance in the outgoing beamline to avoid excessive losses from coherent pairs. In Fig. 14, the total powers above a certain angle are shown for the spent beam, the beamstrahlung photons and the coherent pairs. As can be seen, an exit aperture of about $10\ \mathrm{mrad}$ is well sufficient to avoid losses of more than 1 W in the detector. It should be noted that $1\ \mathrm{W}$ corresponds to a total energy of $400\ \mathrm{TeV}$ per bunch crossing or roughly the energy of 300 beam particles.

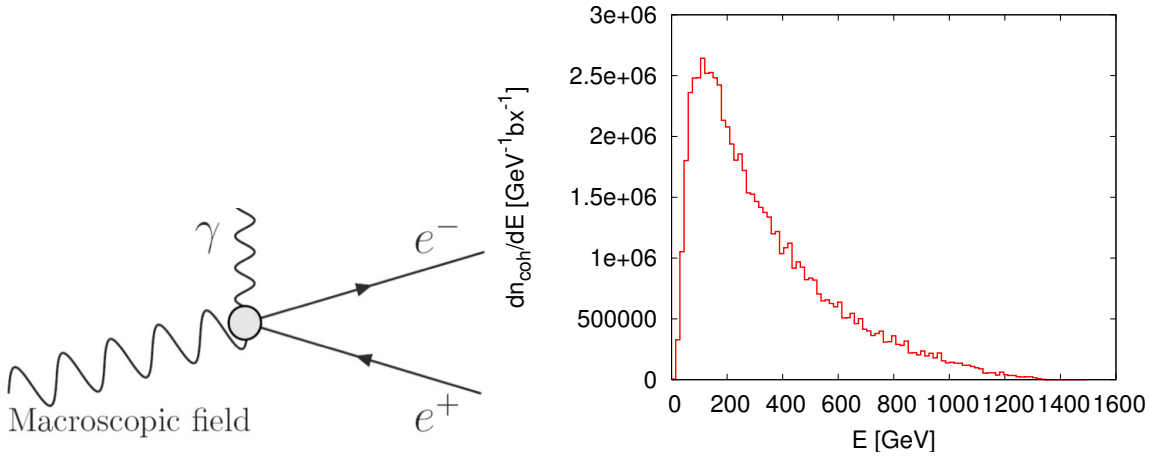


Fig. 13: Left: the Feynman diagram for the coherent pair creation. Right: the energy spectrum of the coherent pairs in CLIC at 3 TeV.

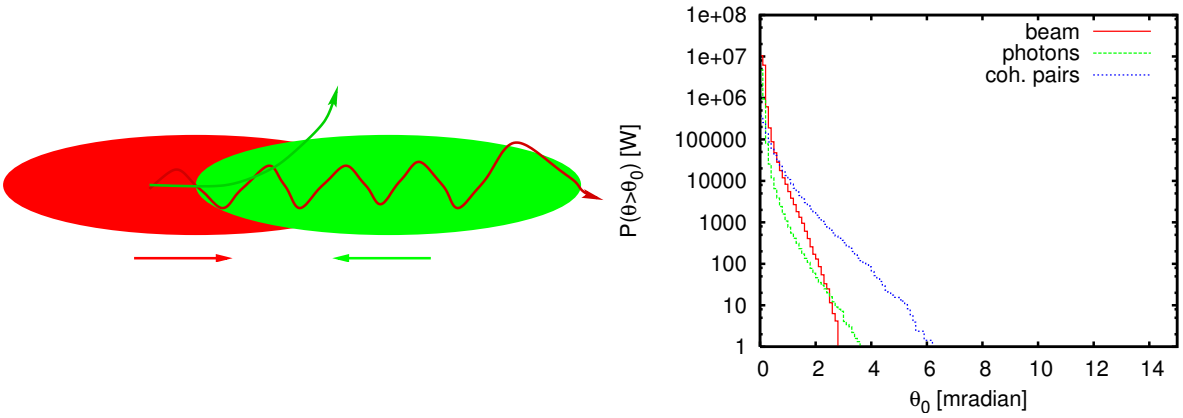


Fig. 14: Left: the particle from coherent pair creation can be focused or defocused by the beams. Right: the total power in the spent beam above a given angle in the case of CLIC at 3 TeV.

It should be noted that not only the real photons from beamstrahlung can be turned into pairs. If the field is strong enough, even the virtual photons that are co-moving with the beam particles can be turned into pairs. This is called the trident cascade process, because a single electron or positron generates an additional pair in the external field. For $\Upsilon \gg 1$, the trident cascade becomes a large source of pairs. This is particularly problematic because the beam particles can lose a large part of their energy in this process, degrading the luminosity spectrum. In the present linear collider designs the trident cascade process does not play an important role. At higher energies or with shorter bunches it would become substantial. GUINEA-PIG++ allows us to simulate the trident cascade process.

6.2 Incoherent pair creation

Colliding photons can turn into an electron–positron pair. In linear colliders two main sources of these photons exist. One is the above-mentioned beamstrahlung. The other are virtual photons that accompany each electron and positron. As a consequence three production mechanisms of pairs exist: the Breit–Wheeler process (two real photons form a pair), the Bethe–Heitler process (one real and one virtual photon form a pair) and the Landau–Lifshitz process (two virtual photons form a pair). The most fundamental Feynman graphs are shown in Fig. 15. It is interesting to note that the actual calculation of the pair production is complicated by the so-called beam-size effect. That is, that the size of the

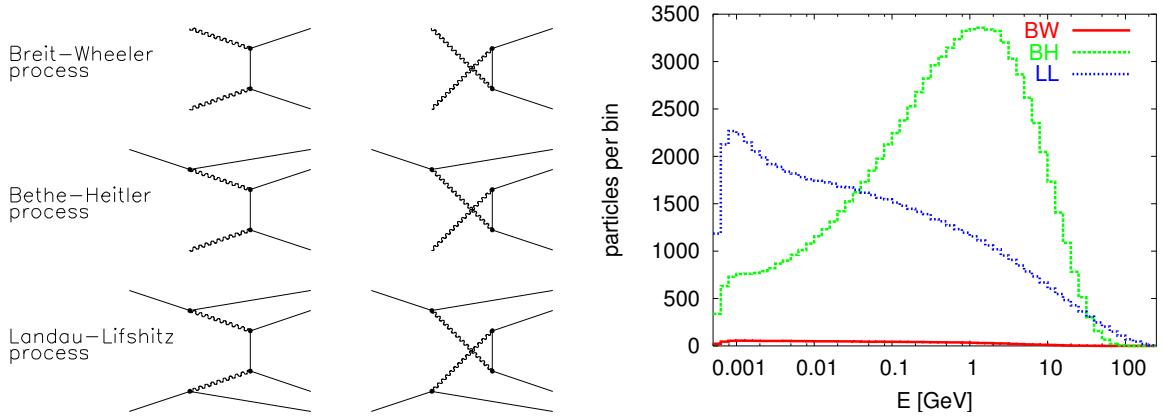


Fig. 15: Left: the Feynman diagrams for the incoherent pair creation. Right: the energy spectrum of the incoherent pairs for an older CLIC design at 500 GeV.

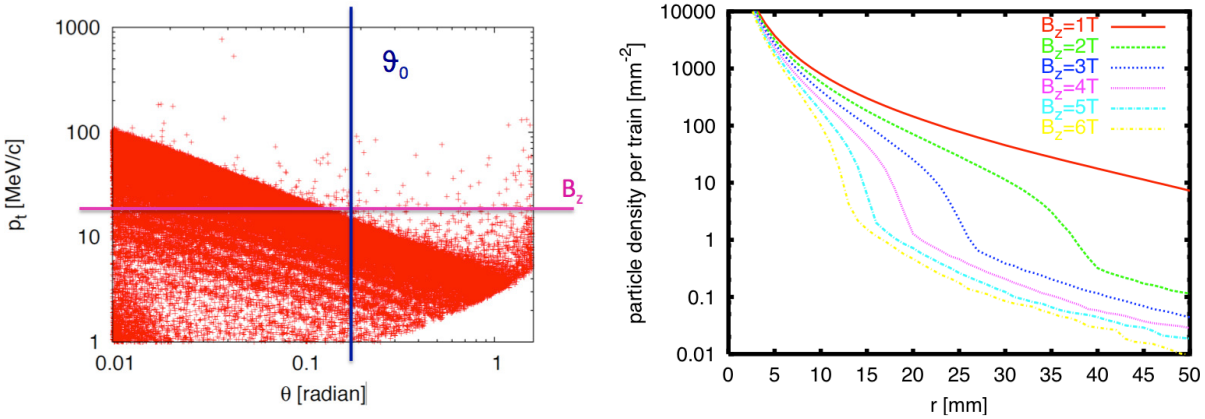


Fig. 16: Left: the angle and transverse momentum of incoherent particles after the beam collision. Right: the number of hits in the vertex detector as a function of the detector radius and for different detector main solenoid fields.

beam modifies the cross-section. In addition the strong fields of the beams also modify the cross-section. Conventional programs that simulate pair production do not take these effects into account. However, GUINEA-PIG and CAIN can model these effects. Typically the number of pairs is $O(10^5)$ per bunch crossing; see Table 1. An example spectrum can be seen in Fig. 15.

Some of the incoherent pair particles are produced at larger angles, but most are produced at small angles. The pairs are deflected in the same fashion as the coherent pairs. However, the deflection is more important because the energy of the incoherent pairs is typically smaller. Pair particles can therefore be deflected to relatively large angles. The maximum deflection is given by the particle energy. An example of the final transverse momentum and angle of the pair particles is shown in Fig. 16. One can see the few particles that have been produced at large angles. The sharp edge in the plot corresponds to the maximum deflection that a pair particle of a given energy can obtain from the beams.

The most important effect of the incoherent pairs is that the particles can hit the vertex detector that is as close to the beam as reasonably possible (typically a couple of centimetres). This detector identifies the origin of particles in the detector. It can determine whether a track originates directly from the colliding beams or if it starts at some distance from the beams because it is produced by a decay product of a particle.

A significant number of background electrons and positrons hitting this detector will compromise its performance. The impact of these particles on the detector is suppressed by two measures. First, the detector solenoid field forces particles on a helical trajectory. The radius of the helix is given by the transverse momentum and hence particles with small $p_\perp$ will not hit the detector. Second, particles at small angles will also not hit the detector. The number of hits in a vertex detector of fixed angular coverage is shown in Fig. 16 for different detector solenoid fields. As the radius of the vertex detector increases, the number of hits decreases and they are distributed over a larger area. The relatively steep edge corresponds to the maximum deflection of the pairs by the beams. One typically chooses a radius sufficiently larger than the edge to guarantee good detector performance. Here, the detector design is a direct consequence of the beam–beam effect and the corresponding background.

7 Conclusion

Beam–beam effects are important drivers for linear collider designs. The beamstrahlung is one of the main limitations for the luminosity and affects the capabilities of the experiments. Background from the beam–beam interaction also impacts the detector design, in particular the choice of beam-pipe radius. It also affects the low-angle detector design.

References

[1] N. Phinney, SLC final performance and lessons, eConf C **00082** (2000) MO102 [physics/0010008 [physics.acc-ph]].

[2] T. Behnke *et al.*, The international linear collider technical design report – Vol. 1: Executive summary, arXiv:1306.6327 [physics.acc-ph] (2013).

[3] H. Baer *et al.*, The international linear collider technical design report – Vol. 2: Physics, arXiv:1306.6352 [hep-ph] (2013).

[4] C. Adolphsen *et al.*, The international linear collider technical design report – Vol. 3.I: Accelerator R&D in the technical design phase, arXiv:1306.6353 [physics.acc-ph] (2013).

[5] C. Adolphsen *et al.*, The international linear collider technical design report – Vol. 3.II: Accelerator baseline design, arXiv:1306.6328 [physics.acc-ph] (2013).

[6] T. Behnke *et al.*, The international linear collider technical design report – Vol. 4: Detectors, arXiv:1306.6329 [physics.ins-det] (2013).

[7] M. Aicheler *et al.*, A multi-TeV linear collider based on CLIC technology: CLIC conceptual design report (2012). http://dx.doi.org/10.5170/CERN-2012-007

[8] L. Linssen, A. Miyamoto, M. Stanitzki and H. Weerts, Physics and detectors at CLIC: CLIC conceptual design report, arXiv:1202.5940 [physics.ins-det] (2012). http://dx.doi.org/10.5170/CERN-2012-003

[9] P. Chen, G. Horton-Smith, T. Ohgaki, A.W. Weidemann and K. Yokoya, CAIN: Conglomerat d'ABEL et d'interactions nonlineaires. SLAC-PUB-6583 (1995).

[10] D. Schulte, Electronmagnetic and Hadronic Background in the Interaction Region of the TESLA Collider. DESY-TESLA-97-08 (1997).

[11] R. Brinkmann, O. Napoly and D. Schulte, Beam–beam instabilities driven by wakefield effects in the main linac, PAC-2001-TPPH153 (2001).

[12] D. Schulte, Luminosity in future linear collider in the presence of static wakefield effects in the main linac, CLIC-Note-544 (2002).

[13] D. Schulte, An update on the banana effect, Nanobeam 2002 and CERN-AB-2003-009 (2003).

[14] P. Chen and V. Telnov, *Phys. Rev. Lett.* **63** (1990) 1796. http://dx.doi.org/10.1103/physrevlett.63.1796

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Machine Protection

R. Schmidt
CERN, Geneva, Switzerland

Abstract

The protection of accelerator equipment is as old as accelerator technology and was for many years related to high-power equipment. Examples are the protection of powered equipment from overheating (magnets, power converters and high-current cables), of superconducting magnets from damage after a quench and of klystrons. The protection of equipment from beam accidents is more recent. It is related to the increasing beam power of high-power proton accelerators (e.g. ISIS at Rutherford Appleton Laboratory, the Spallation Neutron Source (SNS) at Oak Ridge, the European Spallation Source at Lund (ESS) and the cyclotron at the Paul Scherrer Institute in Switzerland), to the emission of synchrotron light by electron–positron accelerators and free-electron lasers and to the increase of energy stored in the beam (in particular for hadron colliders such as the Large Hadron Collider at CERN, Switzerland). Designing a machine protection system requires an excellent understanding of accelerator physics and operation to anticipate possible failures that could lead to damage. Machine protection includes beam and equipment monitoring, a system to safely stop beam operation (e.g. dumping the beam or stopping the beam at low energy) and an interlock system providing the glue between these systems. The most recent accelerator, the LHC, will operate with about 3×10^{14} protons per beam, corresponding to an energy stored in each beam of 360 MJ. This energy can cause massive damage to accelerator equipment in case of uncontrolled beam loss, and a single accident damaging vital parts of the accelerator could interrupt operation for years. This article provides an overview of the requirements for protection of accelerator equipment and introduces the various protection systems. Examples are mainly from LHC and ESS.

Keywords

Machine protection; interlock system; high-power accelerator; beam loss; accident.

1 Introduction

In general, risks come from energy stored in an accelerator (measured in joules) and power when operating an accelerator (measured in watts). When we are talking about a very powerful accelerator, we need to consider that the power flow needs to be controlled.

Particle accelerators use large amounts of power; large accelerators operate with a few megawatts to many megawatts. The question to be addressed is where does the power go in case of failure? An uncontrolled release of energy or power flow can lead to unwanted consequences, such as damage of equipment and loss of time for operation. There is also the risk of activation of equipment when operating with particle beams in case of beam losses. Handling large amounts of power/energy is an issue in many domains, but a particular challenge for complex systems such as accelerators.

https://doi.org/10.23730/CYRSP-2017-003.447

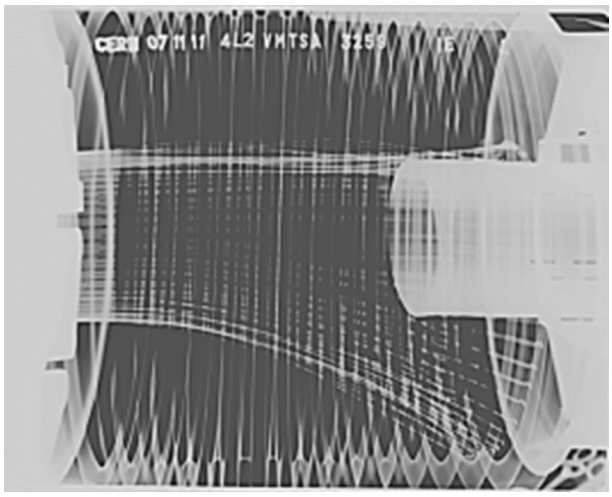


Fig. 1: RF fingers in LHC interconnects; some of them are bent. The metallic coating on the fingers had melted: the temperature had reached more than 800°C [1].

For an accelerator, there are several sources of energy or power that need to be considered.

- The energy stored in the particle beam, e.g. in a synchrotron or storage ring.
- The beam power of a particle beam or in a linac, cyclotron or synchrotron. The beam power is given by the number of particles multiplied with their energy, per unit time.
- The energy stored in superconducting magnets.
- The power required to operate accelerator systems, e.g. magnets, power converters and RF systems.

2 Energy transfer from beam to equipment

There are several mechanisms for energy transfer from the beam to the equipment, one of them by losing particles into the equipment. There will always be particle losses during operation, due to collisions with the residual gas or with a counter-rotating beam, or particle losses at the aperture (e.g. due to emittance growth and other effects). Accidental particle losses can occur due to a large number of possible failure mechanisms.

Another mechanism is energy deposited by electromagnetic interaction between beam and environment (vacuum chamber, RF, beam instrumentation, kicker magnets etc). The deposited power depends on the beam intensity and bunch structure, as well on the impedance seen by the beam. Problems have been frequently observed at accelerators. Figure 1 shows the damage of RF fingers in interconnects at LHC. Due to an incorrect installation of RF fingers, the beam deposited energy into the fingers and they were bent [1].

Energy can also be deposited by the synchrotron radiation emitted by particle beams, in particular in electron or positron accelerators. The power increases with the particle energy to the power of four and can be very high, up to several tens of megawatts. The radiation can be very focused; in particular, radiation from wiggler and undulator magnets can increase the power by orders of magnitude (e.g. for free-electron lasers (FELs)). Normally, this is considered in the design of the accelerator and experiments; however, there are a number of failure scenarios that can lead to accidents (e.g. operating with too high current).

Machine protection is essentially to prevent consequences of accidental beam losses, but other mechanisms for damage should also be addressed.

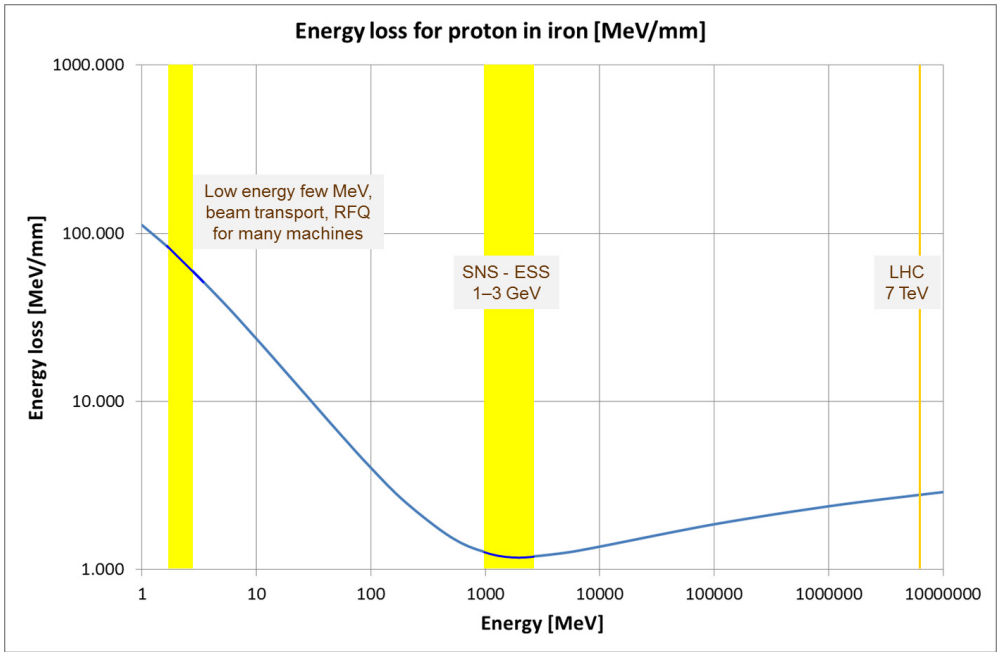


Fig. 2: Energy loss for protons when entering into an iron block, calculated by the Bethe–Bloch equation. Effects of the hadron shower are not included.

2.1 Energy deposition from high-energy particles

Charged particles moving through matter interact with electrons of atoms in the material, exciting or ionizing the atoms. Energy deposition starts when a charged particle enters the material. If the particle energy is high enough, it leads to particle cascades in the material, increasing the deposited energy along the length. The maximum energy deposition can be deep in the material at the maximum of the hadron or electromagnetic shower depending on the particle momentum and material.

To get an idea of the damage potential, the energy loss by ionization at the surface of the target can be estimated using the Bethe–Bloch equation, ignoring particle cascades (see Fig. 2). It is interesting to observe that the energy loss for low-energy particles is very high. For a proton at 7 TeV/c the energy loss at the entrance is much lower, but deeper in the material the energy deposition is dominated by hadron showers with the maximum energy deposition deep in the material.

There is no straightforward expression for the energy deposition of high-energy particles, since this depends on the particle type, momentum, beam parameters and material parameters (atomic number, density and specific heat). Programs such as FLUKA [2], MARS [3] or GEANT4 [4] are being used for the calculation of energy deposition (and subsequent temperature increase) as well for the activation of the material that is exposed.

The following example illustrates the calculation of the beam particles entering a material block or traversing a window.

i) A proton beam travels through a thin window of thickness d.
ii) Assume a beam area of 4 $\sigma_x \times \sigma_y$, with σ_x and σ_y the root-mean-square (r.m.s.) beam sizes assuming Gaussian beams.
iii) Assume a homogeneous beam distribution.

The energy deposition can be calculated; mass and specific heat are known. The temperature can be calculated (rather good approximation), assuming a fast loss and no cooling. With low-energy protons (3 MeV/c) and a beam size of $\sigma_x = \sigma_y = 1$ mm, the following parameters are calculated:

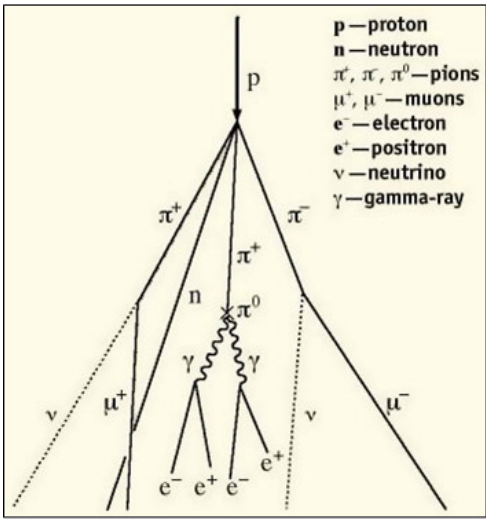


Fig. 3: Illustration of a hadronic shower

i) iron specific heat = 440 J/(kg $\times$ K);
ii) iron specific weight = 7860 kg/m^3;
iii) dE/dx = 56.7 MeV/mm;
iv) $N_{\mathrm{p}} = 1.16 \times 10^{12}$.

The temperature increase in this example for these parameters is 763 K. For high-energy protons the energy deposition by hadronic showers (see Fig. 3) dominates:

i) pions are created when the protons travel through matter;
ii) the decay of pions creates electromagnetic showers;
iii) there is an exponential increase in number of created particles;
iv) the final energy deposition is to a large extent due to the large number of electromagnetic particles;
v) the energy deposition scales roughly with total energy of incident particles;
vi) the maximum of the energy deposition can be deep in the material;
vii) energy deposition is a function of the particle type, its momentum and parameters of the material (atomic number, density and specific heat).

A simple approximation for the temperature increase in material for a 7 TeV/c proton beam impact is given in the following example: for copper, the maximum longitudinal energy deposition for a single 7 TeV/c proton at about 25 cm inside the material is $E_{\mathrm{dep}} = 1.5 \times 10^{-5}$ J/kg (calculation with FLUKA). The energy required to heat and melt copper is $E = 6.3 \times 10^5$ J/kg. Assuming a pencil beam, the number of particles required to damage (melt) copper is of the order of 10^{10}. For graphite, the number of particles needed to cause damage is about one order of magnitude larger.

2.2 Beam loss and consequences

The energy deposition leads to a temperature increase in the material that can be vaporized, melted, deformed or lose its mechanical properties, depending on the material and the beam impact.

Relevant parameters to be considered for heating and possible damage to material are the momentum of the particle, the particle type, the energy stored in the beam, the beam power, the beam size, the beam power/energy density (MJ/mm^2, MW/mm^2), the time structure of the beam and cooling conditions. In order to estimate the order of magnitude for possible damage:

i) 1 MJ can heat and melt about 1.5 kg of copper;
ii) 1 MJ corresponds to the energy stored in about 0.25 kg of TNT [5];
iii) 1 MW during 1 s corresponds to 1 MJ.

When particles interact with equipment, material can be activated with the subsequent risk for hand-on maintenance. It is considered to be acceptable if beam losses do not exceed, say, 1 W/m (assuming proton beams with high energy, say above some 100 MeV/c). Another principle is 'ALARA': exposure of personnel to radiation should be 'as low as reasonably achievable'. If a further reduction of the beam losses below 1 W/m is reasonably possible, this is recommended in order to minimize exposure of service personnel. Radioactive activation of material is mainly an issue for hadron accelerators; it is less problematic for electron–positron machines.

For accelerators with superconducting magnets there is a specific problem: even with beam loss much below the damage threshold, superconducting magnets can quench (beam loss of mJ to J). In case of a quench, beam operation is interrupted for some time (possibly up to many hours) leading to downtime. In order to avoid beam-induced quenches, beam losses are monitored and the beam is dumped if a predefined threshold is exceeded before a magnet quenches, reducing the downtime since the time to recover from a quench is avoided. The damage threshold is far above the quench threshold; this strategy also protects magnets from beam-induced damage.

Superconducting cavities' performance degradation is observed after beam losses of some 10 J. With higher beam losses, material can vaporize, melt, deform or lose its mechanical properties.

There is some risk of damage to sensitive equipment for an energy deposition of some 10 kJ (beam impact for a short time, say a maximum of a few milliseconds). Risk to damage sensitive equipment exists for less than 1 kJ; risk for damage of any structure for some MJ (depends on beam size).

More refined and complete calculations can be made to determine real-world scenarios on a case-by-case basis, where the distribution of the impacting particles and the details of the material and geometry are important. After the calculation of the temperature increase, the response of the material to beam impact needs to be addressed (deformation, melting etc). Mechanical codes such as ANSYS and hydrodynamic codes such as BIG2 and others can be used.

Beams at very low energy have limited power; however, the energy deposition is very high and can lead to (limited) damage in the case of a beam impact at the initial stage of an accelerator, after the source, in the low-energy beam transport and in the radio-frequency quadrupole (RFQ). This might lead to a long downtime, depending on availability of spares.

There is another risk from beam loss, radiation-induced effects in electronics (single-event effects) that could stop the operation of an accelerator.

3 Accelerators that require protection systems

Not all accelerators require protection systems; in this section we discuss when protection needs to be considered.

– Synchrotrons and storage rings for the acceleration of hadron beams with a large amount of energy stored in the beam. Examples are LHC, RHIC and in the past the TEVATRON and HERA. Other examples are synchrotrons accelerating beams for fixed target experiments or used as injectors for other machines. The energy stored in the beam for different accelerators as well as the energy stored in the LHC magnet system is shown in Fig. 4.
– High-power proton accelerators (e.g. spallation sources) with beam power of some 10 kW to above 1 MW. There is the risk of beam-induced damage and activation. The beam power of existing machines exceeds 1 MW (SNS, PSI cyclotron and in the future JPARC), but also for lower power

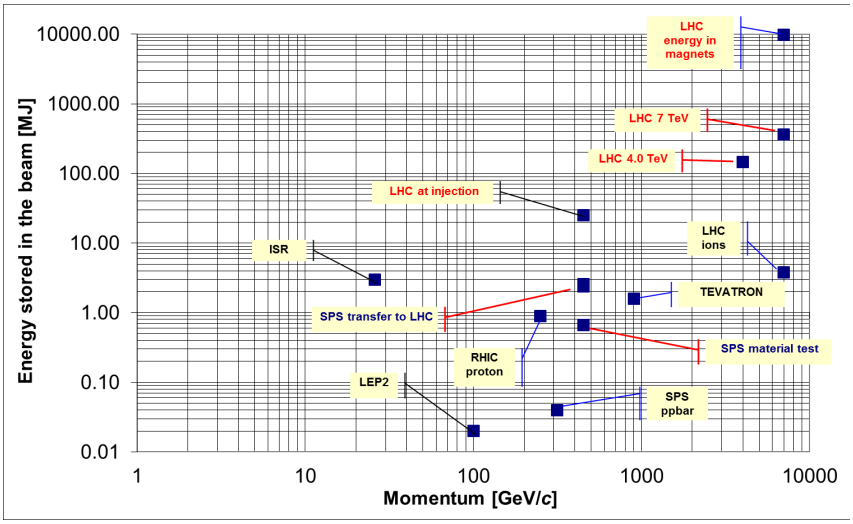


Fig. 4: Energy stored in the beams for different accelerators, and the energy stored in the LHC magnet system

machine protection needs to be considered (ISIS). Several high-power machines are under construction (ESS with a power of 5 MW, IFMIF and MYRRHA are in the project planning phase).

– Synchrotron light sources with high-intensity beams and secondary photon beams. Damage can come from the primary beam but also from the synchrotron radiation.
– Energy-recovery linacs, as an example the Daresbury ERL prototype: one bunch train cannot damage equipment, but in case of beam loss the next train must not leave the (injector) station.
– Linear colliders were expected to operate with very high beam power densities due to small beam size. One beam pulse can lead to damage. After a time interval large enough to allow a substantial change in the beam trajectory (fraction of a second), a pilot beam must be used to prove the integrity [6].
– High average power in linear accelerators: FLASH 90 kW, European XFEL 600 kW, JLab FEL 1.5 MW, ILC 11 MW.
– Medical accelerators: a too high dose to patients needs to be prevented. The techniques for protection are similar.

There is a large interest in the exploitation of high-power hadron accelerators. In spallation sources high-intensity proton beams are accelerated and directed to a target. The protons interact with the target material and spallation neutrons are produced. Other accelerators are using high-intensity proton beams for neutrino production. Rare-isotope beams are produced by accelerating ions (e.g. FRIB is a folded linac to accelerate ions). Accelerator-driven systems (ADSs) are being developed with several projects around the world. A very energetic particle beam is used to stimulate a reaction in a subcritical reactor, which in turn releases enough energy to power the particle accelerator and leaves an energy profit for power generation. Figure 5 shows beam current, beam power and particle momentum for different high-power proton accelerators.

There is a difference between accelerators operating with high-power beams and those with large stored energy. For hadron colliders, the energy stored in the beams can be very high, as has been shown for the LHC. In case of a failure, the energy stored in beam and magnets needs to be safely deposited.

For linear accelerators with high beam power, the beam must be stopped fast at the source in case of a failure causing beam losses. As an example, a continuous beam with a power of 5 MW could deposit the energy of 5 kJ in 1 ms. Even if the beam is stopped at the source, there are still particles present between the source and the location of the beam impact that might damage equipment.

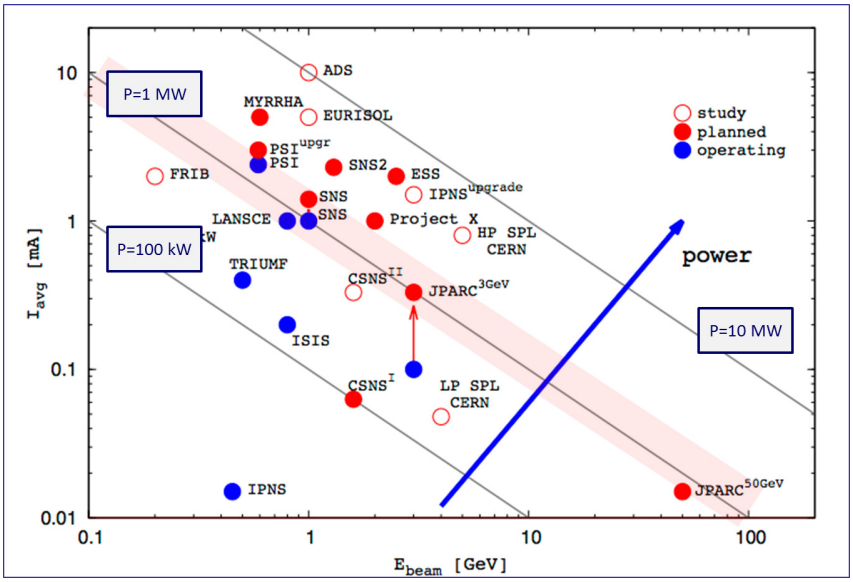


Fig. 5: Current versus particle momentum for high-power proton accelerators around the world

The beam power increases along the accelerating structure proportionally to the particle momentum, which needs to be considered in the estimation of possible damage.

4 Hazard and risk

A 'hazard' is a situation that poses a level of threat to the accelerator. Hazards are dormant or potential, with only a theoretical risk of damage. Before designing a protection system, the hazards need to be identified and the risk needs to be quantified. Once a hazard becomes 'active' it becomes an incident or accident. Consequences and probability of an incident interact together to create a risk that can be quantified:

$$\text{Risk} = \text{Consequences} \times \text{Probability}. \tag{1}$$

Related to accelerators, the consequences and the probability of an uncontrolled beam loss need to be estimated to get an idea about the risk. Machine protection systems prevent damage to equipment after a failure, thus reducing the risk to an acceptable value. With increasing risks for hazards, the reliability of the machine protection systems must increase in order to keep the risk at the acceptable value. Machine protection needs to be considered during design, construction and operation of the accelerator.

If a specific failure is considered, the consequences of the failure can be estimated, in terms of damage to equipment (repair requiring investment, e.g. in money), in downtime of the accelerator (e.g. in days) and in radiation dose to personnel accessing equipment (e.g. in mSv).

In the estimation of downtime of the accelerator for repairs the availability of spare parts needs to be considered. If the accelerator was operating with beam, radioactive activation of material must be taken into account. It may be necessary to wait for cool-down of irradiated components to reduce the dose before accessing the equipment.

The second factor entering into the risk is the probability of such a failure happening (e.g. measured in the probability of a failure per year).

For beam operation, a list of all possible failures that could lead to beam loss in equipment should be established. This is not obvious, since there is a nearly infinite number of mechanisms for losing the beam. However, the most likely failure modes and in particular the worst-case failures and their probabilities must be considered.

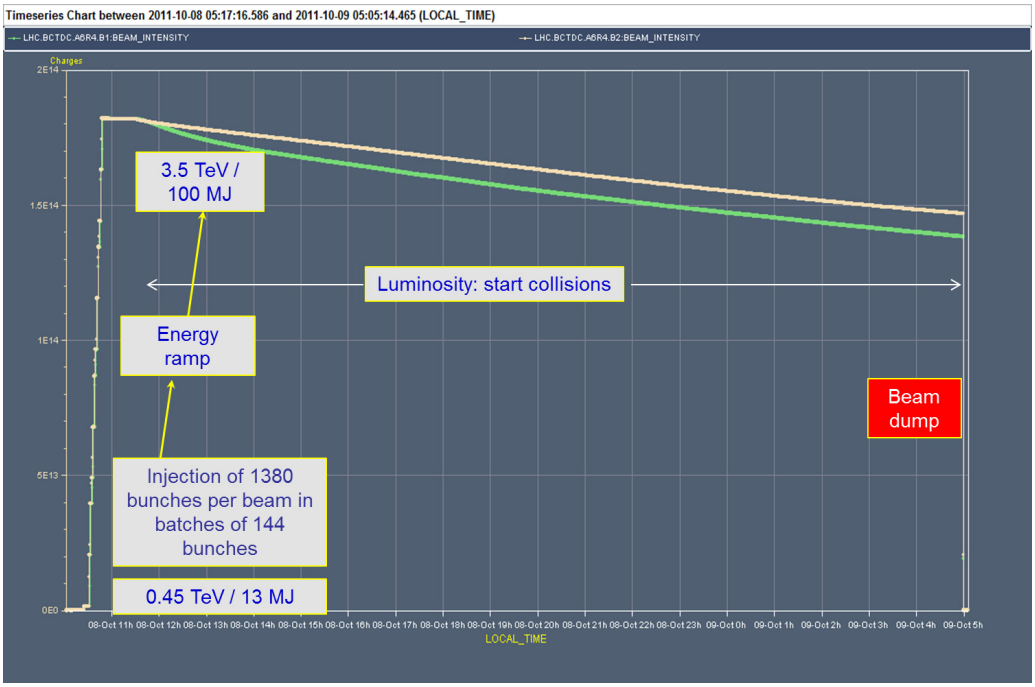


Fig. 6: LHC operation cycle, from injection to beam dump

5 Hazards for a synchrotron: LHC

The LHC is used as an example to discuss hazards for a synchrotron. The operational cycle is shown in Fig. 6. In the LHC, beams are injected at 450 GeV from the SPS, then accelerated to high energy (3.5 GeV and 4 GeV during Run 1 between 2010 and 2012, since 2015 to 6.5 GeV). The beams are then brought into collisions for many hours. At the end of a fill and in case of failure the beams must be extracted by the beam dumping system.

Three different phases during a fill can be identified.

- Injection of the high-intensity beam from the SPS. The energy stored in the injected beam is very high and threatens to cause damage in case of failure.
- When the beams are circulating, there is always the risk of a failure, e.g. a trip of a magnet power converter. Such a trip would change current and magnetic field and particles will be ill-deflected. In this case, the beams need to be safely extracted.
- At the end of a fill or in case of failure, the beams need to be extracted. This is a critical process with a number of hazards.

The general architecture for the essential elements in the machine protection system at LHC is shown in Fig. 7. The LHC has eight sectors and eight insertions; three sectors are related to machine protection: two cleaning insertions with a large number of collimators and one insertion for the beam dumping system. More than 3600 beam loss monitors are installed around the machine. In case of a failure detected by hardware or beam monitors, the beam interlock system transmits a beam dump request to the beam dumping system and the beams are extracted.

The role of the LHC beam dumping system [7] is to safely dispose of the beam when beam operation must be interrupted for any reason. Fifteen fast kicker magnets with a pulse rise time of less than 3 µs deflect the beam by an angle of 280 µrad in the horizontal plane; see Fig. 8. To ensure that all particles are extracted from the LHC without losses, the beam has a particle-free abort gap with a length of 3 µs corresponding to the kicker rise time. The extraction kickers are triggered such that the field increases from zero to the nominal value during this gap when there should be no particles.

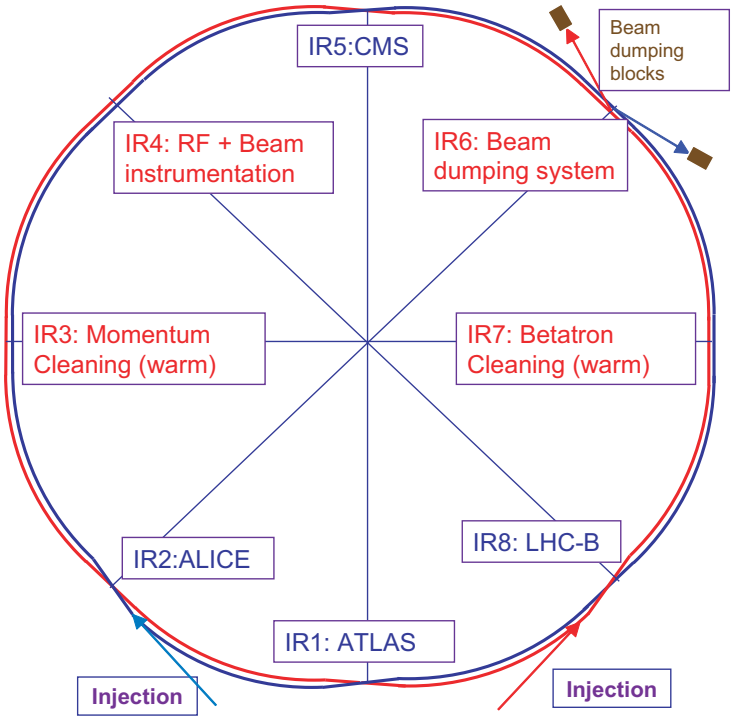


Fig. 7: Layout of LHC with some of the systems for machine protection

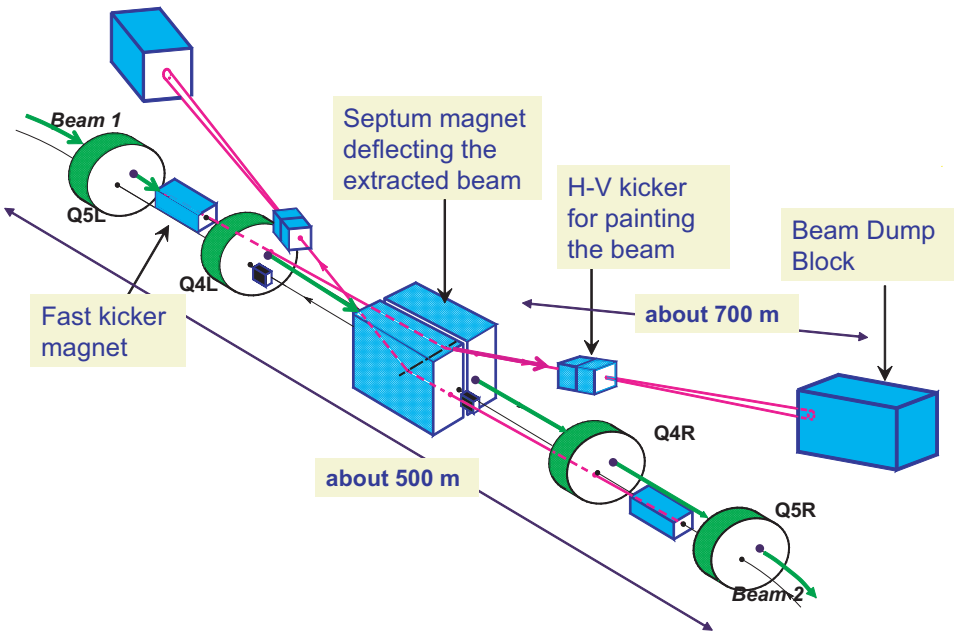


Fig. 8: Layout of the beam dumping systems for both LHC beams (courtesy of M. Gyr)

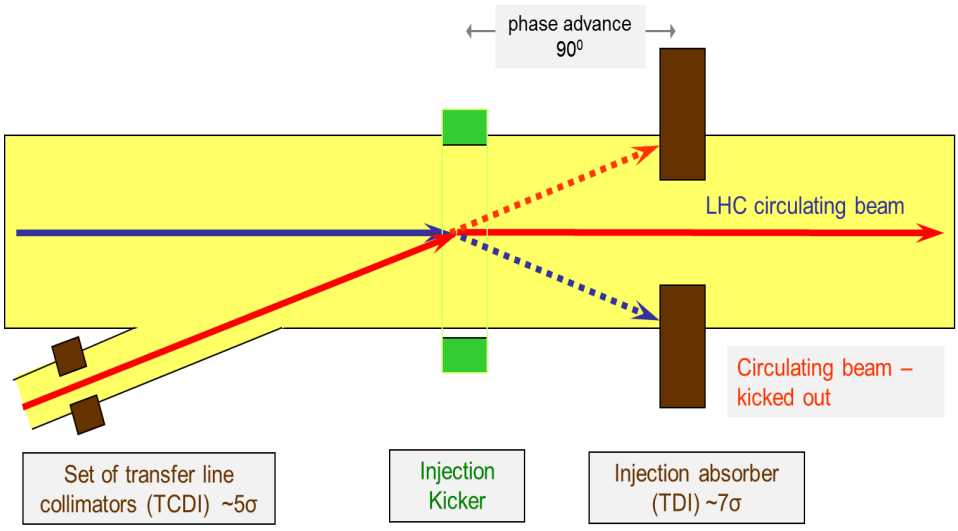


Fig. 9: LHC injection, two failure cases: 1) the injection kicker does not fire to deflect the incoming beam. 2) the injection kicker fires at the wrong time and deflects the circulating beam. In both cases the beam would hit vacuum chamber and equipment. To prevent such accident, the injection absorbers must be correctly positioned to ensure that the beam hits the absorbers for both failure cases.

Downstream of the kickers the beam is deflected vertically by 2.4 mrad towards the beam dump block by 15 septum magnets. A short distance further downstream, 10 diluter kicker magnets are used to paint the bunches in both horizontal and vertical directions to reduce the beam density on the dump block. The beam is transferred through a 700 m long extraction line to increase the transverse r.m.s. beam size from approximately 0.2 to 1.5 mm and to spread the bunches further on the dump block.

The overall shape is produced by the deflection of the extraction and dilution kickers. For nominal beam parameters, the maximum temperature in the beam dump block is expected to be in the order of about 800°C. Access to the dump block becomes difficult since the material will become increasingly activated.

Protection during the injection process is also mandatory. The energy stored in the LHC beam at injection is about one order of magnitude higher than the stored energy at top energy in the beam for other accelerators. An example of a critical failure for the LHC at injection is a failure of the injection kicker. If the kicker does not fire or fires at the wrong time, the incoming beam will not be deflected or the circulating beam will be deflected (see Fig. 9). In both cases, without protection, the beam would hit the vacuum chamber and damage equipment. During the injection process injection collimators are positioned in the vacuum chamber to capture the mis-kicked beam. This type of failure, e.g. a wrong kick of the injection kicker, has happened already several times. Since the injection absorber was always at the correct position there was no damage; however, due to grazing beam incidence superconducting magnets downstream of the injection region quenched.

5.1 Damage potential of a high-energy hadron beam

Beams at very high energy can have a tremendous damage potential for an accelerator such as LHC; damage to metals is expected for about 10^{10} protons. One LHC bunch has about 1.5×10^{11} protons, in total up to 2808 bunches. In case of catastrophic beam loss, the LHC could be possibly damaged beyond repair. The penetration of 2808 bunches (e.g. after a kicker failure) into a metal block such as a magnet has been estimated to be about 20 to 30 m (hydrodynamic beam tunnelling) [8].

In order to estimate the consequences of the beam impacting on a target, the time structure of the beam plays an essential role. Bunches arrive every 25 or 50 ns. The first bunches arrive and deposit their

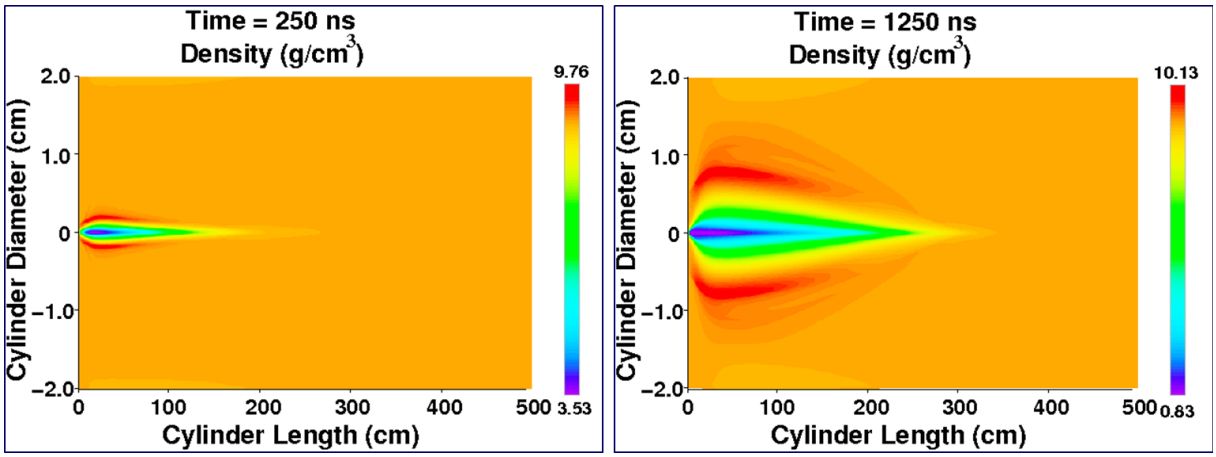


Fig. 10: Density of a copper target after an impact of 250 and 1250 ns for the FCC

energy. This can lead to an increase of the temperature, pressure and finally to a reduction of the target material density. Bunches arriving later travel further into the target since the material density is reduced (predicted also for SSC [9]).

The calculations were performed in the following way. A LHC beam impact on a solid cylindrical target is assumed, with 2808 bunches, each bunch with 1.1×10^{11} protons, a beam size of $\sigma = 0.5$ mm, 25 ns bunch distance, a target length of 6 m and a radius of 5 cm. Various target materials were considered, e.g. graphite with a density of 2.3 g/cm^3.

The energy deposition for a few bunches is calculated with FLUKA. The hydrodynamic code BIG2 uses the 3D energy deposition to calculate temperature, pressure and density of the target. The programs are run iteratively with the FLUKA 3D energy loss data used as input to BIG2, and the BIG2 3D density data used as input for FLUKA. The modified density distribution is used in FLUKA to calculate the energy loss corresponding to this new density distribution. The new energy loss distribution is used in BIG2, which is run for a time step.

For the FCC, a 100 km long accelerator to provide proton collisions at a centimetre energy of 100 TeV was used; calculations were performed for a particle energy of 40 TeV [10]. A copper target with a length of 5 m and a radius of 2 cm was taken. The simulations are very time consuming; this one took about 15 months. The density of the target after 250 ns and 120 ns is shown in Fig. 10; it decreased to 3.5 g/cm^3 after 250 ns beam impact and to 0.83 g/cm^3 after 1250 ns.

Figure 11 shows the density versus depth in the target for the simulation. In the figure, we present the density versus axis at different times during irradiation. Curve 'a' represents the time when two bunches have been delivered while later curves are plotted using an interval of 150 ns. We consider 6 g/cm^3 as the reference point on the density curve and calculate the speed with which this point moves towards the right. It is seen that at 800 ns, the depletion front achieves a steady average speed of 1.1×10^6 m/s. The total beam duration considering all 10 600 bunches is 265 µs that leads to a penetration distance of about 290 m. This means that in case of wrong deflection of the beam, the beam and the shower will penetrate through about 290 m of solid copper. If one considers a 50 TeV proton beam, the penetration distance could be up to 350 m.

An experiment was performed at the CERN-SPS HiRadMat facility with a 450 GeV proton beam to validate the simulation technique. Solid copper targets were facially irradiated by the beam and measurements confirmed hydrodynamic tunnelling of the protons and their showers. Simulations have been done by running the energy deposition code FLUKA and the 2D hydrodynamic code, BIG2, iteratively. Very good agreement has been found between the simulations and the experimental results [11] providing confidence in the validity of the studies for the LHC and FCC.

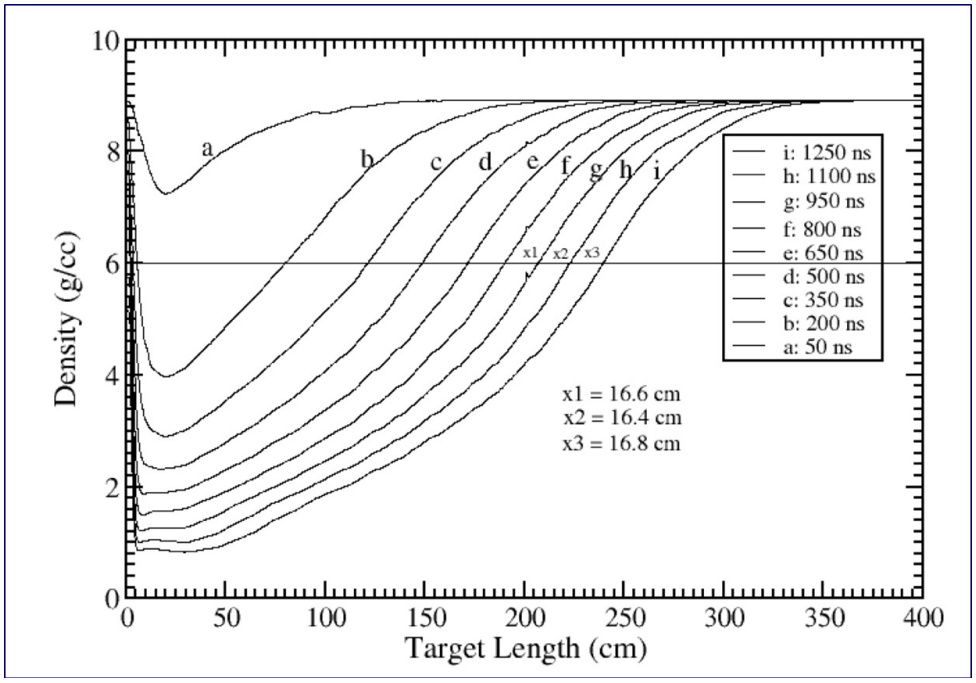


Fig. 11: Density of a copper target after an impact of 250 and 1250 ns

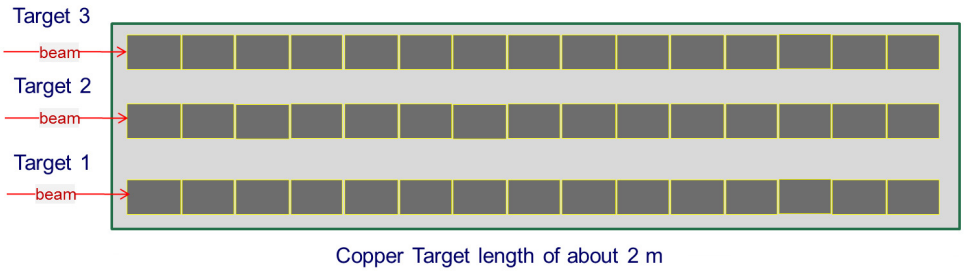


Fig. 12: Layout of the experiment with three different targets

Three targets (see Figs. 12, 13 and 14) were irradiated with bunch trains, each bunch of 50 ns, of different intensities:

- Target 1: 144 bunches, 1.9×10^{11}, $\sigma = 2.0$ mm, no tunnelling expected;
- Target 2: 108 bunches, 1.9×10^{11}, $\sigma = 0.2$ mm, tunnelling expected;
- Target 3: 144 bunches, 1.9×10^{11}, $\sigma = 0.2$ mm, tunnelling expected.

The measured penetration length of the beam and the results from the hydrodynamic simulations using FLUKA and BIG2 were obtained for all targets. Excellent agreement between hydrodynamic simulations and experimental results was found.

6 Protection for a high-intensity proton linac: the European Spallation Source

The European Spallation Source (ESS) being built at Lund, Sweden is designed to accelerate a proton beam with an average power of 5 MW and to direct the protons onto a target. Operation of the ESS will be at a frequency of 14 Hz, with a pulse length of 2.86 ms and a peak power of 125 MW. The layout of the ESS accelerator is shown in Fig. 15.

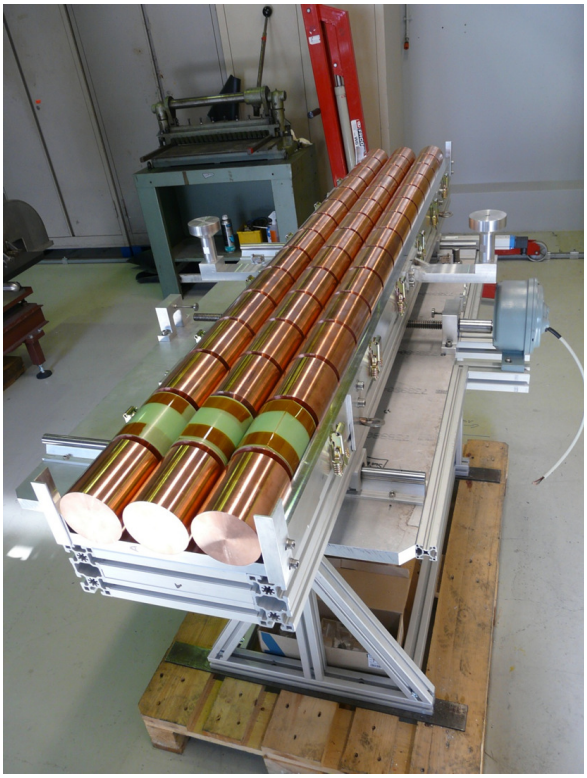


Fig. 13: Picture of the targets before irradiation

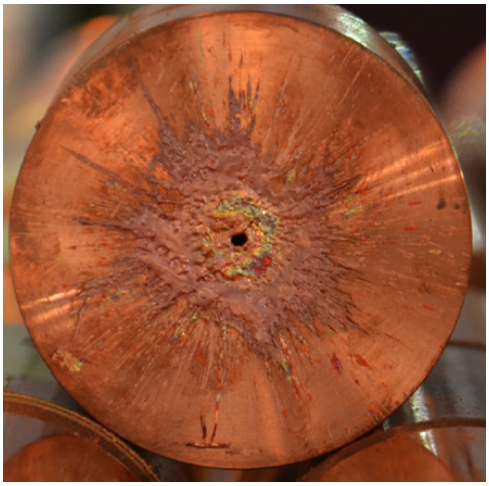


Fig. 14: Picture of the front face of one target block after irradiation

It is assumed that the beam is lost due to a failure after the acceleration section. An example is a trip of the power converter for the vertically deflecting magnets. If the beam loss takes, say, 1 ms, the deposited energy is up to 125 kJ (peak power of 125 MW during 1 ms), for 1 s up to 5 MJ. It is required to inhibit the beam after detecting uncontrolled beam loss as fast as possible. There is some delay between the detection of a failure (e.g. detection of beam losses by a beam loss monitor) and 'beam off'. Figure 16 shows the time to melt copper and steel in the case where the proton beam hits a metal surface between 3 and 80 MeV/c [12]. For example, after the drift tube normal-conducting linac (DTL), the proton energy is 78 MeV/c. In case of a beam size of 2 mm radius, melting would start after a beam impact of about 200 µs. Inhibiting of the beam after a failure is detected should be in about 10% of this time; see [13]).

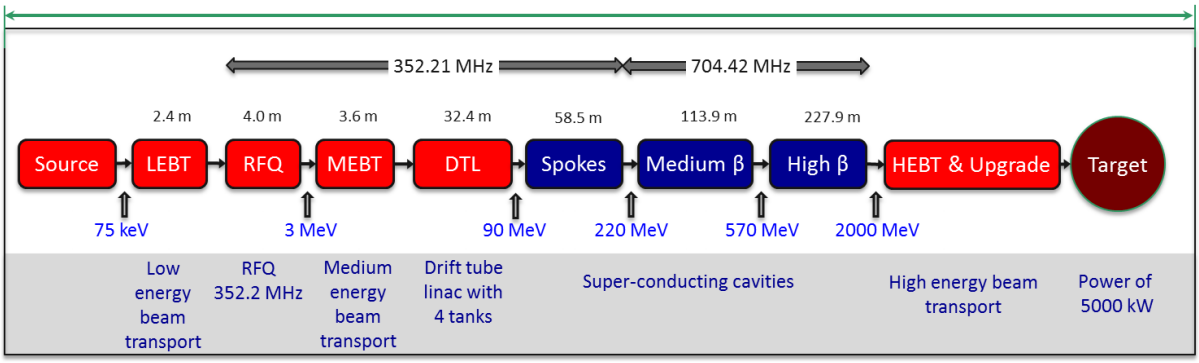


Fig. 15: Layout of the ESS accelerator: the source, low-energy beam transport and RFQ are followed by the medium-energy beam transport. The protons are accelerated by a normal-conducting linac, followed by three sections of superconducting cavities. In the high-energy beam transport line the protons are transported to the target.

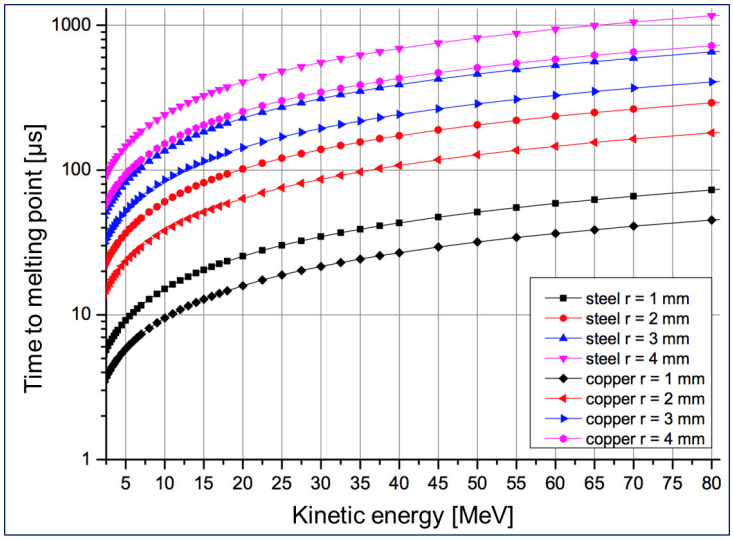


Fig. 16: Time to melt copper and steel, as a function of proton momentum for different beam sizes [12]

The energy stored in the beam at a given moment is relatively small in the low-energy part, in the medium-energy part and in the high-energy part. In case of a failure, the beam needs to be switched off at the source. In between two pulses (about 70 ms), it must be ensured that the parameters of the accelerator allow for correct beam transmission, or do not start the next pulse. If something is wrong and is not detected before the pulse by monitors, the beam must be stopped as soon as possible.

A realistic example for a failure is a trip of the power converter for the bending magnet deflecting the beam vertically Fig. 17. Assume that the power supply for the bend in HEBT-S2 fails and the magnet stops deflecting the beam. A mean time between failures (MTBF) for a power converter of 100 000 hours (15 years) is assumed, already a very good value. If this happens, the beam is not deflected and hits the vacuum chamber. The consequences can be serious; damage of magnet or vacuum pipe, possibly pollution of superconducting cavities.

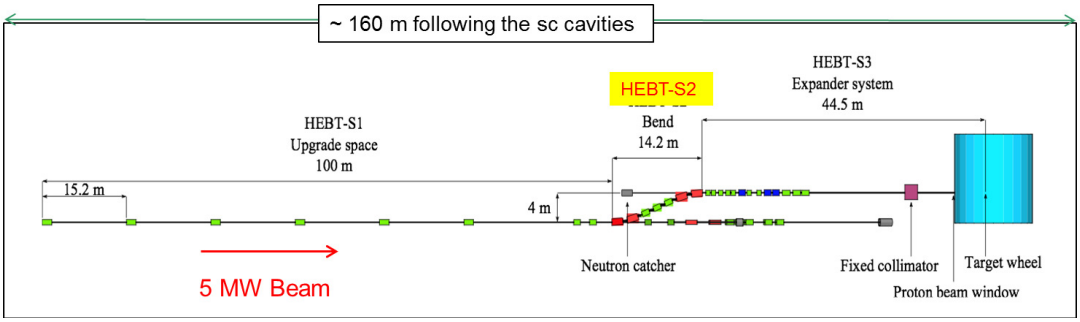


Fig. 17: Illustration of the vertical bending magnet for ESS, deflecting the 5 MW beam into the target plane

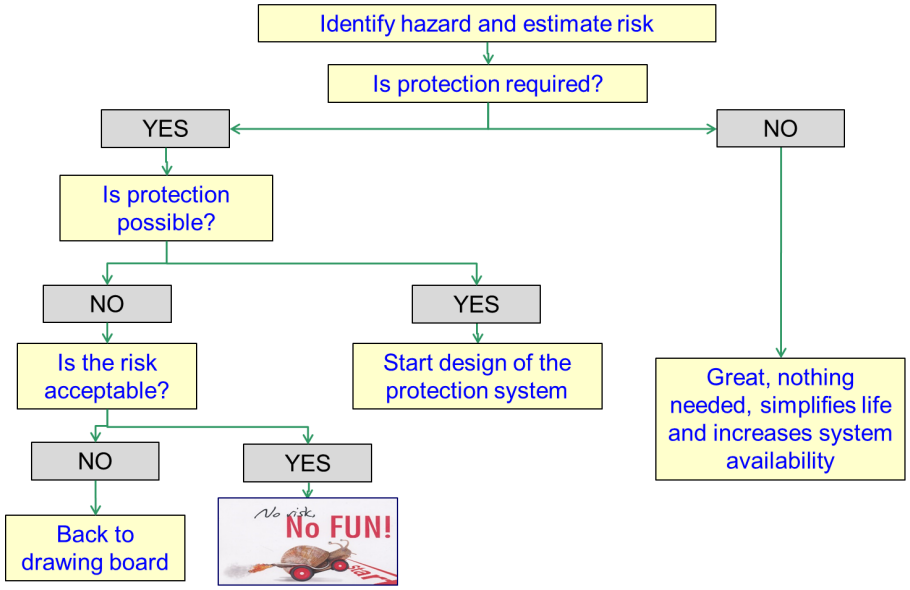


Fig. 18: Analysing if a machine protection system is required

7 Machine protection

7.1 Analysing the need for a machine protection system

The design of a machine protection system requires substantial resources, it can be expensive and it can reduce the availability of the accelerator since it will contribute to false trips. Therefore, it should only be built if absolutely necessary. Figure 18 gives an idea of how to analyse if such a system is required. Different outcomes are possible: there is no need for machine protection or there is a need for machine protection. In some cases protection is not feasible, and the system design must be modified.

7.2 Classification of failures

In the first step different types of failures that can cause beam losses are identified and classified:

- hardware failure (trip of a power converter, magnet quench, AC distribution failure, object in vacuum chamber, vacuum leak, RF trip, kicker magnet misfire etc);
- control failure (wrong data, wrong magnet current function, trigger problem, timing system failure, feedback failure etc);
- operational failures (chromaticity/tune/orbit wrong values etc);
- beam instability (due to too high beam current/bunch current/e-clouds etc).

The most important parameters for a failure are:

- time constant for beam loss after the occurrence of the failure;
- probability of the failure occurring;
- damage potential in case no mitigation is applied.

An accurate understanding of the time constant is required, since this determines the reaction time of the machine protection systems. The risk defined as $\mathrm{risk} = \mathrm{consequences} \times \mathrm{probability}$ is another important input determining the required reliability for the protection systems. For very high risk the protection systems must be extremely reliable.

7.3 Time constant for failures

The time constant for beam loss after a failure varies from nanoseconds to many seconds.

Single-passage beam losses in the accelerator complex have a time constant of a few nanoseconds to some tens of microseconds. In a circular accelerator such losses are related to failures of fast kicker magnets for injection and extraction. If other fast kicker magnets are present, for example for diagnostics, failures of such devices must also be considered. For failures of fast kicker magnets it is not possible to extract the beam or to stop the beam at the source; the particles will travel determined by the electromagnetic field along their path.

Single-passage beam losses are also an issue for any accelerator operating with pulsed beams. In between two pulses, equipment parameters can change (e.g. a magnet power supply can trip). During the following beam pulse, the beam would be mis-steered and can cause damage. This is typically the case for failures in a transfer line between accelerators (e.g. from SPS to LHC) or from an accelerator to a target station (target for secondary particle production or beam dump block). This is also an issue for linear accelerators operating with pulsed beams.

Very fast beam losses with a time constant in the order of 1 ms, e.g. multiturn beam losses in circular accelerators. Such losses can appear due to a large number of possible failures, mostly in the magnet powering system, with a typical time constant of about 1 ms to many seconds.

Fast beam losses with a time constant of 10 ms to seconds, due to many different effects. Beam instabilities in LHC are in general in this time range.

Slow beam losses take many seconds, e.g. due to non-optimized parameters, but also due to a failure.

7.4 Principles for machine protection

There are some principles for machine protection that need to be considered:

i) protect the machine;
ii) protect the beam;
iii) provide the evidence (at CERN, by the so-called post-mortem system).

7.4.1 Protect the machine

The highest priority is clearly to avoid any damage to accelerator equipment.

7.4.2 Protect the beam

The objective is to maximize beam time, but complex protection systems reduce the availability of the machine. The number of 'false' interlocks stopping operation must be minimized. This is a trade-off

between protection and operation. A 'false' interlock is defined as an interlock that stops operation even though there is no risk (example: a temperature sensor reading a wrong value, therefore switching off the power converter of a magnet and stopping beam operation).

7.4.3 Provide the evidence

If the protection systems stop operation (e.g. dump the beam or inhibit injection), clear diagnostics should be provided [14]. If something goes wrong (leading to damage, but also a near miss), it should be possible to understand the event. This needs synchronized transient recording of all the important parameters in all relevant systems, as well as long-term logging of parameters with reduced frequency (such as 1 Hz). Examples are the current in all magnets, beam position, beam losses and beam intensity. The frequency of transient recording depends on the system and can go from Hz to MHz.

7.5 Active and passive protection

The best strategy is to prevent a specific failure from happening. As an example, fast diagnostic kicker magnets that could deflect the beams into the vacuum chamber wall should only be installed in high-intensity machines if they are indispensable.

7.5.1 Active protection

Failure should be detected as early as possible, with priority at the hardware level. For most failures, this strategy allows stopping beam operation before the beam is affected. This requires monitoring of the hardware (such as state signals, parameters etc). As an example, a trip of a magnet power converter should be detected as early as possible.

It is not always possible to detect failures at the hardware level. The second method is to detect the initial consequences of a failure with beam instrumentation and to stop the beam before equipment is damaged. This requires reliable beam instrumentation.

When a failure is detected, beam operation must be stopped. For synchrotrons and storage rings the beam is extracted by a fast kicker magnet into a beam dump block. Injection must be stopped. For linacs the beam is stopped in the low-energy part of the accelerator by switching off the source, deflecting the low-energy beam by electrostatic plates ('choppers') or by switching off the RFQ for proton linacs.

An electronic system (beam interlock system) links the different protection systems. It ensures that the beam is extracted from a synchrotron, injection is stopped, RF acceleration might be stopped (for linacs). The interlock system might include complex logic.

7.5.2 Passive protection

There are failures (e.g. ultra-fast losses) when active protection is not possible. One example is the protection against mis-firing of an injection or extraction kicker magnet. A beam absorber or collimator is required to stop the mis-kicked beam in order to avoid damage. All possible beam trajectories in such case must be considered, and the absorbers must be designed to absorb the beam energy without being damaged. Another example is a fast extraction of a high-intensity beam from a circular accelerator into a transfer line. When the extraction takes place, the parameters of the transfer line must be correctly set since in case of a wrong magnet current the beam could be deflected into the vacuum chamber.

7.6 LHC strategy for machine protection

Machine protection starts with a careful commissioning of the magnet powering system, considering that an energy of about 10 GJ is stored in the superconducting magnets. Magnet protection and powering interlocks must be operational long before starting beam operation. The strategy for LHC machine protection when operating with beam reflects many of the principles that have been discussed above.

i) Definition of aperture by collimators (beam cleaning system).
ii) Early detection of failures of equipment acting on beams generates a beam dump request, possibly before the beam is affected (interlocks for the different hardware system, for LHC essentially the magnet and powering system).
iii) Active monitoring of the beam parameters with beam instruments detecting abnormal beam conditions and generating beam dump requests within a single machine turn (using beam loss and other beam monitors).
iv) Reliable transmission of beam dump requests from a large variety of systems to the beam dumping system by a beam interlock system. An active signal is required for operation; the absence of the signal is considered as a beam dump request and injection inhibit.
v) Reliable operation of the beam dumping system for dump requests or internal faults, safely extracting the beams onto external dump blocks.
vi) Passive protection by beam absorbers and collimators for specific cases of failure.

7.7 Design considerations for protection systems

There are several principles that should be considered in the design of protection systems, although it might not be possible to follow all these principles in all cases.

i) If the protection system does not work, it is better stopping operation rather than continuing and risking damaging equipment.
ii) Fail-safe design: in case of a failure in the protection system, protection functionalities should not be compromised. As an example, if the cable that triggers the extraction kicker of the beam dumping system is disconnected, operation must stop.
iii) Detection of internal faults: the protection system must monitor the internal status. In case of an internal fault, the fault should be reported. If the fault is critical, operation must be stopped.
iv) Remote testing should be an integral part of the design, for example between two runs. This allows verification of the correct status of the system.
v) Critical equipment should be redundant (possibly diverse redundancy, with the same or similar functions executed by different systems).
vi) Critical processes for protection should not rely on complex software running under an operating system and requiring the general computer network.
vii) It should not be possible to remotely change the most critical parameters. If parameters need to be changed, the changes must be controlled and logged and password protection should ensure that only authorized personnel can do the change.
viii) Safety, availability and reliability of the systems should be demonstrated. This is possible by using established methods to analyse critical systems and to predict failure rates.
ix) Operate the protection systems early on before they become critical, to gain experience and to build up confidence. This could be done before beam operation, or during early beam operation when the beam intensity is low.
x) It is inevitable to disable interlocks (e.g. during the early phase of commissioning and for specific tests). Managing interlocks (e.g. disabling) is common practice. Keep track and consider it in the system design. Example for LHC: masking of some interlocks is possible, but only for low-intensity/low-energy beams (‘safe beams’).

Fig. 19: Beam loss monitor at LHC

8 Machine protection subsystems

8.1 Beam loss monitors

Beam loss monitors (BLMs) are used for monitoring beam losses to understand the performance of the accelerator as well as for machine protection. If used for protection, it is important that the monitors cover the entire accelerator and there is no region without BLMs where beam losses can occur.

The monitors should be fast, for LHC down to 40 µs, in order to detect beam losses in time to stop operation. They should be designed such that they can trigger a beam dump and stop operation before very fast beam losses damage equipment. There are about 3600 chambers distributed over the ring to detect abnormal beam losses and if necessary trigger a beam abort [15].

For LHC (Fig. 19) and several other accelerators, ionization chambers are used to detect beam losses. The reaction time is down to microseconds; they can have a very large dynamic range exceeding 10^8.

Figure 20 shows the beam losses around LHC during regular luminosity operation. Losses are low in the arcs with the superconducting magnets, higher in the insertions with the experiments due to the debris of the collisions and very high in the betatron cleaning insertions. Figure 21 shows the losses recorded after a beam dump during regular operation. A UFO (unidentified falling object) in the arc between the betatron cleaning insertion and the LHC experiment caused locally an increase of beam losses. The UFO, very likely dust particles getting into the beam (see [16]) causes very fast losses and when the threshold of one BLM is exceeded, the beams are dumped (see Fig. 22).

8.2 Beam cleaning system

The LHC operates with a stored beam with an energy of 360 MJ (nominal parameters). A beam lifetime of 10 min corresponds to a beam loss of 500 kW that should not be lost in superconducting magnets. The strategy to keep beam losses at an acceptable level is as follows:

i) avoid beam losses as far as possible;
ii) define the aperture by collimators;
iii) capture continuous particle losses with collimators at specific locations.

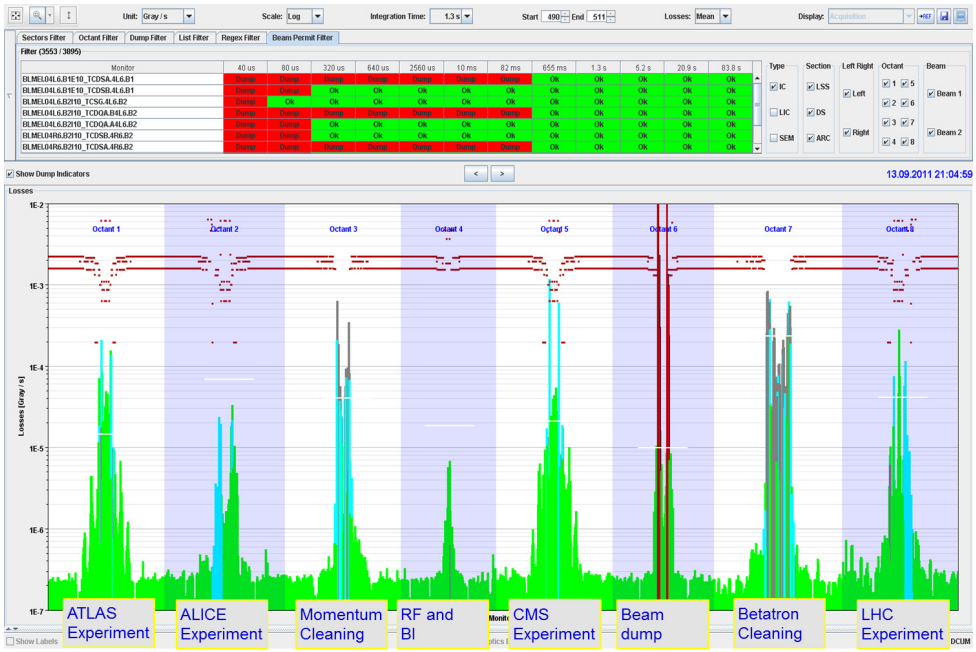


Fig. 20: Regular beam losses during luminosity operation

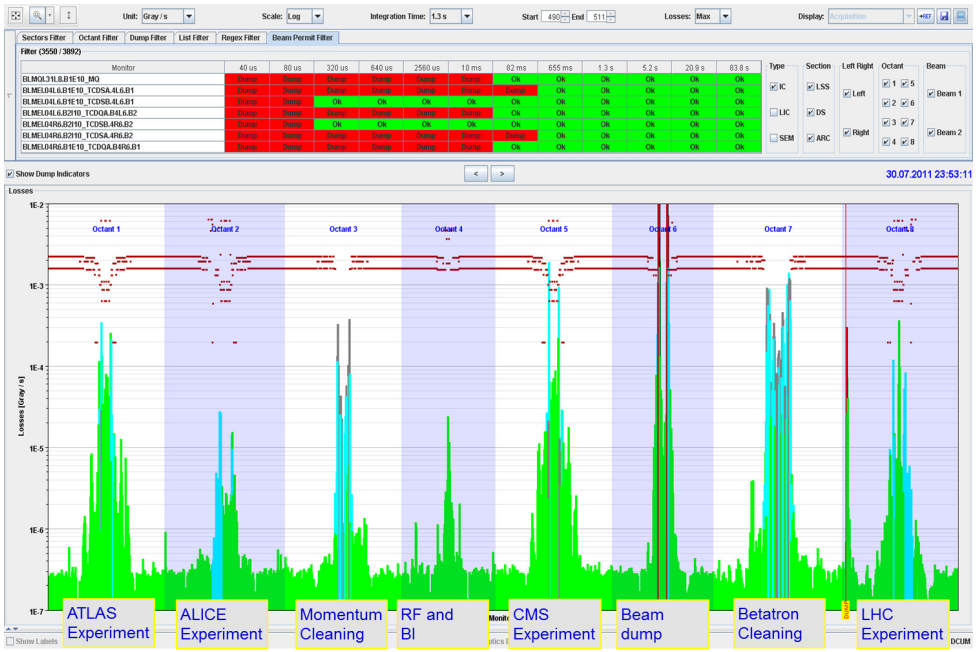


Fig. 21: Beam losses with UFO during luminosity operation

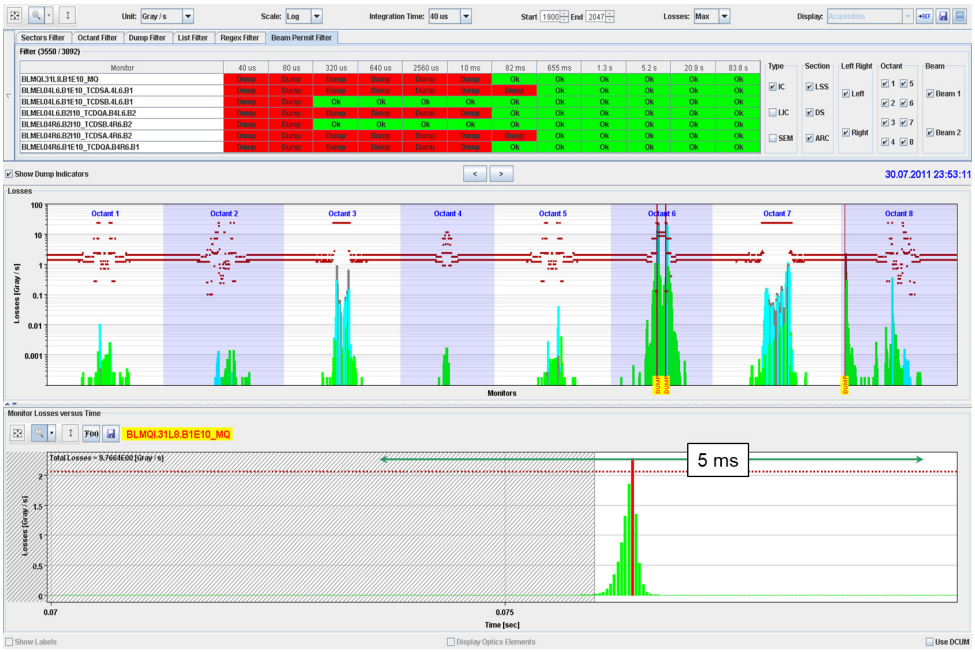


Fig. 22: Time structure of the beam losses caused by a UFO

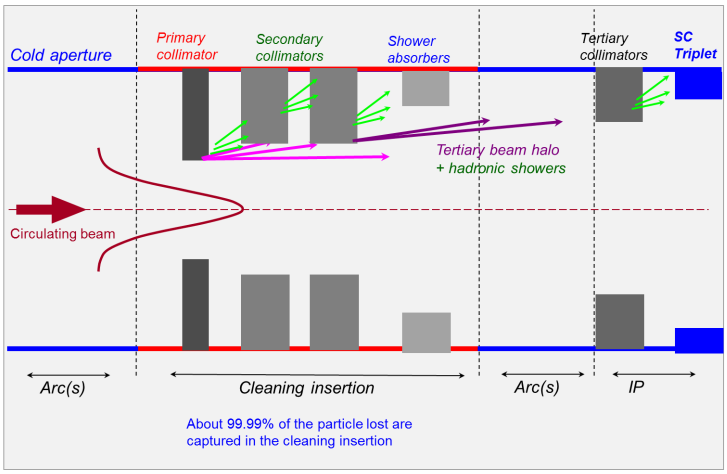


Fig. 23: Illustration of beam cleaning at LHC

A collimation system is a very efficient system to avoid too high beam losses in the accelerator (so-called beam cleaning) [17]. It can be very complex with (massive) material blocks close to the beam installed in an accelerator to capture halo particles. Figure 23 illustrates the LHC beam cleaning system.

The collimation system at LHC reduces the losses by four orders of magnitude and also captures fast accidental beam losses. About 100 collimators are installed in LHC. Figure 24 shows the view of one of the two-sided collimators; it is closed down to 2 mm when operating at 7 TeV/c. Collimators (or beam absorbers) are equally important to capture mis-steered beam.

8.3 Interlock systems

Figure 25 illustrates the interlock systems for LHC. The heart is the beam interlock system that receives beam dump requests from many connected systems. If a beam dump request arrives, a signal is sent to

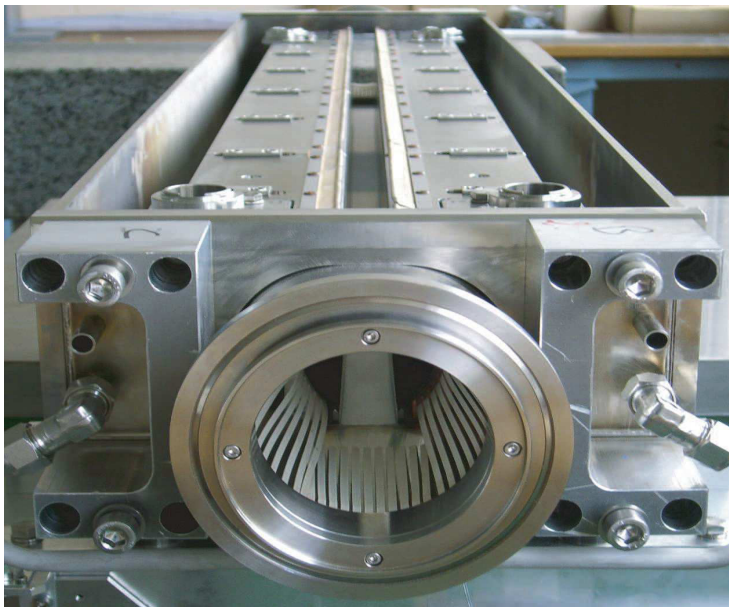


Fig. 24: Collimator at LHC

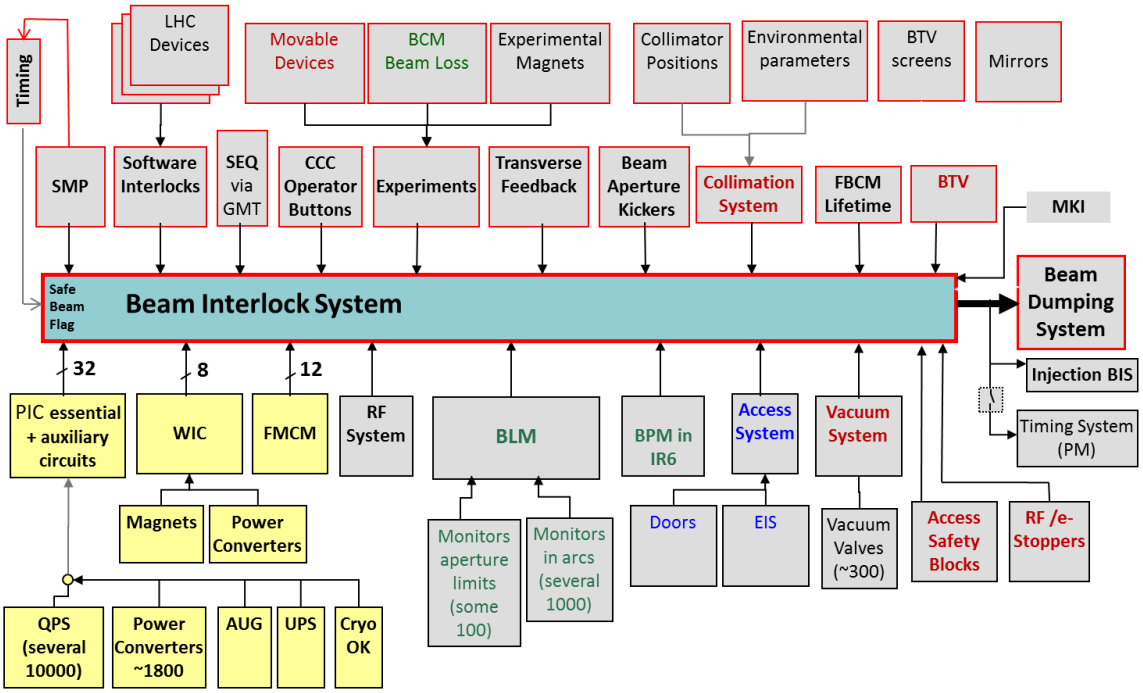


Fig. 25: Beam interlock system at LHC as well as connected systems

the beam dumping system to request the extraction of the beams. At the same time, a signal is sent to the injection system to block injection into LHC as well as extraction of beam from the SPS. A third signal is provided for the timing system that sends out a request to many LHC systems for providing data that were recorded before the beam dump, to understand the reasons for the beam dump (typically beam loss, beam position, beam current, magnet currents etc).

The most complex system of LHC is the superconducting magnets and powering system. The powering interlock system (PIC) ensures communication between systems involved in the powering of the LHC superconducting magnets. This includes power converters, magnet protection system, UPSs

(un-interruptible power supplies), emergency stop of electrical supplies (AUG) and the cryogenic system. As an example, in case a magnet quench is detected by the quench protection system (QPS), the power converter must stop. In total, there are several tens of thousands of interlock signals. When a failure is detected that risks stopping the powering of magnets, a beam dump request is sent to the beam interlock system. A second system manages interlocks from the normal-conducting magnets and their power supplies (WIC) that ensures protection of normal-conducting magnets in case of overheating.

The machine interlock system is strictly separated from interlocks for personnel safety such as the personnel access system; however, an interlock from the access system is sent to the beam interlock system.

As shown in Fig. 25, many other systems also provide beam dump requests in case of failure: beam loss monitors, other beam monitors, movable devices and LHC experiments.

9 Conclusions

Machine protection goes far beyond the equipment protection and across many systems. It requires the understanding of many different types of failures that could lead to beam loss. It requires fairly comprehensive understanding of all aspects of the accelerator (accelerator physics, operation, equipment and instrumentation) and touches many aspects of accelerator construction and operation.

Machine protection is becoming increasingly important for future projects, with increased beam power and energy density (W/mm^2 or J/mm^2) and increasingly complex machines.

Protection of equipment, even when operating without beam, must not be forgotten. The largest accident happening at an accelerator was the rupture of a superconducting cable at LHC in 2008 due to the very large energy stored in the superconducting magnet system.

Acknowledgements

I wish to thank many colleagues from CERN and ESS and the authors of the listed papers for their help and for providing material for this article. Thanks to Laura Grob for reading and commenting on the manuscript. This paper is an updated version of the lecture given at a CAS in 2013 [18].

References

[1] E. Metral *et al.*, Lessons learnt and mitigation measures for the CERN-LHC equipment with RF fingers, 4th International Particle Accelerator Conf., IPAC2013, Shanghai, China, 2013.

[2] A. Fasso *et al.*, The physics models of FLUKA: status and recent development, Computing in High Energy and Nuclear Physics, La Jolla, California, USA, 2003, MOMT005 [hep-ph/0306267] (2003).

[3] https://mars.fnal.gov/, Nikolai Mokhov, last update August 3, 2016.

[4] http://geant4.web.cern.ch/geant4/, last update August 3, 2016.

[5] Wikipedia, keyword: fuel efficiency.

[6] M. Ross, The next linear collider machine protection system, 1999 Particle Accelerator Conference, New York 1999.

[7] J. Uyhoven *et al.*, Results from the LHC beam dump reliability run, Technical Report LHC-Project-Report-1120, CERN, Geneva (2008).

[8] N.A. Tahir, V.E. Fortov, B. Goddard, D.H.H. Hoffmann, V. Kain, I.V. Lomonosov, A.R. Piriz, R. Schmidt, A. Shutov and M. Temporal, *J. Appl. Phys.* **97**(8) (2005) 83532-1. https://doi.org/10.1063/1.1888031

[9] D.C. Wilson, C.A. Wingate, J.C. Goldstein, R.P. Godwin and N.V. Mokhov, Hydrodynamic calculations of 20-TeV beam interactions with the SSC beam dump, US Particle Accelerator Conf., Washington, DC, 17–20 May 1993, CERN-ATS-2011-064 (1993).

[10] N.A. Tahir, A.R. Piriz, R. Schmidt, A. Shutov, F. Burkart and D. Wollmann, *Phys. Rev. Accel. Beams* **19**, 081002 (2016).

[11] F. Burkart, A. Piriz, R. Schmidt, A. Shutov, N. Tahir, D. Wollmann and M. Zerlauth, Experimental and simulation studies of hydrodynamic tunneling of ultra-relativistic protons, Proceeding of IPAC 2015, Richmond, VA, USA, 2015.

[12] L. Tchelidze, In how long the ESS beam pulse would start melting steel/copper accelerating components, Technical Report ESS AD Technical Note ESS/AD/0031, ESS, Lund, Sweden (2012).

[13] A. Nordt, A. Apollonio and R. Schmidt, Overview on the design of the machine protection system for ESS, 5th International Particle Accelerator Conf., IPAC 2014, Dresden, Germany, 15–20 June 2014.

[14] M. Zerlauth *et al.*, The LHC post mortem analysis framework, ICALEPCS 2009, Kobe, Japan, 2009.

[15] B. Dehning *et al.*, Overview of LHC beam loss measurements, 2nd International Particle Accelerator Conf., IPAC 2011, San Sebastian, Spain, 4–9 September 2011, CERN-ATS-2011-064 (2011), p. 3.

[16] T. Baer, J. Wenninger and E. Elsen, Very fast losses of the circulating LHC beam, their mitigation and machine protection, Ph.D. thesis, Hamburg University, October 2013, presented 25 November 2013.

[17] S. Redaelli, in Proceedings of the Joint International Accelerator School: Beam Loss and Accelerator Protection, Newport Beach, United States, 5–14 November 2014, edited by R. Schmidt, CERN-2016-002 (CERN, Geneva, 2014), pp. 403-437, https://doi.org/10.5170/CERN-2016-002.403.

[18] R. Schmidt, in Proceedings of the CAS–CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 18–29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp. 221-243, https://doi.org/10.5170/CERN-2016-009.221.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Multi-bunch Feedback Systems[1]

M. Lonza, presented by H. Schmickler
Elettra Synchrotron Light Laboratory, Sincrotrone Trieste S.C.p.A., Trieste, Italy

Abstract
Coupled-bunch instabilities excited by the interaction of the particle beam with its surroundings can seriously limit the performance of circular particle accelerators. These instabilities can be cured by the use of active feedback systems based on sensors capable of detecting the unwanted beam motion and actuators that apply the feedback correction to the beam. Advances in electronic technology now allow the implementation of feedback loops using programmable digital systems. Besides important advantages in terms of flexibility and reproducibility, digital systems open the way to the use of novel diagnostic tools and additional features. We first introduce coupled-bunch instabilities, analysing the equation of motion of charged particles and the different modes of oscillation of a multi-bunch beam, showing how they can be observed and measured. Different types of feedback systems will then be presented as examples of real implementations that belong to the history of multi-bunch feedback systems. The main components of a feedback system and the related issues will also be analysed. Finally, we shall focus on digital feedback systems, their characteristics, and features, as well as on how they can be concretely exploited for both the optimization of feedback performance and for beam dynamics studies.

Keywords
Multi-bunch; feedback systems.

1 Coupled-bunch instabilities

1.1 Introduction

The demand for high brightness in synchrotron light sources and luminosity in circular colliders require the storage of intense particle beams with many bunches. The interaction of these beams with the environment in which they move can give rise to collective effects called ‘coupled-bunch instabilities’. Collective instabilities can be seen as follows: charged particles in a storage ring are subject to transverse (betatron) and longitudinal (synchrotron) oscillations that are normally damped by a natural damping mechanism. The electromagnetic field created by the bunches can interact with the surrounding metallic structures generating ‘wake fields’, which act back on the beam producing growth of the oscillations. If the growth is stronger than the damping the oscillation becomes unstable. Since the electromagnetic fields created by the bunches are proportional to their charge, the onset of instabilities, and their amplitude are normally current-dependent.

Large-amplitude instabilities can cause beam loss, thus limiting the maximum stored current. If non-linear effects prevent beam loss by saturating the oscillation amplitude, instabilities degrade the beam quality and reduce the brightness or the luminosity of the accelerator.

[1] This paper has been previously published in the Proceedings of the CAS-CERN Accelerator School: Advanced Physics Course, Trondheim, Norway, 18 - 29 Aug 2013, pp.503-546 (CERN-2014-009), https://doi.org/10.5170/CERN-2014-009.503

https://doi.org/10.23730/CYRSP-2017-003.471

1.2 Sources of coupled-bunch instabilities and cures

1.2.1 Cavity high-order modes

High-Q spurious resonances of the accelerating cavity structure are excited by the bunched beam, and act back on the beam itself. Each bunch affects the following bunches through the wake fields excited in the cavity. In this way the cavity high-order modes (HOMs) can couple with a beam oscillation mode having the same frequency and give rise to instability. Both transverse and longitudinal multi-bunch modes can be excited (Fig. 1).

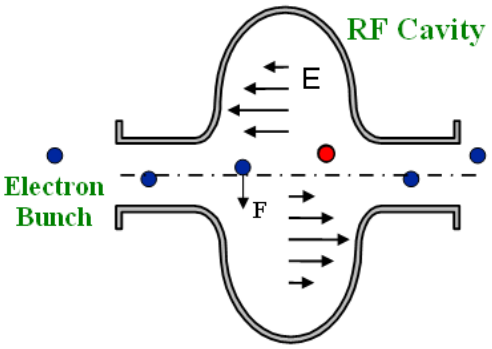


Fig. 1: RF cavity HOMs excited by a particle beam

Excitation of cavity HOMs can be avoided by a thorough design of the RF cavity or by the use of mode dampers made of antennas and resistive loads. Tuning of HOM frequencies through plungers or by changing the temperature of the cavity-cooling water is also currently used in many accelerators [1].

1.2.2 Resistive wall impedance

Resistive wall transverse instabilities are generated by the interaction of the beam with the vacuum chamber through the so-called 'skin effect' [2] (Fig. 2). They are particularly strong where low-gap chambers and in-vacuum insertion devices (undulators and wigglers) are used.

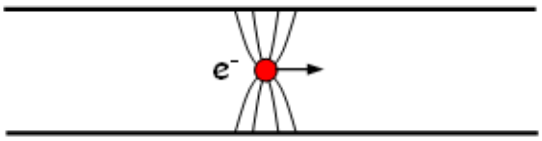


Fig. 2: Interaction of a bunch with the vacuum chamber

This effect can be reduced with an optimization of the vacuum chamber geometry and by using vacuum pipes made of low-resistivity materials.

1.2.3 Interaction with vacuum chamber objects

Coupled-bunch instabilities in both the transverse and longitudinal planes can be created by discontinuities of the vacuum chamber or small cavity-like structures located, for example, in beam position monitors (BPMs), vacuum pumps, bellows, etc. (Fig. 3).

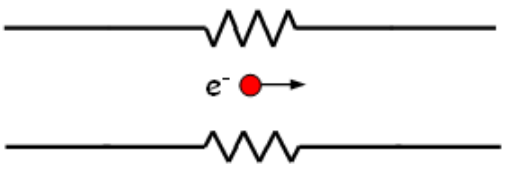


Fig. 3: Discontinuity of the vacuum chamber created by a bellows

The detrimental effect on coupled-bunch instabilities can be reduced with a proper design of the vacuum chamber and of the various installed objects [3].

1.2.4 Ion instabilities

Molecules of gas in the vacuum chamber can be ionized by collision with the electron beam. The generated positive ions remain trapped in the negative electric potential and produce electron–ion coherent oscillations [4].

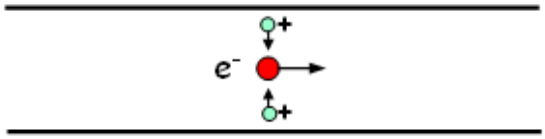


Fig. 4: Interaction of positive ions with electron bunches

A common practice currently used to counteract ion instabilities is so-called 'ion cleaning', obtained by operating the machine with a gap in the bunch train.

1.2.5 Generic countermeasures

Besides passive cures that reduce unwanted effects by acting directly on the sources of instability, it is possible to counteract the onset of coupled-bunch instabilities by increasing the Landau damping, which can be obtained by enhancing the betatron/synchrotron tune spread. This can be accomplished, for example, by acting on the chromaticity, by the use of higher harmonic RF cavities producing bunch lengthening [5] or by modulating the amplitude or phase of the accelerating radio frequency [6]. When all of the above-mentioned techniques are insufficient to reduce or eliminate coupled-bunch instabilities, active feedback mechanisms are required [7–12].

1.3 Equation of motion of particle bunches

1.3.1 Equation of motion of charged particles

The motion of a charged particle in a storage ring can be described with the harmonic oscillator analogy following the equation of motion

$$\ddot{x}(t)+2D\,\dot{x}(t)+\omega^2 x(t)=0 \tag{1}$$

where x is the oscillation coordinate (transverse or longitudinal displacement), D the natural damping, and ω the betatron/synchrotron tune frequency given by the $\nu\omega_0$ where ν is the betatron/synchrotron tune, and ω_0 the revolution frequency. In the transverse planes, oscillations are periodic horizontal or vertical displacements from an ideal trajectory, while in the longitudinal plane they are periodic 'timing' differences with respect to a reference 'stable particle' running in the accelerator at constant speed. Longitudinal oscillations are often referred to as 'phase oscillations'.

If $\omega >> D$, an approximated solution of the differential equation Eq. (1) is a damped sinusoidal oscillation

$$x(t)= k\, \mathrm{e}^{-\frac{t}{\tau_D}} \sin(\omega t+\phi) \tag{2}$$

where $\tau_D = 1/D$ is the 'damping time constant' (D is the damping rate).

Any excited oscillation (e.g. by quantum excitation) is damped by the natural damping (e.g. due to radiation of synchrotron light). The oscillation of individual particles is uncorrelated and manifests itself as an emittance growth.

1.3.2 Coherent bunch oscillations

Coupling with other bunches through the interaction with surrounding metallic structures adds a 'driving force' term $F(t)$ to the equation of motion

$$\ddot{x}(t)+2D\,\dot{x}(t)+\omega^2 x(t)=F(t)\ . \tag{3}$$

Under given conditions, the oscillation of individual particles becomes correlated and the centroid of the bunch oscillates coherently with the other bunches giving rise to coherent bunch (coupled-bunch) oscillations. Each bunch oscillates according to the equation of motion:

$$\ddot{x}(t)+2(D-G)\,\dot{x}(t)+\omega^2 x(t)=0 \tag{4}$$

where $\tau_G = 1/G$ is the 'growth time constant' (G is called the growth rate). Similarly to Eq. (2), if $\omega >> (D - G)$, an approximated solution of Eq. (4) is

$$x(t) = k\,\mathrm{e}^{-\frac{t}{\tau}} \sin(\omega t + \phi) \tag{5}$$

where $\frac{1}{\tau} = \frac{1}{\tau_D} - \frac{1}{\tau_G}$. If $D > G$ the oscillation amplitude decays; if $D < G$ it grows exponentially.

Since G is proportional to the stored beam current, if the latter is lower than a given current threshold the beam remains stable; if higher a coupled-bunch instability is excited.

1.3.3 Feedback action

The feedback action adds a damping term D_{fb} to the equation of motion

$$\ddot{x}(t)+2(D-G+D_{\mathrm{fb}})\,\dot{x}(t)+\omega^2 x(t)=0\ . \tag{6}$$

In order to damp the oscillation D_{fb} must be such that $D - G + D_{\mathrm{fb}} > 0$.

A multi-bunch feedback system detects an instability by means of one or more BPMs and acts back on the beam by applying electromagnetic 'kicks' to the bunches.

Since the bunch oscillation is sinusoidal, the turn-by-turn position of a bunch measured at a given location is a sampled sinusoid. From Eq. (6), in order to introduce damping, the force applied by the feedback must be proportional to the derivative of the bunch oscillation. Consequently, the kick signal applied by the actuator to each bunch can be generated by shifting by $\pi/2$ the 'sampled' signal of the position of the same bunch when it passes through the 'kicker'. This is the basic principle of bunch-by-bunch feedback systems that will be treated in more detail in Section 2.

1.4 Multi-bunch modes

When a coupled-bunch instability is established, although each bunch oscillates at the tune frequency, there can be different modes of oscillation, called multi-bunch modes, depending on how each bunch oscillates with respect to the other bunches [9, 13].

Let us consider M bunches equally spaced around the ring. Each multi-bunch mode is characterized by a bunch-to-bunch phase difference of

$$\Delta\Phi = m\,\frac{2\pi}{M} \tag{7}$$

where m is the multi-bunch mode number ($m = 0, 1, ..., M - 1$). Each multi-bunch mode is associated to a characteristic set of frequencies

$$\omega = p\,M\,\omega_0 \pm (m+\nu)\,\omega_0 \tag{8}$$

where p is an integer number ($-\infty < p < \infty$), ω_0 is the revolution frequency, $\omega_{rf} = M\omega_0$ is the RF frequency (bunch repetition frequency), and ν is the tune. In turn, there are two sidebands at $\pm(m + \nu)\omega_0$ for each multiple of the RF frequency.

In the following sections we shall give some examples of transverse oscillation modes and examine the associated spectra.

1.4.1 Examples of transverse multi-bunch modes and their spectra

1.4.1.1 One single stable bunch

In this example a single bunch of particles is stored in the ring. Figure 5 shows the linear trajectory of the bunch during its run along the ring (dotted line in Fig. 5(a)) and the time domain signal captured by a pickup (Fig. 5(b)). The location of the pickup is identified in the figure by triangles placed at longitudinal coordinate 0.

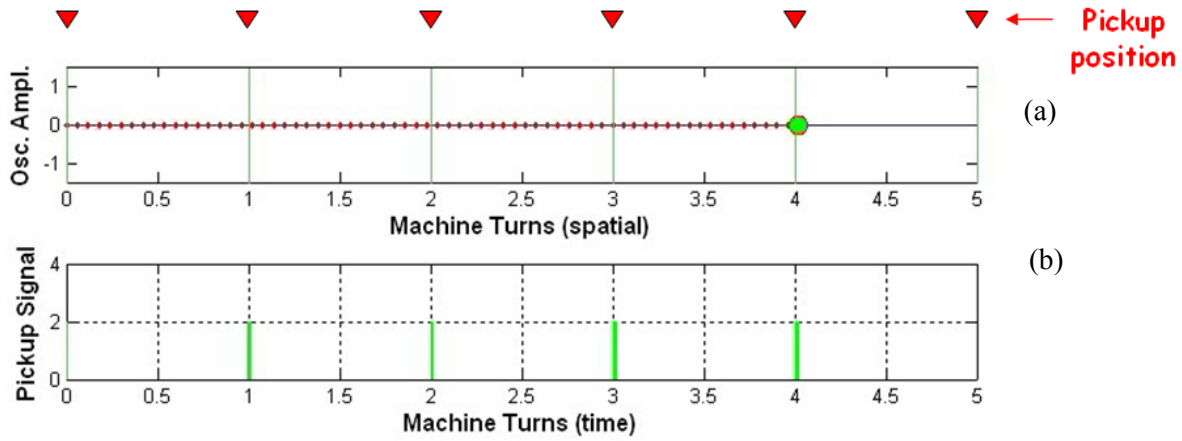


Fig. 5: (a) Motion of a single stable bunch; (b) time domain signal captured by a pickup

Every time the bunch passes through the pickup, an electrical pulse with constant amplitude is generated (vertical bars). If we think of it as a Dirac impulse, the spectrum of the pickup signal (Fig. 6) is a repetition of frequency lines at a multiple of the revolution frequency $p\omega_0$ for $-\infty < p < \infty$.

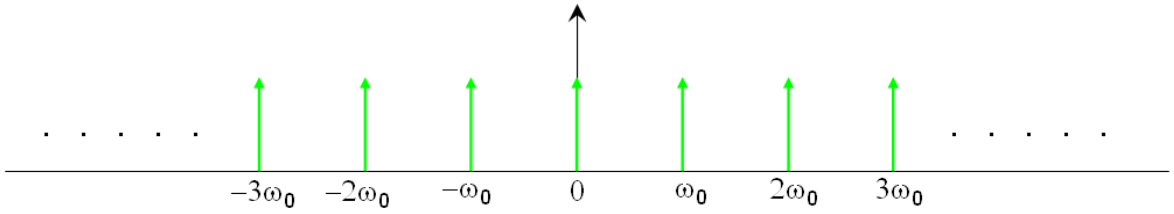


Fig. 6: Spectrum of the signal captured by the pickup with a single stable bunch

1.4.1.2 One single unstable bunch

In this case a single bunch stored in the ring oscillates at the tune frequency $\nu\omega_0$. For simplicity, in this example we consider a tune value lower than unity, $\nu = 0.25$. With a single bunch $M = 1$, only mode number 0 exists. The bunch performs one complete oscillation in four machine turns and the pickup signal is a sequence of pulses modulated in amplitude with frequency $\nu\omega_0$ (Fig. 7).

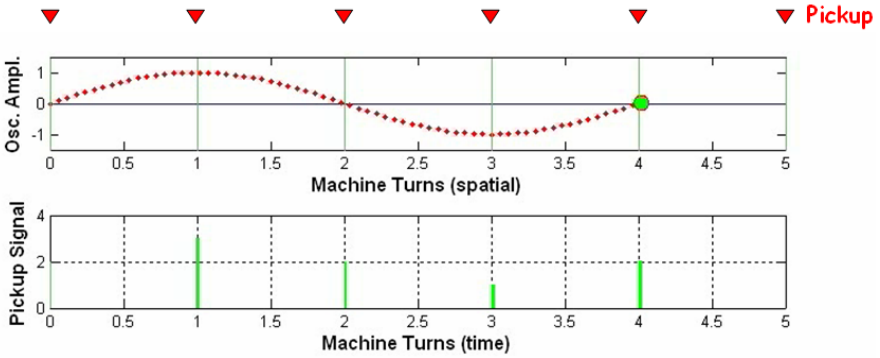


Fig. 7: (a) A single bunch oscillating at the tune frequency; (b) corresponding pickup signal

In the spectrum two sidebands at $\pm\nu\omega_0$ appear at each of the revolution harmonics (Fig. 8).

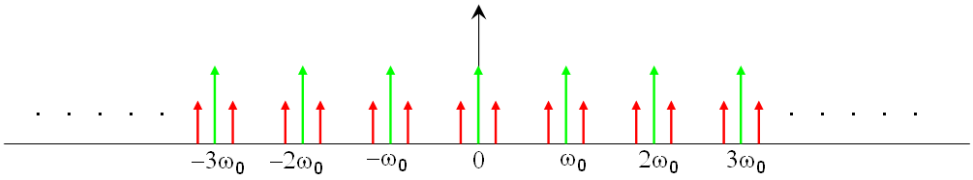


Fig. 8: Spectrum of the pickup signal with a single unstable bunch

1.4.1.3 Ten stable bunches

Let us consider now ten identical, equally spaced, stable bunches ($M = 10$). The pickup signal is a sequence of pulses at the bunch repetition frequency (RF frequency) $\omega_{rf} = 10\ \omega_0$.

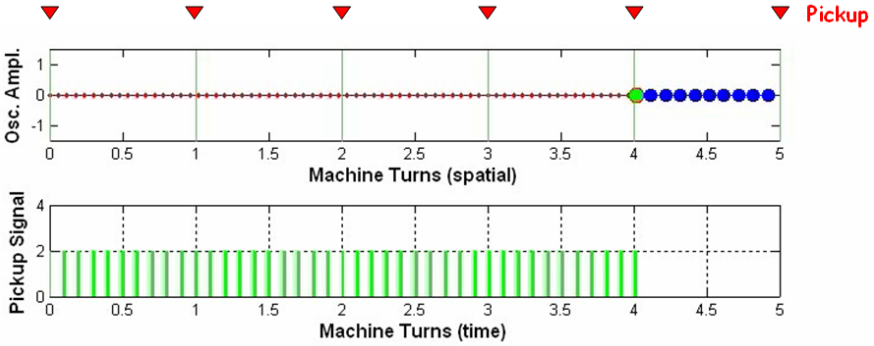


Fig. 9: Ten equally spaced stable bunches

The spectrum is a repetition of frequency lines at multiples of the bunch repetition frequency.

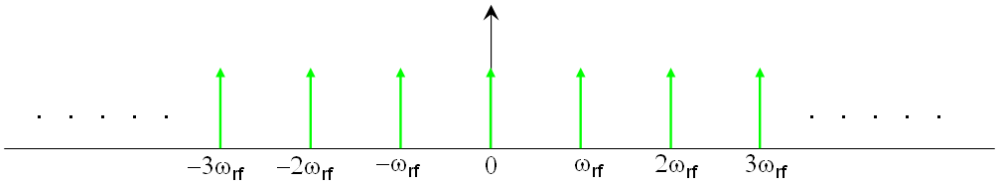


Fig. 10: Spectrum of the pickup signal with ten stable bunches

1.4.1.4 Ten unstable bunches, mode #0

In the case where the bunches oscillate at the tune frequency $\nu\omega_0$ with $\nu = 0.25$, since $M = 10$ there are 10 possible modes of oscillation characterized by a phase difference between adjacent bunches of $\Delta\Phi = m\,\frac{2\pi}{10}$ with $m = 0, 1, ..., 9$.

If $m = 0$ (mode number 0) then $\Delta\Phi = 0$, namely all the bunches oscillate with the same phase (Fig. 11). The pickup signal is a sequence of pulses at the bunch repetition frequency modulated in amplitude with frequency $\nu\omega_0$.

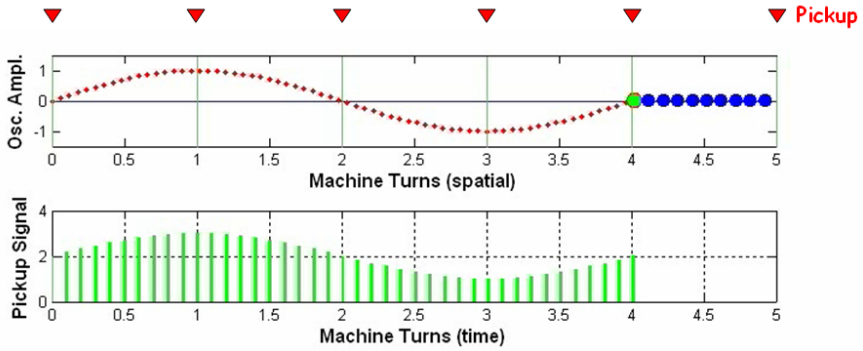


Fig. 11: Ten identical bunches oscillating in phase at the tune frequency. The dotted line is the trajectory of one of them.

The spectrum is a repetition of frequency lines at multiples of the bunch repetition frequency with sidebands at $\pm\nu\omega_0$: $\omega = p\omega_{rf} \pm \nu\omega_0$ with $-\infty < p < \infty$ (Fig. 12).

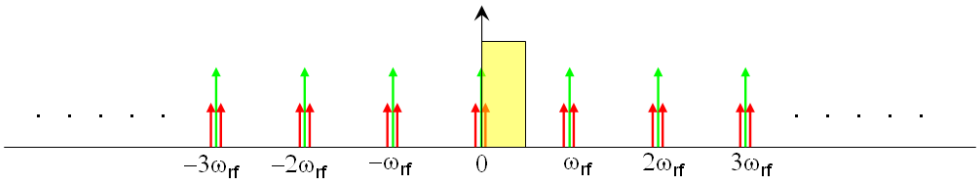


Fig. 12: Spectrum of ten unstable bunches: mode #0

Since the spectrum is periodic and the mode appears twice (lower and upper sideband) in a ω_{rf} frequency span centred on every harmonic of ω_{rf}, we can limit the spectrum analysis to a $0–\omega_{rf}/2$ frequency range. Figure 13 is an enlargement of the frequency portion denoted by the rectangular box in Fig. 12.

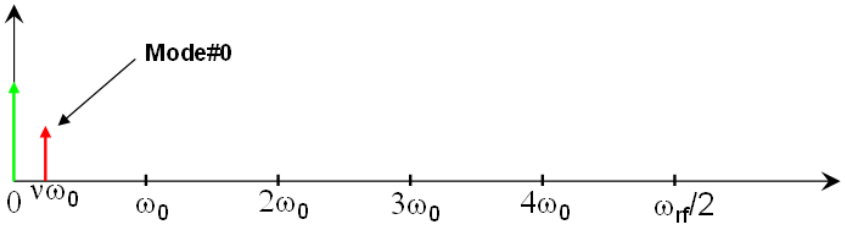


Fig. 13: Spectrum of mode #0 restricted to a $\omega_{rf}/2$ wide frequency range

1.4.1.5 Ten unstable bunches, mode #1

In mode #1, the phase difference of the bunches oscillating at the betatron frequency is $\Delta\Phi = 2\pi/10$ and the spectrum is characterized by frequency lines at $\omega = p\omega_{rf} \pm (\nu + 1)\omega_0$, with $-\infty < p < \infty$ (Figs. 14 and 15).

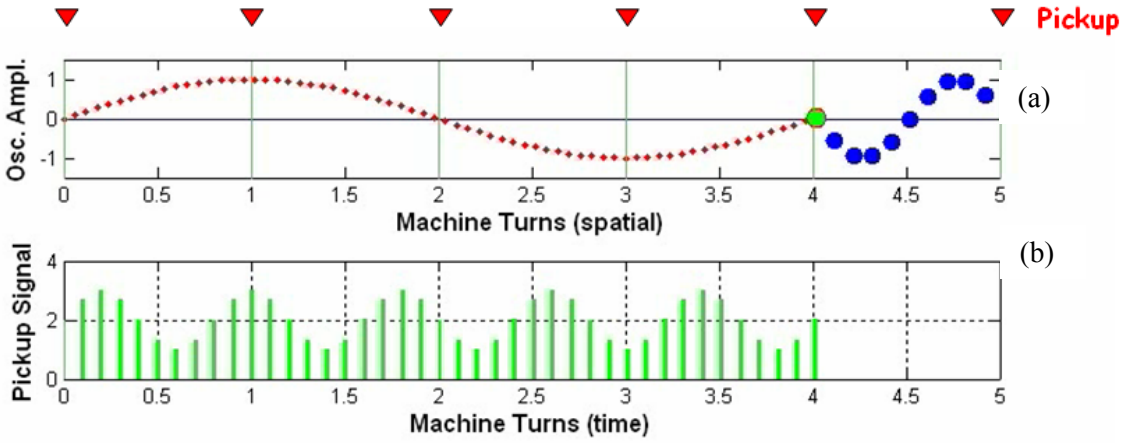


Fig. 14: Ten unstable bunches: mode #1

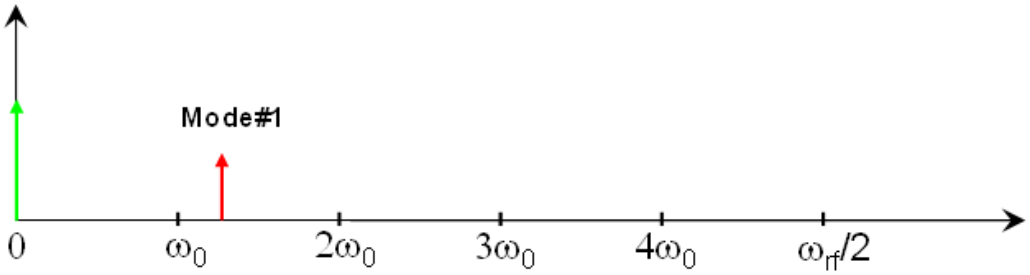


Fig. 15: Spectrum of mode #1

1.4.1.6 Ten unstable bunches, modes #2 to #9

Figures 16–23 show the motion, time domain signal, and spectrum of multi-bunch modes #2 to #9.

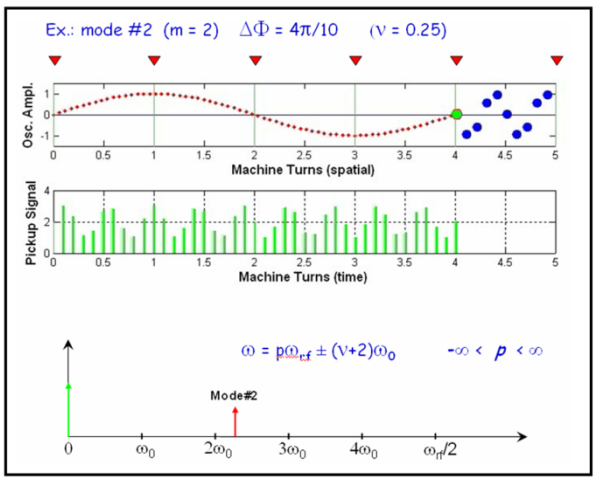


Fig. 16: Mode #2

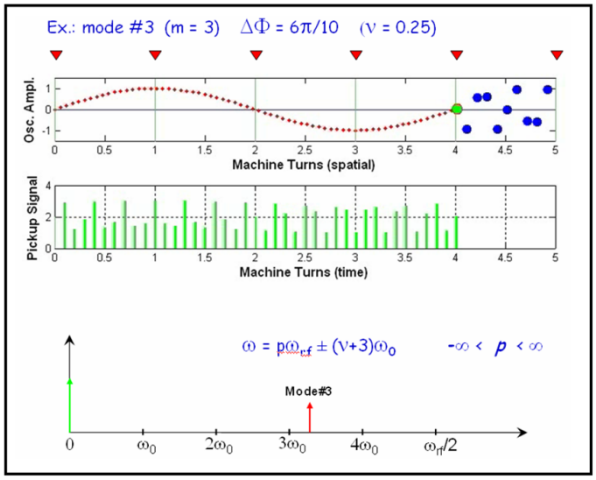


Fig. 17: Mode #3

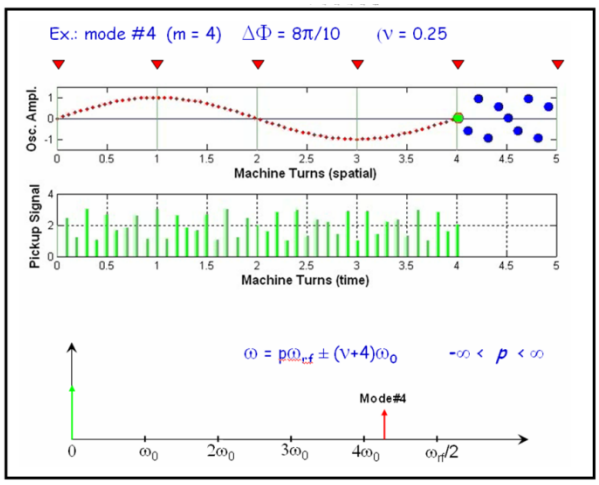


Fig. 18: Mode #4

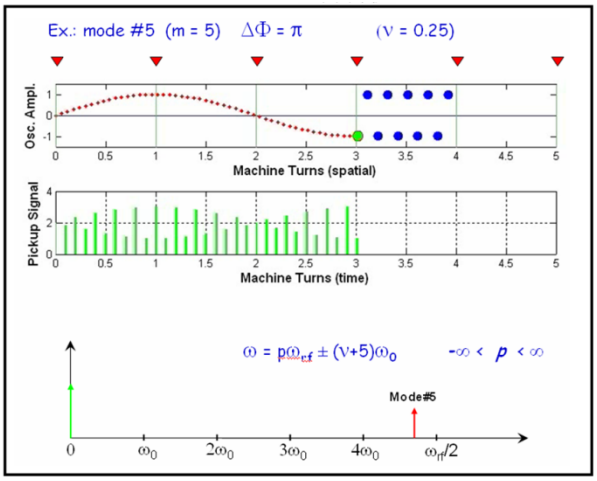


Fig. 19: Mode #5

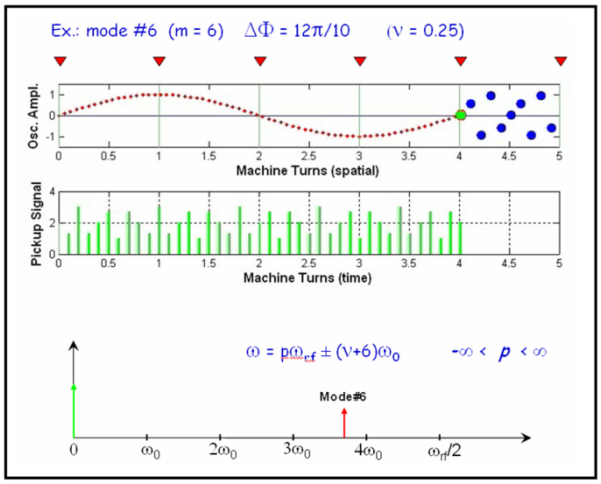


Fig. 20: Mode #6

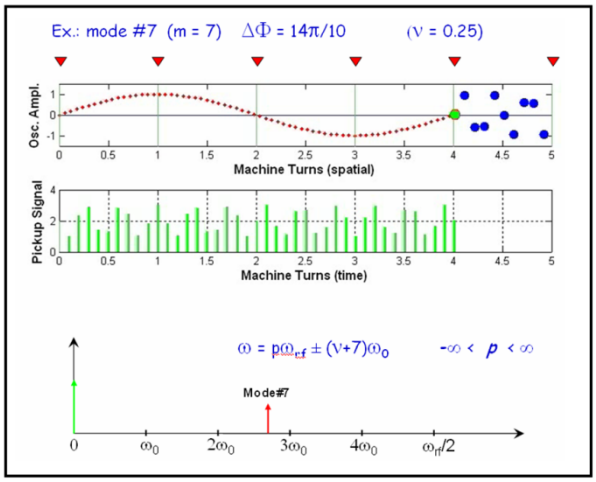


Fig. 21: Mode #7

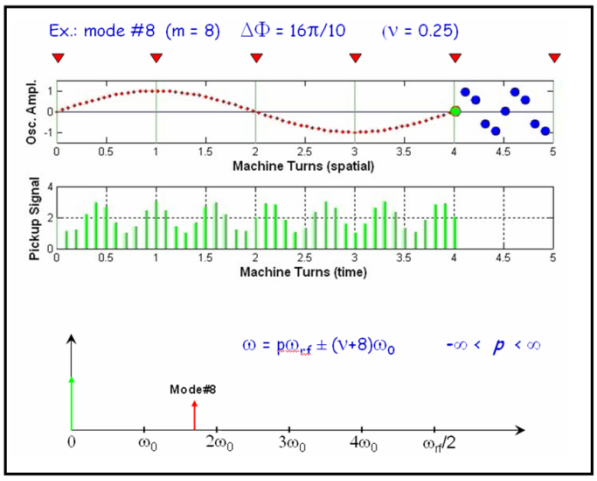


Fig. 22: Mode #8

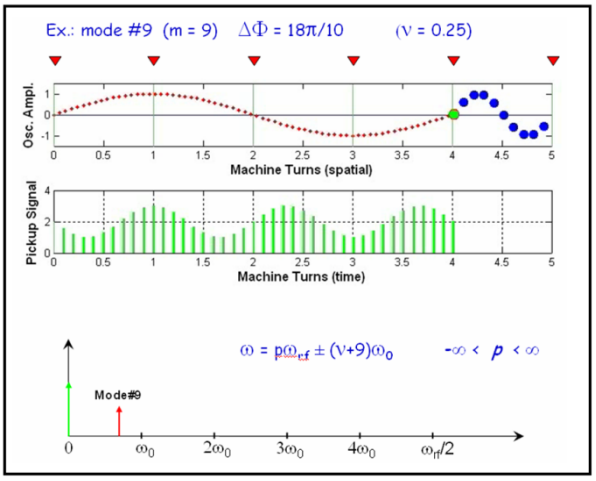


Fig. 23: Mode #9

1.4.2 Uneven filling

If the bunches do not have the same charge, i.e. the buckets are not equally filled (uneven filling), the spectrum also has frequency components at the revolution harmonics (multiples of ω_0) (Fig. 24). The amplitude of the revolution harmonics depends on the filling pattern of one machine turn.

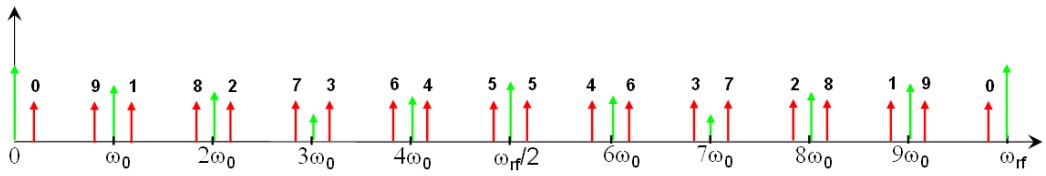


Fig. 24: Spectrum of a beam with uneven filled buckets

1.4.3 Longitudinal multi-bunch modes

Compared with transverse oscillations where multi-bunch instabilities produce amplitude modulation of the stable beam pickup signal, longitudinal multi-bunch modes produce phase modulation. Components at $\pm\nu\omega_0$, $\pm2\nu\omega_0$, $\pm3\nu\omega_0$… appear aside the revolution harmonics. Their amplitude depends on the depth of the phase modulation, namely the amplitude of the instability, according to Bessel series expansion (Fig. 25).

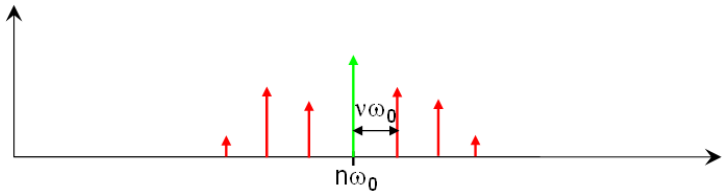


Fig. 25: Spectrum of a longitudinally unstable beam with components near to a revolution harmonic at $\pm\nu\omega_0$, $\pm2\nu\omega_0$, $\pm3\nu\omega_0$….

1.4.4 Coupled-bunch instability

Starting from the spectral representation of all potential multi-bunch modes, one of them can become unstable if one of its sidebands overlaps, for example, with the frequency response of a HOM. The HOM couples with the sideband giving rise to a coupled-bunch instability, with a consequent increase of the mode amplitude (Fig. 26).

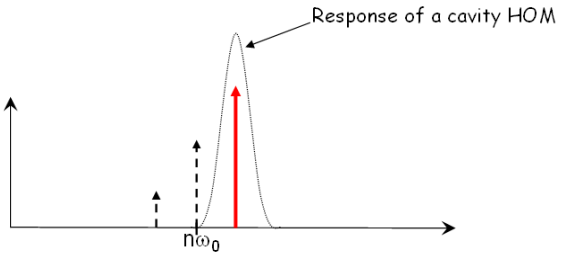


Fig. 26: Coupling of a cavity HOM with the sideband corresponding to a given multi-bunch mode

More generally, in order to characterize the interaction of the beam with the machine structures, the 'impedance model' is often used. The impedance is the frequency spectrum of the forces generated by wake fields acting back on the beam. It summarizes the electromagnetic effects of the machine components (vacuum chamber, cavities, bellows, BPMs, etc.) on the beam. The shape and amplitude of the machine impedance give us information about which multi-bunch modes could be excited and the strength of the excitation (growth time constant).

1.4.5 Effects of coupled-bunch instabilities

Besides the risk of beam loss or the limitation of the maximum stored current, the onset of instabilities produces an increase of the transverse beam size (also in the case of longitudinal instabilities because of dispersion) and thus of the effective emittance. This is detrimental for the beam quality since it lowers the brightness of the produced photon beams in synchrotron light sources and the luminosity in circular colliders. Besides these negative effects, however, the presence of multi-bunch instabilities is often

associated with an increase of the beam lifetime. Saturation effects due to Landau damping, in fact, produce dilution of particles inside the bunches with a consequent decrease of Touschek scattering.

1.4.6 Real examples of multi-bunch modes and their measurement

Let us consider two real examples of transverse and longitudinal instabilities. We refer to the Elettra synchrotron storage ring. The RF frequency is 499.654 MHz; having a harmonic number of 432, the revolution frequency is 1.157 MHz. The horizontal tune ν_{hor} is 12.30 (fractional tune frequency = 345 kHz), the vertical tune ν_{ver} is 8.17 (fractional tune frequency = 200 kHz), and the longitudinal tune ν_{long} is 0.0076 (tune frequency = 8.8 kHz). Keeping in mind Eq. (8) ($\omega = pM\omega_0 \pm (m+i)\omega_0$) let us analyse two beam spectra measured with a spectrum analyser connected to a strip line pickup.

1.4.6.1 Example 1

In this example we analyse a spectral line at 512.185 MHz (Fig. 27). It is a lower sideband of the second harmonic of the RF frequency, 200 kHz apart from the 443rd revolution harmonic at 512.355 MHz, denoting that it is a vertical instability. From Eq. (8) it can be easily calculated that it corresponds to the vertical multi-bunch mode #413.

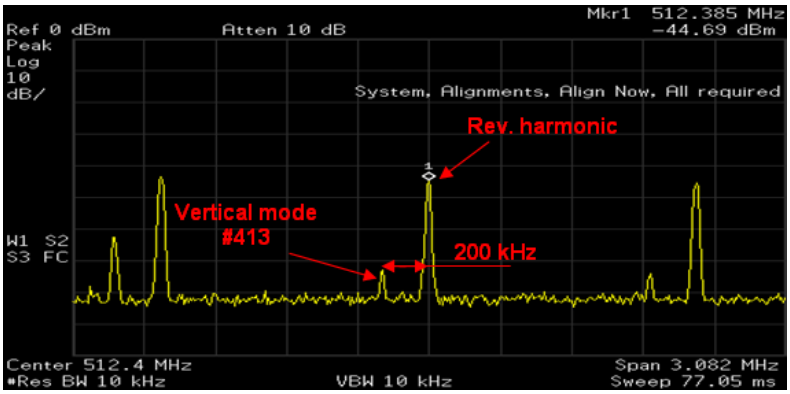


Fig. 27: Example of beam spectrum in the presence of vertical instabilities measured by a spectrum analyser

1.4.6.2 Example 2

In this case we analyse a spectral line at 604.914 MHz (Fig. 28). It is an upper sideband of the RF frequency, 8.8 kHz apart from the 523rd revolution harmonic, denoting a longitudinal instability. From Eq. (8) we can calculate that it corresponds to the longitudinal multi-bunch mode #91.

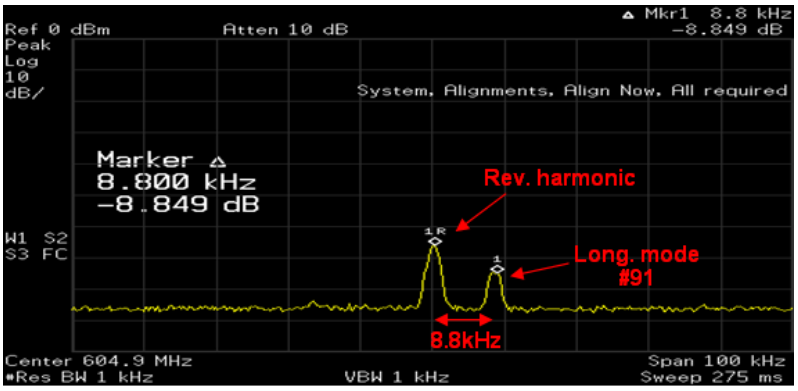


Fig. 28: Example of beam spectrum in the presence of a longitudinal instability measured by a spectrum analyser

2 Multi-bunch feedback systems

2.1 Introduction

As already seen in Section 1.3.3, a multi-bunch feedback system detects multi-bunch instabilities using one or more BPMs and acts back on the beam through an electromagnetic actuator called a 'kicker'. A simplified block diagram of a feedback system is depicted in Fig. 29.

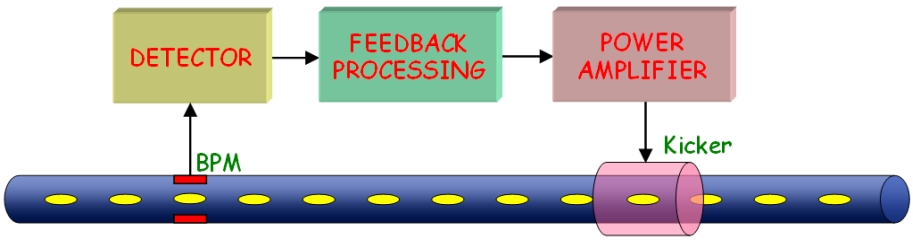


Fig. 29: Block diagram of a multi-bunch feedback system

BPMs and detectors measure the oscillation of the bunches and the processing unit generates the correction signal, which is amplified by the power amplifier and sent to the kicker.

2.2 Types of feedback systems

Depending on how a feedback system detects and controls instabilities, there are two main feedback categories: 'mode-by-mode' and 'bunch-by-bunch' feedback. The former are often referred to as 'frequency domain' feedback, the latter as 'time domain' feedback [8]. Moreover, the feedback implementation can be analogue or digital, depending on whether the signal is handled in the analogue domain or sampled and processed digitally. Despite the fact that most of the latest feedback systems are digital bunch-by-bunch systems, analogue or mode-by-mode implementations are still in use in a number of accelerators.

In the following sections the basic principles of the above-mentioned types of feedback will be illustrated; they are part of the history of multi-bunch feedback systems over a period of at least thirty years.

2.2.1 Mode-by-mode feedback

Mode-by-mode (frequency domain) feedback acts separately on each of the controlled unstable modes [14]. A simplified block diagram is illustrated in Fig. 30.

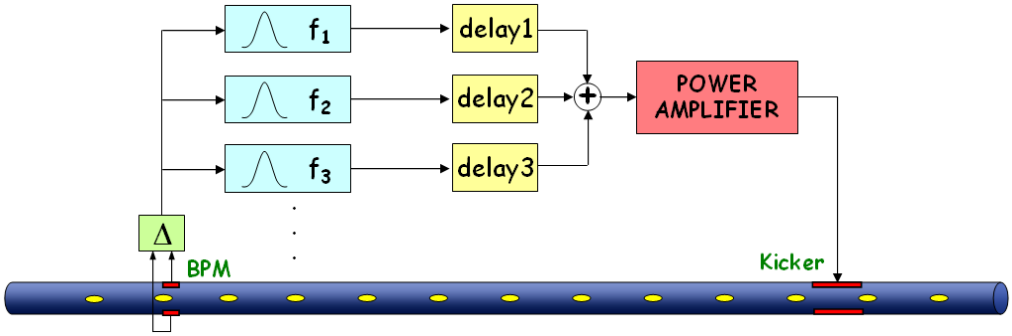


Fig. 30: Block diagram of a mode-by-mode feedback system

An appropriate electronic system generates the position error signal by combining the signals of the BPM buttons. A number of processing chains, each dedicated to one of the controlled modes, work in parallel. These are composed of a narrow band-pass filter that selects the proper frequency and an

adjustable delay line to phase-shift the signal in order to produce a negative feedback. The channels are then combined, amplified, and sent to the kicker.

2.2.2 Bunch-by-bunch feedback

A bunch-by-bunch (time domain) feedback individually steers each bunch by applying 'small' electromagnetic kicks every time the bunch passes through the kicker. The correction signal for a given bunch is generated based on the motion of that bunch. The result is a damped oscillation lasting several turns. Bunch-by-bunch feedback can be realized as shown in Fig. 31.

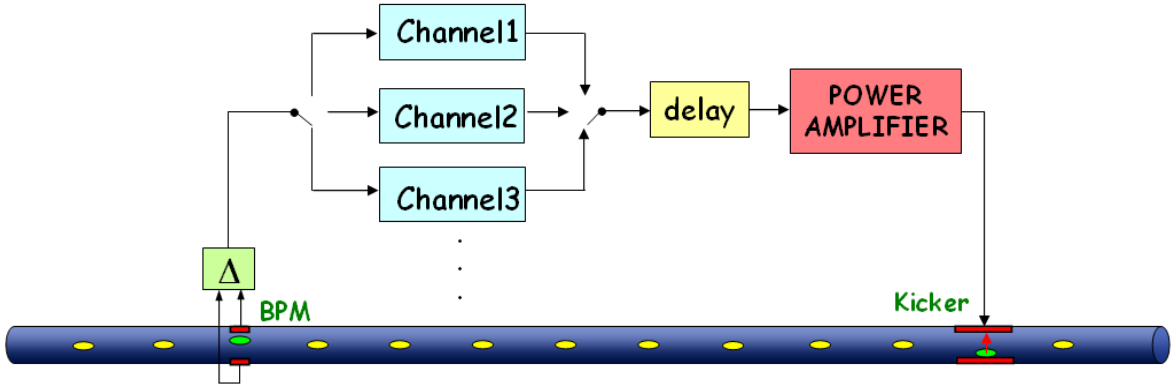


Fig. 31: Block diagram of a bunch-by-bunch feedback system based on a time division scheme

There are as many processing channels as the number of bunches. A demultiplexer provides each channel with the samples for the corresponding bunch, while a multiplexer recombines the samples at the end of the channels. Following the principle explained in Section 1.3.3, the main task of each channel is to properly phase-shift the sampled position signal. A delay line assures that each bunch is kicked with its own correction samples.

Every bunch is measured and corrected at every machine turn but, owing to the delay of the feedback chain, the correction kick corresponding to a given measurement is applied to the bunch one or more turns later.

An important principle behind bunch-by-bunch feedback is that it is possible to stabilize all coupled-bunch modes by damping the oscillation of each bunch.

2.2.3 Analogue bunch-by-bunch feedback

In the following sections we shall see some real analogue implementations of transverse bunch-by-bunch feedback, starting from a basic very simple architecture to a system that has been deployed in many accelerators.

2.2.3.1 Feedback using one BPM

In bunch-by-bunch feedback the correction signal applied to a given bunch must be proportional to the derivative of the bunch oscillation at the kicker. The correction signal is therefore a sampled sinusoid shifted by $\pi/2$ with respect to the oscillation of the bunch when it passes through the kicker. Starting from this principle, a very simple way to obtain such a phase shift is to have the BPM and kicker placed at points in the ring with the appropriate betatron phase difference and to use the detected signal to directly feed the kicker through an amplifier. The system is shown in Fig. 32.

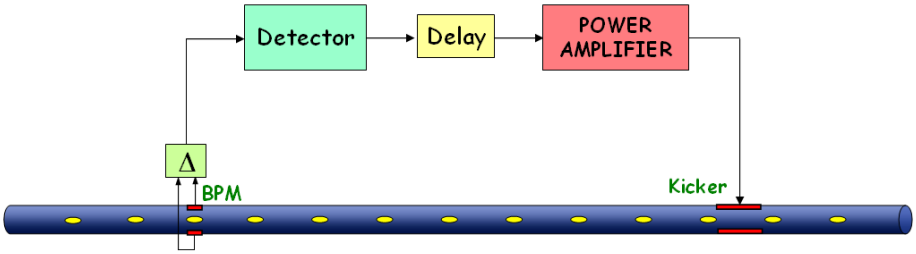


Fig. 32: Block diagram of a simple analogue bunch-by-bunch transverse feedback system

The detector down-converts the high frequency (typically a multiple of the bunch frequency f_{rf}) BPM signal into baseband (range 0–f_{rf}/2). The delay line assures that the signal for a given bunch passing through the feedback chain arrives at the kicker when, after one machine turn, the same bunch again passes through it.

2.2.3.2 Feedback using two BPMs

By using two BPMs, provided that they are separated by about π/2 in betatron phase, it is possible to place them in any ring position with respect to the kicker (Fig. 33).

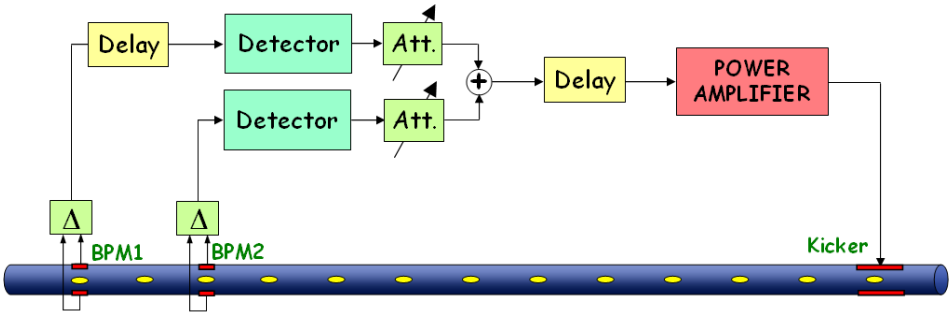


Fig. 33: Analogue bunch-by-bunch transverse feedback using two BPMs

As shown in Fig. 34, if we opportunely combine the two detected signals using variable attenuators, it is possible to vary the phase of the resulting bunch correction signal. The attenuators have to be adjusted so that the bunch oscillation and the corresponding kick signal are in quadrature.

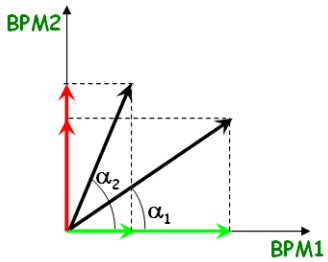


Fig. 34: If the two BPM bunch signals are in quadrature, the phase of the sum signal can be varied by adjusting the variable attenuators.

2.2.3.3 Suppression of revolution harmonics

The revolution harmonics (frequencies at multiples of ω_0) are useless components that can be removed from the correction signal in order not to saturate the RF amplifier. A special module implementing notch filters at the harmonics of ω_0 can be placed before the amplifier to eliminate these components. This operation is also called 'stable beam' rejection. We shall see below a number of possible implementations of this module that are also used in digital feedback systems.

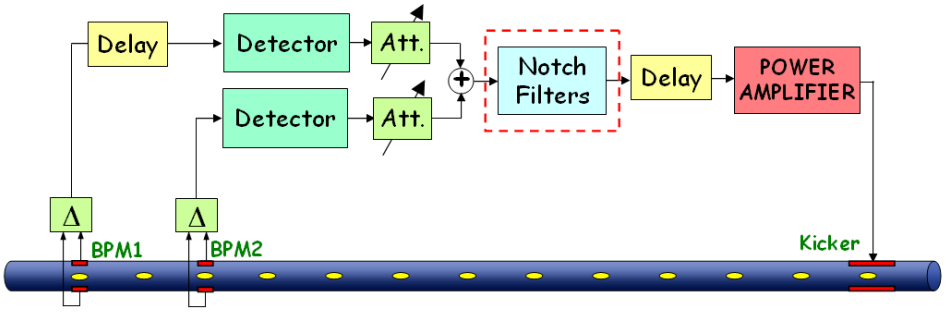


Fig. 35: Analogue bunch-by-bunch transverse feedback with suppression of revolution harmonics

Architectures similar to the above examples have been used to build the ALS [15], Bessy II [16], PLS [17] and ANKA [18] transverse multi-bunch feedback systems.

2.2.4 Digital bunch-by-bunch feedback

Commercially available fast digital electronics (analogue-to-digital converters (ADCs), digital-to-analogue converters (DACs), digital signal processors (DSPs), field programmable gate arrays (FPGAs), etc.) allow one to build digital feedback systems where the signal of the bunch positions is digitized, digitally processed to calculate the corrections, and re-converted to an analogue signal. A block diagram of a generic digital feedback is illustrated in Fig. 36. Thanks to the capabilities offered by digital signal processing, digital feedback usually makes use of only one BPM and a kicker mounted in any ring position.

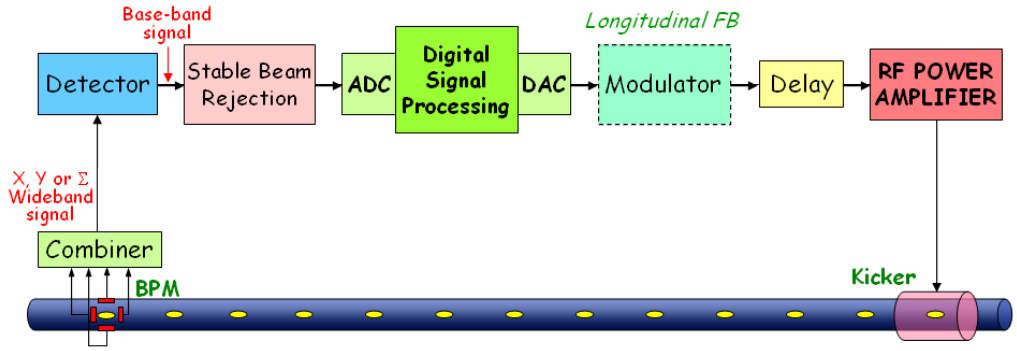


Fig. 36: Block diagram of a digital bunch-by-bunch feedback system

Starting from the BPM button signals, the combiner generates the wideband horizontal (X), vertical (Y) or sum (Σ) signal, which is then demodulated to baseband by the detector (also called 'RF front-end'). Some feedback implementations feature a stable beam rejection module that removes useless stable beam components from the signal, which is eventually digitized, processed, and re-converted to analogue by the digital processor. While in transverse feedback systems the amplifier and kicker operate in baseband, longitudinal feedback requires a modulator that translates the correction signal to the kicker operation frequency. The delay line adjusts the timing of the signal to match the bunch arrival time.

2.2.5 Digital vs. analogue feedback

The following list summarizes the main advantages of digital feedback systems:

i) reproducibility: when the signal is digitized it is not subject to temperature/environment changes or ageing;

ii) programmability: the implementation of processing functionalities is usually made using DSPs or FPGAs, which are programmable via software/firmware;

iii) performance: digital controllers feature superior processing capabilities with the possibility to implement sophisticated control algorithms not feasible in analogue;

iv) additional features: the possibility to combine basic control algorithms and additional useful features like signal conditioning, saturation control, down-sampling, etc.;

v) implementation of diagnostic tools, used for both feedback commissioning and machine physics studies (see Section 6);

vi) easier and more efficient integration of the feedback in the accelerator control system, important for feedback set-up and tuning, fast data acquisition, easy and automated operations, etc.

Among the disadvantages is the higher delay of the feedback chain (due to ADC, digital processing, and DAC) with respect to analogue feedback, although with the use of FPGAs this delay is reduced to acceptable values.

3 Components of feedback systems

In the following sections a more detailed description of components usually employed in (digital) feedback systems will be given, with practical issues, examples, and real implementations.

3.1 BPM and combiner

The four signals from a standard four-button BPM can be opportunely combined to obtain the wideband X, Y and Σ signals used by the horizontal, vertical, and longitudinal feedback systems, respectively. Figure 37 shows an example of how this can be made.

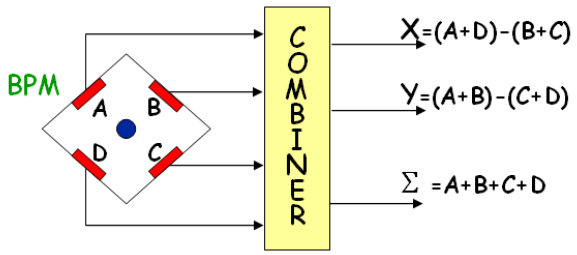


Fig. 37: BPM and combiner are used to provide X, Y, and Σ wideband signals

Any $f_{rf}/2$ portion of the beam spectrum contains the information for all potential multi-bunch modes and can be used to detect instabilities and measure their amplitude. Usually the BPM and combiner work in a frequency range around a multiple of f_{rf} (Fig. 38), where the amplitude of the overall frequency response of the BPM and cables is at a maximum. Moreover, a higher f_{rf} harmonic is preferred for the longitudinal feedback because of the better sensitivity of the phase detection system.

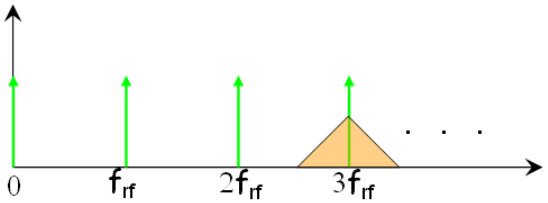


Fig. 38: Example of wideband signal centred on the third harmonic of the RF frequency

3.2 Detector (RF front-end)

3.2.1 Transverse case

The detector translates the wideband signal to baseband (0–f_{rf}/2 range): the operation is an amplitude demodulation (Fig. 39).

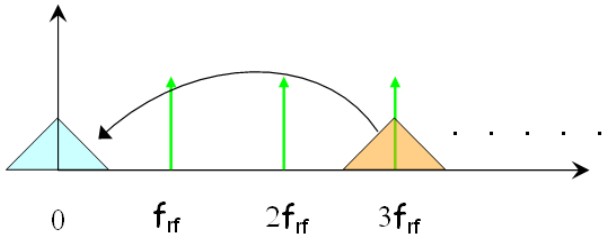


Fig. 39: The amplitude demodulation translates the wideband signal to baseband

A possible implementation of amplitude demodulation, widely used in telecommunication technology, is the 'heterodyne' technique illustrated in Fig. 40. The wideband signal is first band-pass filtered to select the desired frequency range and then mixed with the 'local oscillator' signal, which must be synchronous with the former. The local oscillator signal can be derived from the RF by multiplying its frequency by an integer number corresponding to the chosen harmonic of f_{rf}. The resulting signal is eventually low-pass filtered in the 0–f_{rf}/2 frequency range.

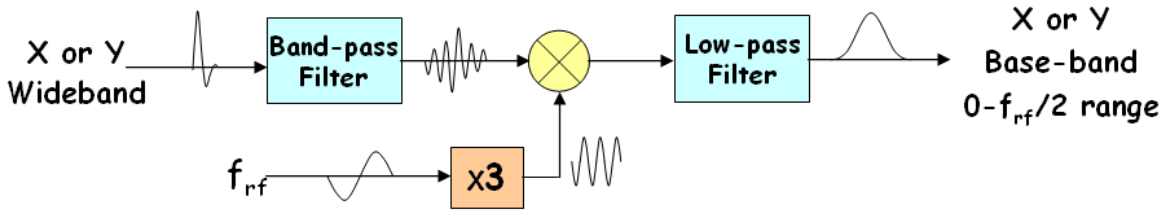


Fig. 40: Heterodyne amplitude demodulation using the third harmonic of the RF frequency

3.2.2 Longitudinal case

The wideband sum signal (Σ) contains only information about the phase (longitudinal position) of the bunches, since the sum of the four button signals has almost constant amplitude.

The baseband phase error signal (0–f_{rf}/2 range) can be generated by phase demodulation, which can be carried out with the same heterodyne technique used for amplitude demodulation but using a 'local oscillator' in quadrature with the wideband signal, namely shifted in phase by π/2 (Fig. 41).

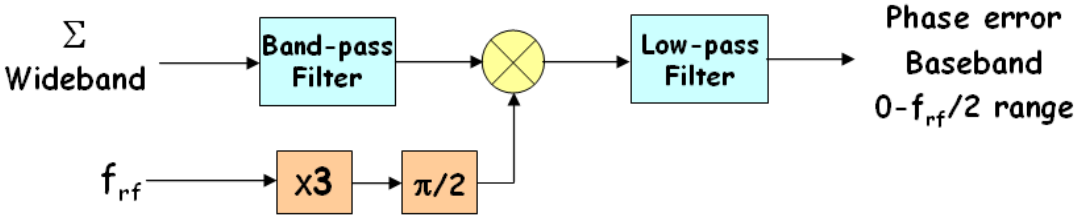


Fig. 41: Phase demodulation using the heterodyne technique

3.2.3 Amplitude and phase demodulation

The following are simple mathematic demonstrations that can help in understanding heterodyne amplitude and phase demodulations.

3.2.3.1 Amplitude demodulation

Let us consider a sinusoidal carrier $\sin(3\omega_{rf}t)$ modulated in amplitude by a signal $A(t)$. If we multiply it by the local oscillator signal $\sin(3\omega_{rf}t)$ (mixing operation) we obtain two frequency components. The component at the higher frequency can be filtered out by a low-pass filter leaving a signal proportional to the modulating signal $A(t)$

$$A(t)\sin(3\omega_{rf}t)\sin(3\omega_{rf}t) \propto A(t)(\cos(0) - \cos(6\omega_{rf}t)) \approx A(t)\,. \tag{9}$$

3.2.3.2 Phase demodulation

In this case the sinusoidal carrier is phase modulated by a signal $\varphi(t)$. If the latter is small, by multiplying the modulated signal by a quadrature-phase local oscillator signal $\cos(3\omega_{rf}t)$, after low-pass filtering we obtain the modulating signal

$$\sin(3\omega_{rf}t + \varphi(t))\cos(3\omega_{rf}t) \propto \sin(6\omega_{rf}t + \varphi(t)) + \sin(\varphi(t)) \approx \varphi(t)\,. \tag{10}$$

3.2.4 Time domain considerations

The baseband signal can be seen as a sequence of 'pulses', each with amplitude proportional to the position error (X, Y or Φ) and to the charge of the corresponding bunch. By sampling this signal with an A/D converter synchronous with the bunch frequency, one can measure X, Y or Φ (Fig. 42).

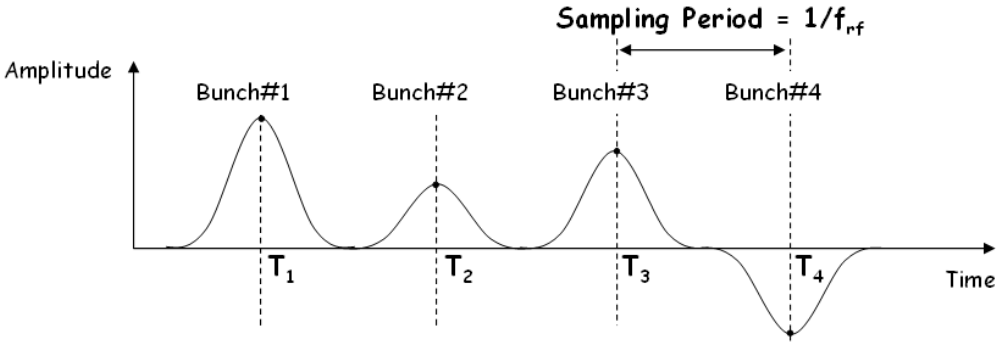


Fig. 42: Detector output signal

Note that the multi-bunch-mode number $M/2$, which is the one with the highest frequency (close to $f_{rf}/2$) in baseband, is characterized by a sequence of pulses with almost the same amplitude but alternating signs.

The design of the detector band-pass and low-pass filters affects the shape of the pulses that should have maximum flatness on the top (to reduce clock jitter noise) and minimum overlap with adjacent pulses (to reduce cross-talk between bunches) (Fig. 43). To achieve this, Bessel filters with linear phase response or a special filter called a 'comb generator' [19] can be used.

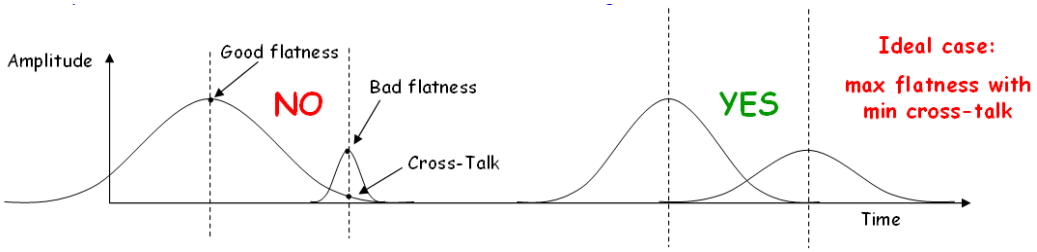


Fig. 43: Examples of badly and well-filtered detector signals

3.3 Rejection of stable beam signal

The amplitude of the pulses in the baseband signal can have a constant offset called the 'stable beam signal'. In the transverse plane this offset can be due to an off-centre beam in the BPM or to unbalanced

BPM electrodes or cables. In the longitudinal plane it can be generated by a mismatch of the local oscillator phase with respect to the bunch phases, which is difficult to avoid when the bunches have different synchronous phases due to beam loading.

In the frequency domain, the presence of a stable beam signal carries non-zero revolution harmonics.

The stable beam signal is useless for feedback since it does not contain information about multi-bunch modes. It is therefore preferable to reduce it in order to avoid saturation of the ADC, DAC or amplifier. A number of techniques have been adopted to reduce this signal; we shall mention three examples.

3.3.1 Balancing of BPM buttons

In the case of transverse feedback, variable attenuators placed on the BPM electrodes can be used to equalize the amplitude of the signals (Fig. 44). In this way closed-orbit offsets or unbalanced signals can be compensated for.

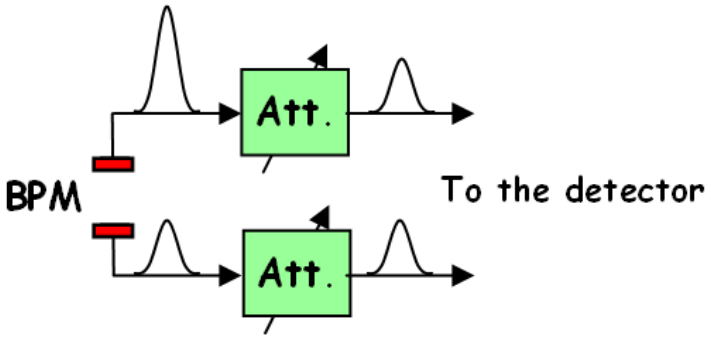
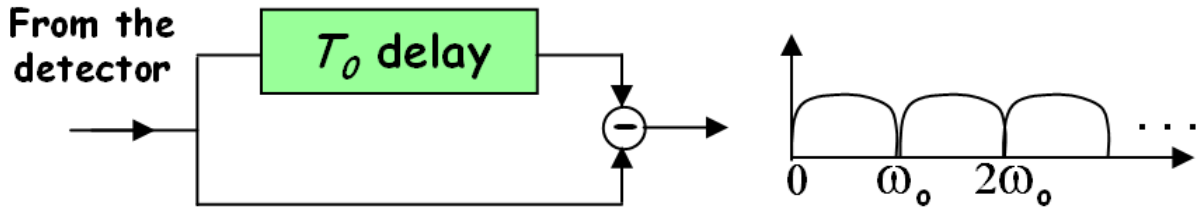


Fig. 44: Example of BPM signal balancing

3.3.2 Comb filter

An analogue filter performing stable beam rejection through the reduction of the revolution harmonics can be simply implemented by means of delay lines and combiners, as shown in Fig. 45(a). Its frequency response (Fig. 45(b), shows a series of notches at multiples of ω_0 able to suppress all the revolution harmonics (DC included) from the detector output signal [18].



(a) (b)

Fig. 45: Stable beam rejection using a comb filter. (a) Block diagram of the filter; (b) frequency response

3.3.3 Digital DC rejection

The stable beam signal can be eliminated by removing the DC component from the turn-by-turn signal of every bunch. This can be done digitally (Fig. 46).

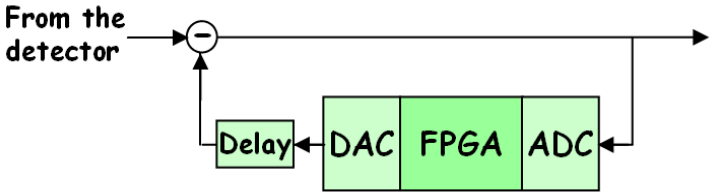


Fig. 46: Block diagram of a digital DC rejection system

The detector output signal is sampled by an ADC clocked at f_{rf}. The turn-by-turn digital signal of each bunch is integrated by a digital processing unit (e.g. a FPGA), recombined with the other bunch signals, converted to analogue, and subtracted from the original signal. In this way the constant component of each bunch is eliminated [20].

3.4 Digital processor

The A/D converter samples and digitizes the detector signal at the bunch repetition frequency; each sample corresponds to the position error (X, Y or Φ) of a given bunch. Precise synchronization of the sampling clock with the bunch signal must be provided.

The digital samples are then de-multiplexed into *M* channels, *M* being the number of bunches in the ring. In each channel the turn-by-turn samples of a given bunch are processed by a dedicated digital filter to calculate the correction samples. Basic processing consists of DC component suppression (if not completely accomplished by the external stable beam rejection) and phase shift of the signal at the betatron/synchrotron frequency (see Section 1.3.3).

The correction sample streams are then recombined and converted to analogue by the D/A converter (Fig. 47).

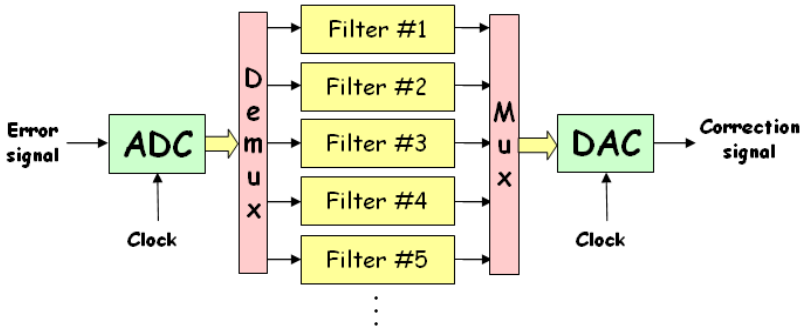


Fig. 47: Block diagram of bunch-by-bunch feedback by digital processor

3.4.1 Digital processor implementation

3.4.1.1 ADC

Existing multi-bunch feedback systems usually employ 8-bit ADCs clocked at up to 500 Msample/s; some implementations use a number of ADCs with higher resolution (e.g. 14 bits) and lower clock rate working in parallel. ADCs with enhanced resolution have the advantage of a lower quantization noise (crucial for low-emittance machines) and a higher dynamic range, which allows, in some cases, the avoidance of external stable beam rejection.

3.4.1.2 DAC

The DACs usually employed convert samples at up to 500 Msample/s and 14-bit resolution.

3.4.1.3 *Digital processing*

Feedback processing can be performed by discrete digital electronics (an obsolete technology), DSPs or FPGAs. Figure 48 summarizes the advantages and disadvantages of the last two options. Although there are still some digital multi-bunch feedback systems employing DSPs, the trend is towards solutions using FPGAs.

	Pros	Cons
DSP	➢ Easy programming ➢ Flexible	➢ Difficult HW integration ➢ Latency ➢ Sequential program execution ➢ A number of DSPs are necessary
FPGA	➢ Fast (only one FPGA is necessary) ➢ Parallel processing ➢ Low latency	➢ Trickier programming ➢ Less flexible

Fig. 48: Summary of advantages and disadvantaged of DSPs and FPGAs

3.4.2 *Examples of digital processors*

The following are examples of digital processor implementation employed in a number of accelerators across the world. We would like to point out that this list is probably incomplete and not up-to-date.

3.4.2.1 *PETRA transverse and longitudinal feedback*

This is one of the first digital implementations of bunch-by-bunch feedback [21]. The digital processor is composed of an ADC, digital processing electronics made of discrete components (adders, multipliers, shift registers, etc.) implementing a FIR filter and a DAC.

3.4.2.2 *ALS/PEP-II/DAΦNE longitudinal feedback*

This is also adopted at SPEAR, Bessy II, and PLS [22, 16, 17]. The A/D and D/A conversions are performed by VXI boards, while the feedback processing is done by DSP boards hosted in a number of VME crates (Fig. 49).

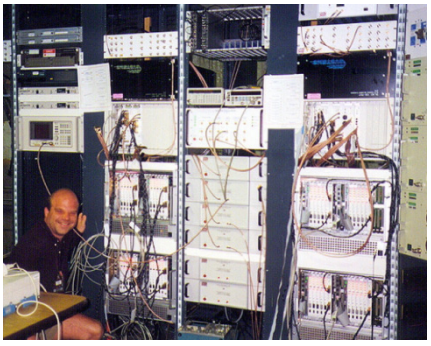


Fig. 49: The ALS/PEP-II/DAΦNE longitudinal feedback system

3.4.2.3 *PEP-II transverse feedback*

The digital part, made of two ADCs, a FPGA and a DAC, features a digital delay and integrated diagnostics tools, while the rest of the signal processing is analogue [23, 24].

3.4.2.4 KEKB transverse and longitudinal feedback

The digital processing unit, made of discrete digital electronics and banks of memories (Fig. 50), performs a two-tap FIR filter featuring stable beam rejection, phase shift, and delay [25, 26].

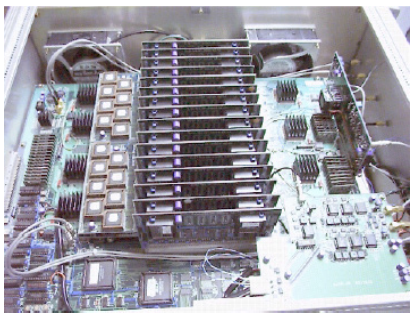


Fig. 50: The digital processing board of the KEKB feedback systems

3.4.2.5 Elettra/SLS transverse and longitudinal feedback

The digital processing unit is made of a VME crate equipped with one ADC, one DAC and six commercial DSP boards with four microprocessors each (Fig. 51) [27, 28]. Further developments at SLS allowed the direct connection of ADC and DAC boards, exploiting the on-board FPGAs to perform feedback processing [29].

Fig. 51: The Elettra horizontal, vertical, and longitudinal feedback processors

3.4.2.6 CESR transverse and longitudinal feedback

These employ VME digital processing boards equipped with ADC, DAC, FIFOs and PLDs [30].

3.4.2.7 HERA-p longitudinal feedback

This is made of a processing chain with two ADCs (for I and Q components), a FPGA and two DACs [31].

3.4.2.8 HLS tranverse feedback

The digital processor consists of two ADCs, one FPGA and two DACs [32].

3.4.2.9 SPring-8 transverse feedback

This is also adopted at TLS, KEK-Photon-Factory, and Soleil [33–36]. It employs a fast analogue de-multiplexer that distributes analogue samples to a number of slower ADC–FPGA channels. The correction samples are converted to analogue by one DAC (Fig. 52).

Fig. 52: The processing unit of the SPring-8 multi-bunch feedback system

3.4.2.10 ESRF transverse/longitudinal and Diamond transverse feedback

These adopt a commercial product called Libera Bunch-by-bunch [20, 37] from Instrumentation Technologies [38] (Fig. 53), which features four ADCs sampling the same analogue signal appropriately delayed, one FPGA and one DAC.

Fig. 53: Libera Bunch-by-bunch by Instrumentation Technologies

3.4.2.11 DAΦNE transverse and KEK-Photon-Factory longitudinal feedback

These feedback systems [39] rely on a commercial product called iGp from Dimtel [40] (Fig. 54), featuring a ADC–FPGA–DAC chain.

Fig. 54: iGp by Dimtel

3.5 Amplifier and kicker

The kicker is the feedback actuator. It generates a transverse/longitudinal electromagnetic field that steers the bunches with small 'kicks' as they pass through the kicker. The overall effect is the damping of the betatron/synchrotron oscillations. The power amplifier provides the kicker with the necessary RF power by amplifying the signal from the DAC (or from the modulator in the case of longitudinal feedback).

An important parameter that measures the efficiency of the kicker is the shunt impedance R, defined as the ratio between the squared voltage seen by the bunch and twice the power at the kicker input

$$R = \frac{V^2}{2P_{\mathrm{IN}}} \,. \tag{11}$$

Since the shunt impedance depends on the frequency of the input signal, it is normally represented as a function of frequency. We shall use this definition below, when calculating the power necessary to damp coupled-bunch oscillations.

The amplifier and kicker need a bandwidth of at least $f_{rf}/2$. In a transverse feedback, for example, frequencies go from ~DC (all kicks of the same sign) to ~$f_{rf}/2$ (kicks of alternating signs). The bandwidth of the amplifier and kicker must be sufficient to correct each bunch with the appropriate kick without affecting the neighbouring bunches. The amplifier and kicker design has to maximize the kick strength while minimizing the cross-talk between corrections given to adjacent bunches.

Another issue is the group delay that should be constant in the working frequency range, namely, the phase response should be linear. If this is not the case, the feedback efficiency is reduced when damping some multi-bunch modes and, under given conditions, the feedback can even become positive. Figure 55 shows an example of amplitude and phase response of an amplifier. Where the phase is 180°, the multi-bunch mode corresponding to that frequency is excited.

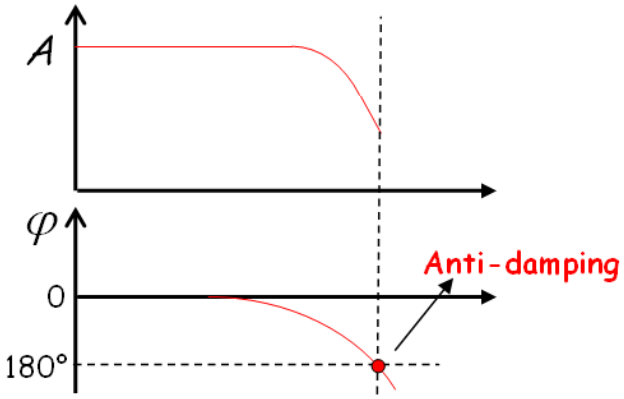


Fig. 55: An example of amplifier amplitude and phase response where, for a given frequency, the feedback performs anti-damping

3.5.1 Transverse feedback

A stripline geometry is usually employed for the transverse feedback kicker. The amplifier and kicker work in baseband, namely in the frequency range from ~DC to ~$f_{rf}/2$. Figure 56 shows an example of a transverse feedback back-end adopting two power amplifiers feeding the downstream ports of two striplines in counter-phase. The upstream ports are connected to 50 Ω power loads. Power low-pass filters can optionally be included to protect the amplifiers from peak voltages picked up from the beam by the kicker.

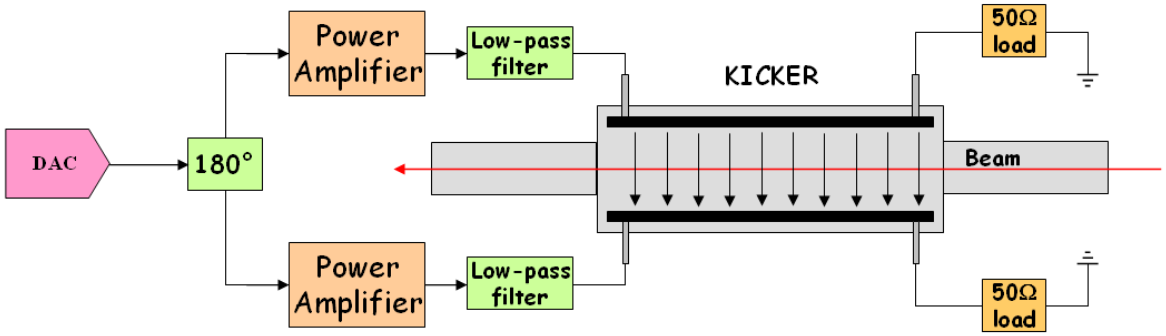
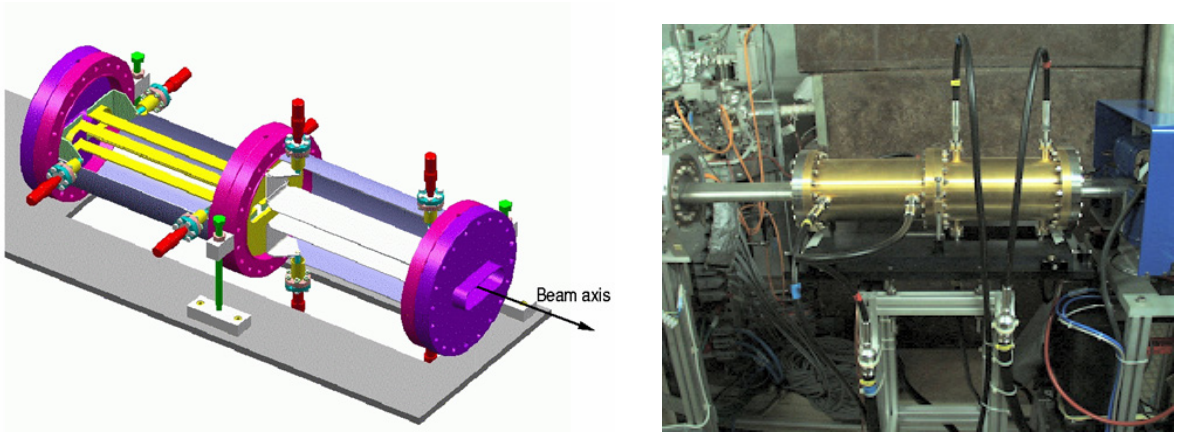


Fig. 56: An example of a transverse feedback back-end using two power amplifiers to feed the kicker

The design of the Elettra/SLS transverse kickers [41] is depicted in Fig. 57(a) together with a picture of the kickers installed on the Elettra vacuum chamber (Fig. 57(b)), while the shunt impedance of the same kickers as a function of frequency is plotted in Fig. 58.



(a) (b)

Fig. 57: (a) The Elettra/SLS horizontal and vertical kickers; (b) kickers installed on the Elettra vacuum chamber

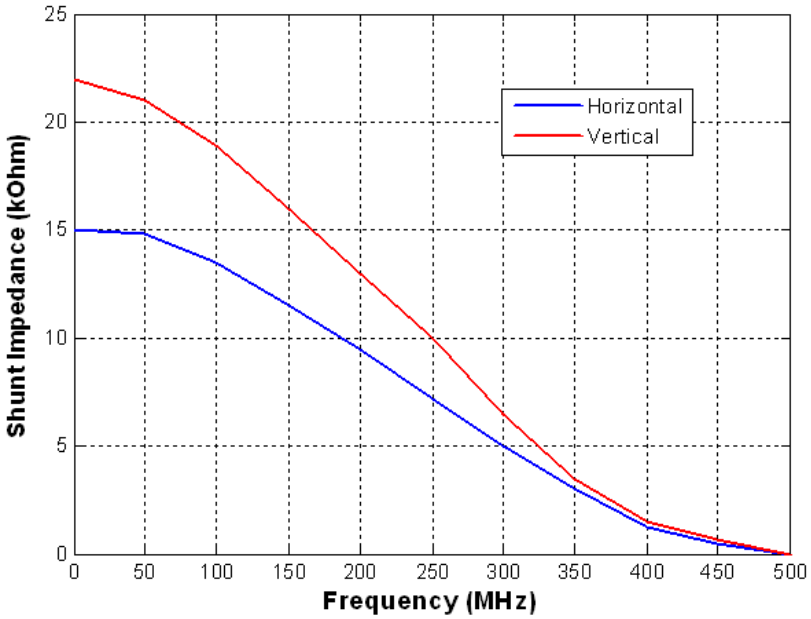


Fig. 58: Shunt impedance of the Elettra/SLS transverse kickers

3.5.2 Longitudinal feedback

For the longitudinal kicker, a 'cavity-like' structure is usually preferred because of the higher shunt impedance and smaller size. As in the case of the transverse kicker, the operating frequency range of the longitudinal kicker is $f_{rf}/2$ wide, but it is usually placed on one side of a multiple of f_{rf} (e.g. from $3\,f_{rf}$ to $3\,f_{rf} + f_{rf}/2$).

The baseband signal from the DAC has to be translated in frequency or, in other words, 'modulated' in order to overlap with the kicker shunt impedance (Fig. 59). Single SideBand (SSB) amplitude modulation or other techniques such as, for example, Quadrature Phase Shift Keying (QPSK) modulation can be adopted. The modulated signal is then amplified and sent to the kicker. In this case a circulator can be employed to protect the amplifier from reverse power (Fig. 60) [42]. The Elettra/SLS longitudinal kicker is illustrated (Fig. 61) [43].

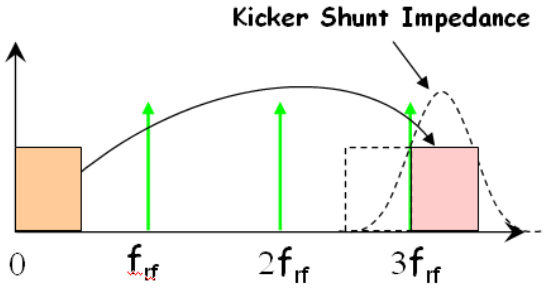


Fig. 59: The baseband signal is modulated to overlap with the kicker shunt impedance

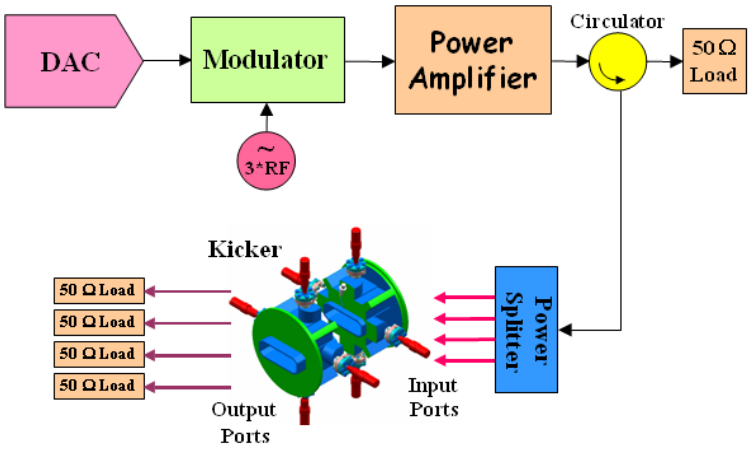
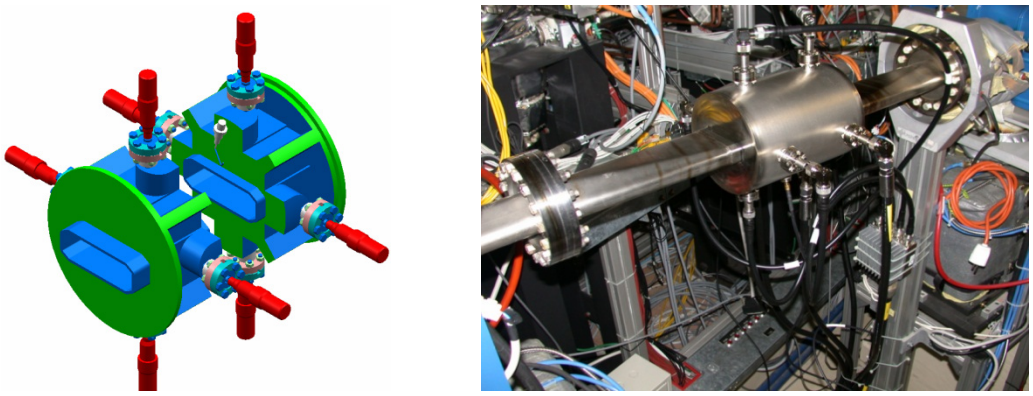


Fig. 60: Block diagram of a typical longitudinal feedback back-end



(a) (b)

Fig. 61: (a) The Elettra/SLS longitudinal kicker; (b) kicker installation on the Elettra vacuum chamber

3.6 Control system integration

Each component of the feedback system that needs to be configured and adjusted should include an interface to the accelerator control system (Fig. 62). There should be the ability to perform any operation remotely to facilitate system commissioning and the optimization of its performance. Moreover, as we shall see below, it is crucial to have an effective data acquisition channel to provide fast transfer of large amounts of data for analysis of feedback performance and for beam dynamics studies. It is also preferable to have direct access to the feedback system from a numerical computing environment and/or a script language like MATLAB [44] or similar products (Octave, Scilab, Python, IGOR Pro, IDL, etc.) for quick development of measurement procedures using scripts as well as for data analysis and visualization.

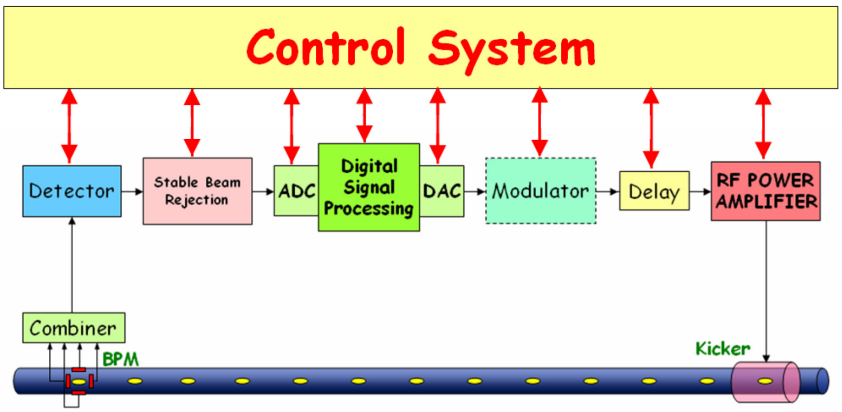


Fig. 62: Integration of the feedback system components into the accelerator control system

4 RF power requirements

In this section we determine the formulae to calculate the RF power required to damp coupled-bunch instabilities.

4.1 Transverse feedback

The transverse motion of a bunch of particles not subject to damping or excitation can be described as a pseudo-harmonic oscillation with amplitude proportional to the square root of the β function

$$x(s)=a\sqrt{\beta(s)}\cos\phi(s), \quad \text{where} \quad \phi(s)=\int_0^s \frac{\mathrm{d}\bar{s}}{\beta(\bar{s})} \tag{12}$$

where s is the longitudinal coordinate. The derivative of the position, namely the angle of the trajectory, is

$$x'=-\frac{a}{\sqrt{\beta}}\sin\phi+\frac{a\beta'}{2\sqrt{\beta}}\cos\phi, \qquad \text{with} \quad \phi'=\frac{1}{\beta} \tag{13}$$

By introducing $\alpha=-\frac{\beta'}{2}$ we can write

$$x'=\frac{a}{\sqrt{\beta}}\sqrt{1+\alpha^2}\,\sin(\varphi+\arctan\alpha) \tag{14}$$

Let us consider now a kicker placed at coordinate s_k; the kicker electromagnetic field deflects the bunch trajectory, which varies its angle by k. As a consequence the bunch starts another oscillation

$$x_1=a_1\sqrt{\beta}\cos\varphi_1 \tag{15}$$

which must satisfy the following constraints

$$\begin{cases} x(s_k)=x_1(s_k) \\ x'(s_k)=x_1'(s_k)+k \end{cases} \tag{16}$$

By introducing

$$A = a\sqrt{\beta},\ A_1 = a_1\sqrt{\beta} \tag{17}$$

Eq. (16) becomes a two-equation, two-unknown-variables system:

$$\begin{cases} A\cos\varphi = A_1\cos\varphi_1 \\ A\dfrac{\sqrt{1+\alpha^2}}{\beta}\sin(\varphi+\operatorname{arctg}(\alpha)) = A_1\dfrac{\sqrt{1+\alpha^2}}{\beta}\sin(\varphi_1+\operatorname{arctg}(\alpha)) + k. \end{cases} \tag{18}$$

The solution of the system gives amplitude and phase of the new oscillation

$$\begin{cases} A_1 = \sqrt{(A\sin\phi - k\beta)^2 + A^2\cos^2\phi} \\ \phi_1 = \arccos\left(\dfrac{A}{A_1}\cos\phi\right). \end{cases} \tag{19}$$

From the first equation, if the kick is small ($k << A/\beta$) then

$$\frac{\Delta A}{A} = \frac{A - A_1}{A} \cong \frac{\beta}{A}\, k\, \sin\phi\ . \tag{20}$$

From the same equation it can be seen that in the linear feedback case, i.e. when the turn-by-turn kick signal is a sampled sinusoid with amplitude proportional to the bunch oscillation amplitude, in order to maximize the damping rate the kick signal must be in phase with sin φ

$$k = g\,\frac{A}{\text{‰}}\sin\varphi \qquad \text{with} \quad 0 < g < 1\ , \tag{21}$$

that is, in quadrature with the bunch oscillation (Eq. (12)).

The optimal gain g_{opt} (which must in any case be <1) is determined by the maximum kick value k_{max} that the kicker is able to generate. If we want the feedback to work in a linear regime, i.e. not in saturation, the feedback gain must be set so that k_{max} is generated when the oscillation amplitude A at the kicker location is maximum: $g_{opt} = \dfrac{k_{max}}{A_{max}}\beta$. Therefore

$$k = \frac{k_{max}}{A_{max}} A\sin\varphi. \tag{22}$$

For small kicks the relative amplitude decrease is

$$\frac{€A}{A} \cong \frac{k_{max}}{A_{max}}\ \text{‰}\sin^2\varphi \tag{23}$$

and its average is

$$\left\langle \frac{\Delta A}{A} \right\rangle \cong \frac{\beta\, k_{max}}{2\, A_{max}}\ . \tag{24}$$

The average relative decrease is therefore constant, which means that, on average, the amplitude decrease is exponential with time constant τ (damping time) given by

$$\frac{1}{\tau} = \left\langle \frac{\Delta A}{A} \right\rangle \frac{1}{T_0} = \frac{\beta\, k_{max}}{2\, A_{max}\, T_0}\ , \tag{25}$$

where T_0 is the revolution period. By referring to the oscillation amplitude at the BPM location

$$\frac{1}{\tau} = \frac{k_{\max}}{2\,T_0\,A_{B\max}}\sqrt{\beta_K \beta_B} \tag{26}$$

where $A_{B\max}$ is the maximum oscillation amplitude at the BPM location.

In the case of relativistic particles, the change of the transverse momentum p of the bunch passing through the kicker can be expressed by

$$\Delta p = \frac{e}{c} V_\perp ,$$

where

$$V_\perp = \int_0^L (\bar{E} + c \times \bar{B})_\perp \, \mathrm{d}z$$

is the kick voltage, e is the electron charge, c is the speed of light, $\bar{E}$ and $\bar{B}$ are the fields in the kicker and L is the length of the kicker. Here, p can be written as $\frac{E_0}{c}$, where E_0 is the beam energy and $V_\perp$ can be derived from the definition of kicker shunt impedance

$$R_k = \frac{V_\perp^2}{2P_k} .$$

The maximum deflection angle in the kicker is given by

$$k_{\max} = \frac{\Delta p}{p} = e\frac{V_\perp}{E_0} = \left(\frac{e}{E_0}\right)\sqrt{2P_k R_k} . \tag{27}$$

From the previous equations we obtain

$$P_K = \frac{2}{R_K \%\varphi}\left(\frac{E_0}{e}\right)^2 \left(\frac{T_0}{\text{TM}}\right)^2 \left(\frac{A_{B\max}}{\sqrt{\%\varphi_B}}\right)^2 . \tag{28}$$

Given a free bunch oscillation with amplitude at the BPM location $A_{B\max}$, Eq. (28) allows us to calculate the power necessary to damp this oscillation with time constant τ. In the case of excited oscillations, in order to damp the instability, the feedback damping time constant must be less than the instability growth time-constant (see Section 1.3.3).

It has to be noted that this equation is only valid if all of the system components are ideal and the feedback is perfectly tuned.

The maximum power is supplied when the oscillation amplitude is at its maximum, namely at the time the feedback is switched on. During the damping process, the power decreases with the square of the oscillation amplitude. This leads us to an important conclusion: in order to determine the required power of the RF amplifier, we have to know not only the strength of the multi-bunch instability we want to damp, but also its maximum amplitude. For the same reason it is preferable to switch on the feedback when the oscillation is small: if we keep the feedback gain high (a small oscillation corresponds to the entire dynamic range of DAC/amplifier) it will be easier to catch and keep it damped.

The following is an example of the calculation of the required power for the Elettra transverse feedback system. The parameters are R_k = 15 kΩ (average value), E_B/e = 2 GeV, T_0 = 864 ns, τ = 120 μs, $\beta_{B\ H,V}$ = 5.2, 8.9 m, $\beta_{K\ H,V}$ = 6.5, 7.5 m.

The plots in Fig. 63 show the required power in the horizontal and vertical planes as a function of the initial oscillation amplitude.

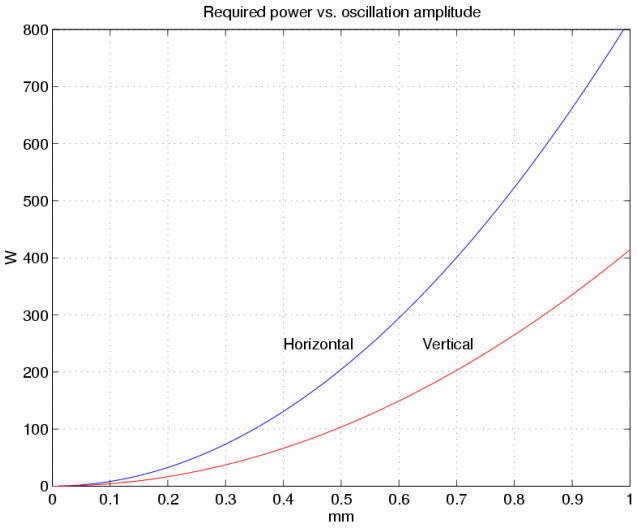


Fig. 63: Power required to damp a transverse instability as a function of the initial oscillation amplitude calculated for the Elettra storage ring

4.2 Longitudinal feedback

A similar procedure can be adopted to calculate the power required for longitudinal feedback. The relationship between the energy oscillation amplitude ε and the phase oscillation amplitude φ of a longitudinal synchrotron oscillation is given by

$$\varepsilon = \frac{E_0\ \omega_s}{2\,\pi\ f_{\mathrm{rf}}\ \alpha}\,\phi\ , \tag{29}$$

where ω_s is the synchrotron frequency, α the momentum compaction factor, and f_{rf} the RF frequency. As in the transverse case, the damping time constant τ is given by the equation

$$\frac{1}{\tau} = \frac{eV_{\max}}{2\,\varepsilon_{\max} T_0}\ , \tag{30}$$

where $V_{\max}$ is the maximum kick voltage and $\varepsilon_{\max}$ is the maximum energy oscillation amplitude.

By introducing the longitudinal kicker shunt impedance R_k,

$$R_k\ = \frac{V^2}{2\,P_k}\ , \tag{31}$$

the power required to damp a longitudinal oscillation is given by

$$P_k\ = \frac{2}{R_k}\left(\frac{\omega_S}{\omega_0}\frac{E_0}{e}\frac{\phi_{\max}}{\alpha\ f_{\mathrm{rf}}\ \tau}\right)^2\ . \tag{32}$$

5 Digital signal processing

As mentioned in Section 3.4, the digital processing in a bunch-by-bunch feedback is performed by *M* separated channels, each dedicated to one bunch (Fig. 64).

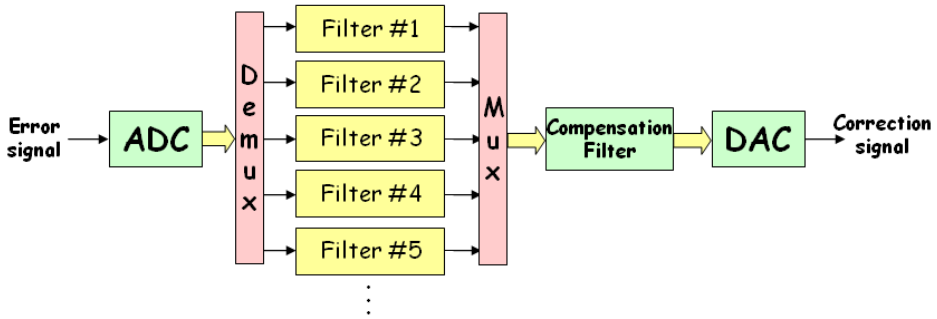


Fig. 64: Block diagram of the digital processor

In order to damp the bunch oscillations, the turn-by-turn kick signal must be the derivative of the bunch position at the kicker, namely, for a given oscillation frequency, a $\pi/2$ phase-shifted signal must be generated. In determining the real phase shift to perform in each channel, the phase advance between BPM and kicker must be taken into account as well as any additional delay due to the feedback latency (multiple of one machine revolution period). Moreover, any residual constant offset (stable beam component) must be rejected from the bunch signal to avoid DAC saturation.

These basic tasks can be accomplished by digitally filtering the streams of turn-by-turn samples in each channel, namely calculating the present correction sample on the basis of the past position samples.

Additionally, to compensate for the not-ideal characteristics of an amplifier and kicker, a digital filter can be put in the full-rate data path before the DAC.

5.1 Digital filter design

Digital filters can be implemented with finite impulse response (FIR) or infinite impulse response (IIR) structures. In general, IIR filters offer better performance than FIR filters. The disadvantage is a more difficult design and an increased complexity of the filter implementation.

Various techniques are used to design digital filters, including frequency domain and model-based design. In the following sections we show some examples of digital filters actually adopted in bunch-by-bunch feedback systems.

5.1.1 Three-tap FIR filter

The minimum requirements for each of the *M* digital filters are:

i) DC rejection (for a FIR filter this means that the sum of the coefficients must be zero);

ii) given amplitude response at the tune frequency;

iii) given phase response at the tune frequency.

A simple three-tap FIR filter can fulfil these requirements. Since it is a very simple filter, the coefficients can be calculated analytically. Let us consider the *Z*-transform of the filter response, namely the filter transfer function $H(z)$; in order to have zero amplitude response at DC, $H(z)$ must have a 'zero' in $z = 1$. Another zero in $z = c$ is needed in order to have the required phase response α at the tune frequency ω:

$$H(z) = k(1 - z^{-1})(1 - cz^{-1}) \quad . \tag{33}$$

The parameter c can be calculated analytically

$$\begin{aligned} H(z) &= k(1-(1+c)z^{-1}+c\,z^{-2}) \qquad z = \mathrm{e}^{\mathrm{j}\omega} \\ H(\omega) &= k(1-(1+c)\,\mathrm{e}^{-\mathrm{j}\omega}+c\,\mathrm{e}^{-2\mathrm{j}\omega}) \\ \mathrm{e}^{-\mathrm{j}\omega} &= \cos\omega - \mathrm{j}\sin\omega\,, \quad \alpha = \mathrm{ang}(H(\omega)) \\ \mathrm{tg}(\alpha) &= \frac{c(\sin(\omega)-\sin(2\omega))+\sin(\omega)}{c(\cos(2\omega)-\cos(\omega))+1-\cos(\omega)} \\ c &= \frac{\mathrm{tg}(\alpha)(1-\cos(\omega))-\sin(\omega)}{(\sin(\omega)-\sin(2\omega))-\mathrm{tg}(\alpha)(\cos(2\omega)-\cos(\omega))}. \end{aligned} \tag{34}$$

Here k is determined given the required amplitude response at tune $|H(\omega)|$:

$$k = \frac{|H(\omega)|}{\sqrt{\left(1-(1+c)\cos(\omega)+c\cos(2\omega)\right)^2 + \left((1+c)\sin(\omega)-c\sin(2\omega)\right)^2}}\,. \tag{35}$$

As an example: $\omega/2\pi = 0.2$, $|H(\omega)| = 0.8$, $\alpha = 222°$.

From the equations we can calculate $c = -0.63$ and $k = 0.14$. The transfer function of the filter is therefore $H(z) = -0.63 + 0.49\,z^{-1} + 0.14\,z^{-2}$. The filter response is depicted in Fig. 65.

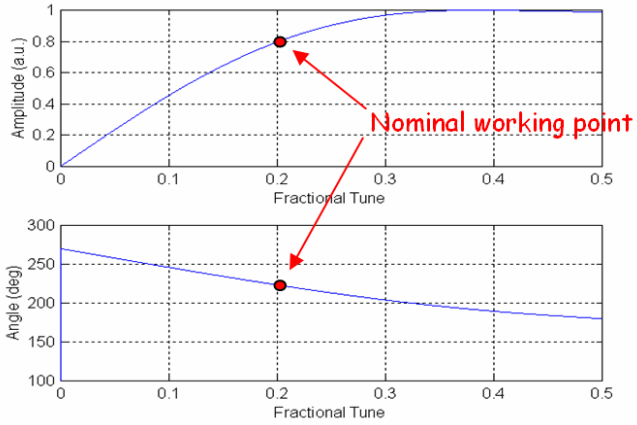


Fig. 65: Amplitude and phase response of a three-tap FIR filter

5.1.2 Five-tap FIR filter

With a longer FIR filter more degrees of freedom are available in the design, and additional features can be added to the filter. The following is an example of a five-tap FIR filter currently used at Elettra for transverse feedback [28].

The tune of an accelerator can significantly change during machine operations; the filter can be designed in order to guarantee the same feedback efficiency in a given frequency range. In this case the requirement is to have an optimal phase response.

In this example the feedback delay is four machine turns, which must be taken into account when calculating the filter phase response. When the tune frequency increases, the phase of the filter must increase as well; this means that the phase response must have a positive slope around the working point.

The filter design can be made using the MATLAB function *invfreqz*(). This function calculates the FIR filter coefficients that best fit the required frequency response using the least-squares method. The desired response is specified by defining amplitude and phase at three different frequencies: the origin and two frequencies, f_1 and f_2, chosen in the proximity of the nominal tune (Fig. 66, lower plot).

As we can see in the upper plot, this is achieved at the expense of a bad amplitude response, which features a minimum at the tune while the maximum is at a different frequency.

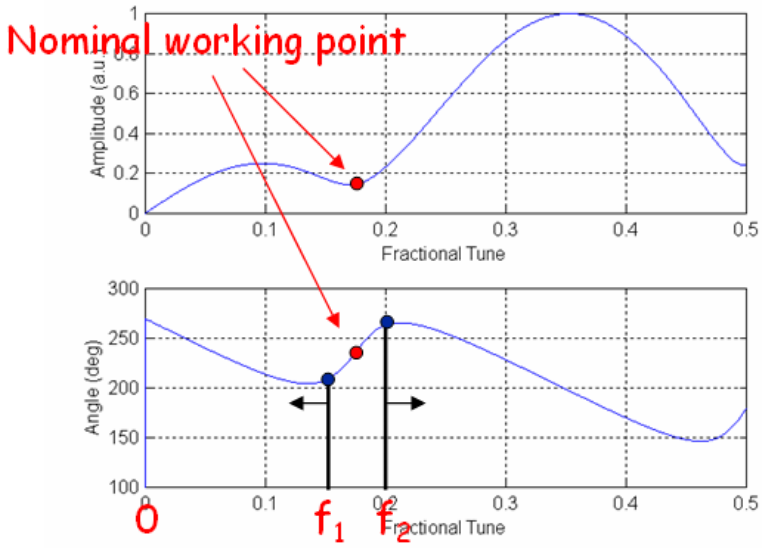


Fig. 66: Amplitude and phase response of a five-tap FIR filter

5.1.3 Selective FIR filter

A filter often employed in longitudinal feedback systems is the selective FIR filter, whose impulse response (the filter coefficients) is a sampled sinusoid with frequency equal to the synchrotron tune (see Fig. 67) [45]. As we can see in Fig. 68, the filter amplitude response has a maximum at the tune frequency and linear phase. The more filter coefficients we use the more selective the filter.

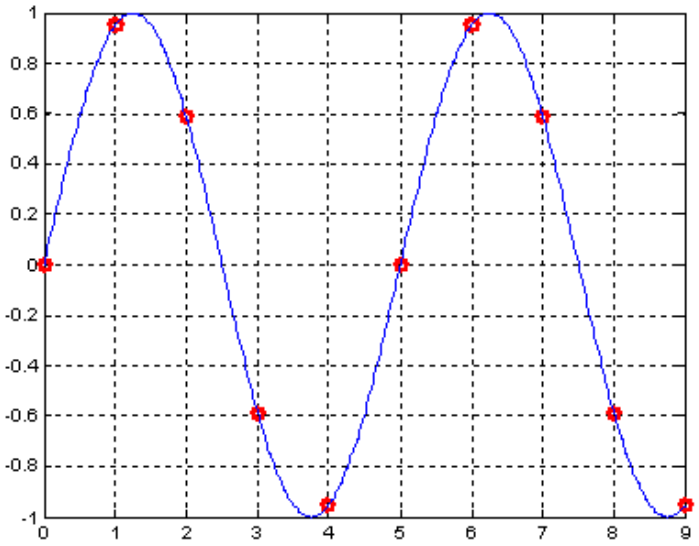


Fig. 67: Impulse response of a selective FIR filter (dots)

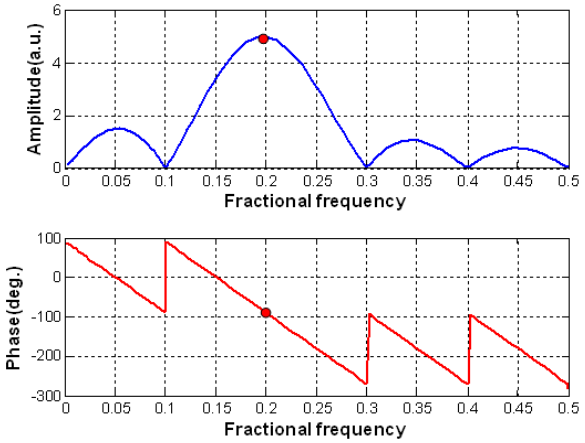


Fig. 68: Amplitude and phase response of a selective FIR filter

5.1.4 Advanced filter design

More sophisticated techniques using longer FIR or IIR filters enable a variety of additional features that exploit the potential of digital signal processing.

It is possible, for instance, to enlarge the working frequency range (as seen in Section 5.1.2) with no degradation of the amplitude response. Moreover, the filter selectivity can be enhanced to better reject unwanted frequency components (noise) or to minimize the amplitude response at frequencies that must not be fed back. An example of this design technique is the Time Domain Least Square Fitting (TDLSF) described in Ref. [33].

Another interesting possibility is to stabilize different tune frequencies simultaneously by designing a filter with two separate working points. This is required, for example, when horizontal and vertical as well as dipole and quadrupole instabilities have to be addressed by the same feedback system [35].

Advanced design techniques can also be employed to improve the robustness of the feedback under parametric changes of accelerator or feedback components (e.g. optimal control, robust control, etc.) [46].

5.2 Down-sampling

The synchrotron frequency is usually much lower than the revolution frequency, which means that one complete synchrotron oscillation is accomplished in many machine turns. In longitudinal feedback systems, in order to be able to properly filter the bunch signal, down-sampling is usually carried out [47]; one out of D bunch samples is used, where D is the down-sampling factor (Fig. 69).

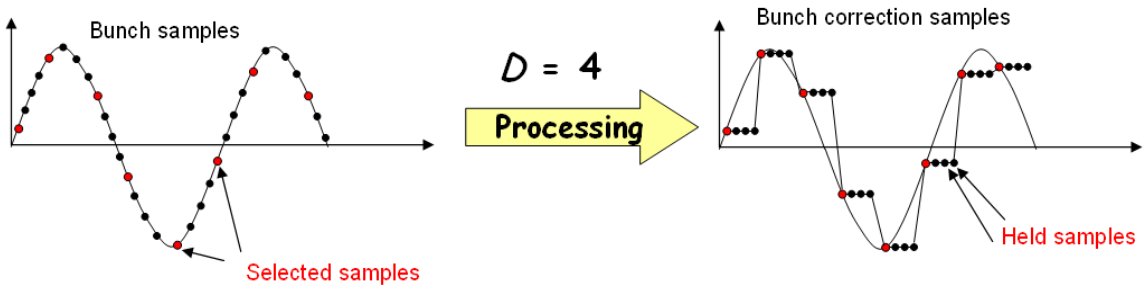


Fig. 69: Down-sampling of the bunch position signal and reconstruction by the hold buffer

Filtering is performed over the down-sampled digital signal, and the filter design is done in the down-sampled frequency domain (the original one enlarged by the down-sampling factor D). After

processing, the turn-by-turn correction signal is reconstructed by means of a hold buffer that keeps each calculated correction value for D turns, as shown in Fig. 69.

The reduced data rate allows for more time to be available to perform the filter calculations, and more complex filters can therefore be implemented.

6 Integrated diagnostic tools

A feedback system can implement a number of powerful diagnostic tools that are useful for both commissioning and tuning of the feedback system as well as for machine physics studies. The following are some of the tools normally implemented in digital feedback systems:

i) ADC data recording (see (1) in Fig. 70): acquisition and recording, in parallel with feedback operation, of a large number of samples for off-line data analysis;

ii) modification of filter parameters 'on the fly' with the required timing and even individually for each bunch (see (2) in Fig. 70): useful for switching feedback on/off, generation of grown/damped transients, optimization of feedback performance, etc.;

iii) injection of externally generated digital samples (see (3) in Fig. 70): for the excitation of single/multi-bunches.

These tools are usually managed by an additional controller with a fast interface to the accelerator control system.

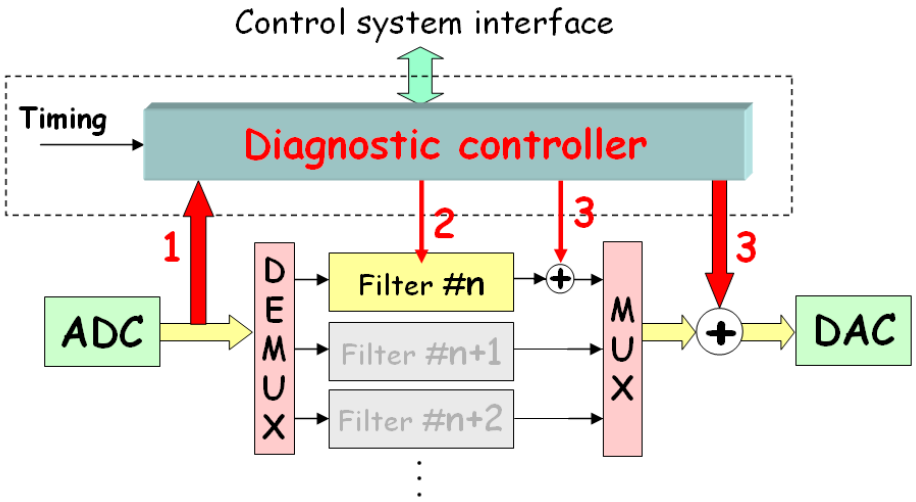


Fig. 70: Integrated diagnostics tools of a digital feedback system

Below we give some examples of measurements carried out using the diagnostics tools.

6.1 Acquisition and recording of ADC data

Figure 71 depicts the amplitude spectrum of a vertical bunch-by-bunch signal sampled at f_{rf} = 500 MHz. The spectrum is calculated with the FFT algorithm. At the moment of acquisition the beam was vertically unstable with some multi-bunch modes excited around the baseband frequency of 5 MHz.

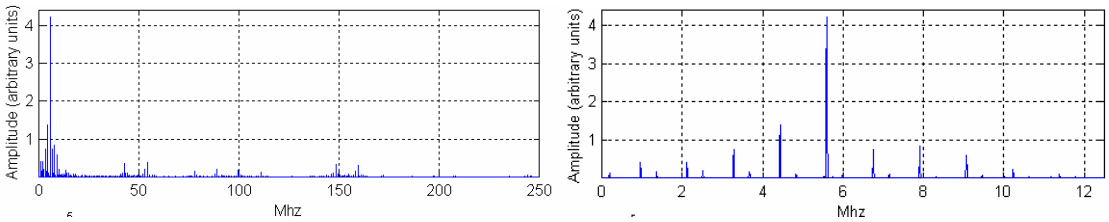


Fig. 71: Amplitude spectrum of bunch-by-bunch data of (a) an unstable beam; (b) a zoom view showing the unstable sidebands

6.2 Excitation of individual bunches

To produce an excitation, the feedback loop is switched off for one or more selected bunches (for example, by setting to zero the filter coefficients) and an excitation signal is injected in place of the bunch correction signal (Fig. 72). Possible excitation signals are white or pink noise and sinusoids.

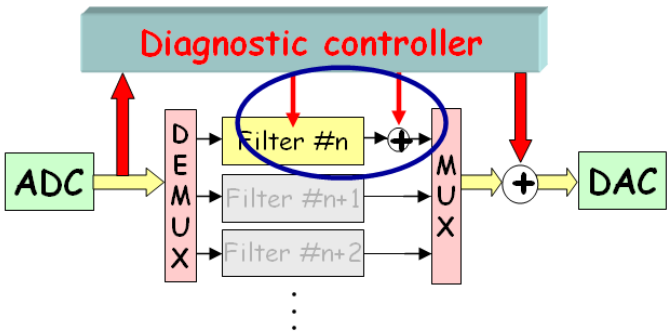


Fig. 72: Excitation of individual bunches through injection of appropriate signals

In the example in Fig. 73, two selected bunches are vertically excited with pink noise in a range of frequencies around the tune, while feedback is applied to the other bunches. The upper plot shows the excitation of the two bunches. The spectrum of one of the excited bunches reveals a peak at the tune frequency (lower plot). This technique can be used to measure the betatron tune of individual bunches with almost no deterioration of the beam quality.

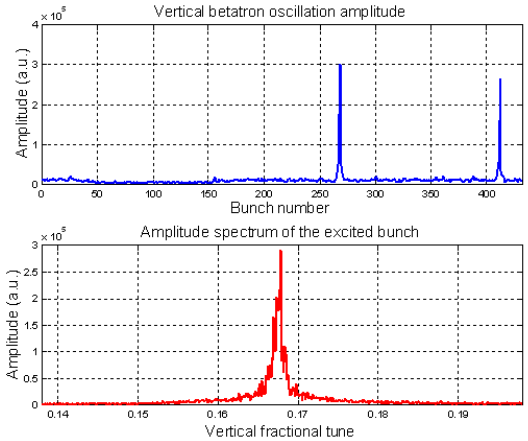


Fig. 73: Excitation of individual bunches by injection of appropriate digital signals

Excitation of individual bunches can also be done with feedback on by setting filter coefficients with the sign inverted. This produces positive feedback and anti-damping of the oscillations.

6.3 Multi-bunch excitation

Interesting measurements can be done by injecting pre-defined signals into the output of the digital processor (Fig. 74). Depending on the particular excitation signal and the filter settings, several experiments can be carried out.

By injecting a sinusoid at a given frequency, for example, the corresponding beam multi-bunch mode can be excited in order to test the performance of the feedback in damping that mode. If we inject an appropriate signal and record the ADC data when filter coefficients are set to zero, the beam transfer function can be calculated. By doing the same but with filter coefficients set to the nominal values, the closed loop transfer function can be determined.

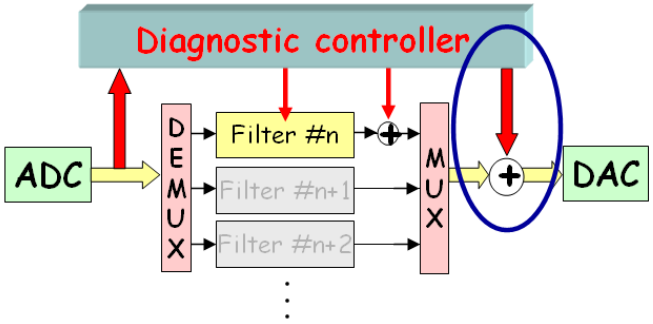


Fig. 74: Multi-bunch excitation through injection of appropriate digital signals

6.4 Transient generation

A powerful diagnostic application is the generation of transients. Transients can be generated by changing the filter coefficients according to a predefined timing and by concurrently recording the oscillations of the bunches (Fig. 75).

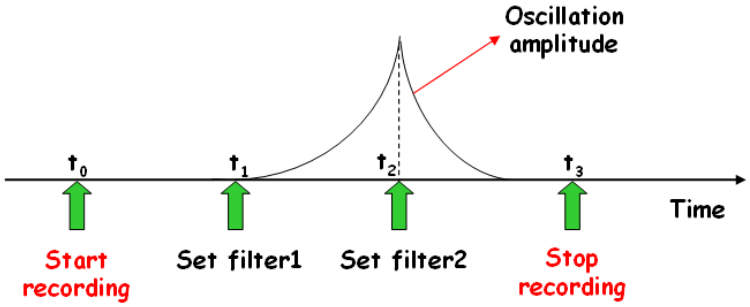


Fig. 75: Transient generation by changing the filter coefficients and recording the oscillations

The following are some examples of transients that can be used to measure damping times and growth rates by exponential fitting of the oscillation amplitude.

Let us suppose we have an excited multi-bunch oscillation that has reached equilibrium, i.e. constant amplitude: we can switch on the feedback and record the damping transient (Fig. 76(a)). If the feedback is already on, we can switch it off and on again after a while and generate a grow/damp transient (Fig. 76(b)).

Transients can also be generated by artificially exciting a stable beam with a proper setting of the filter coefficients in order to realize a positive feedback (anti-damping). The feedback is then switched off letting the oscillation decay by natural damping (Fig. 76(c)).

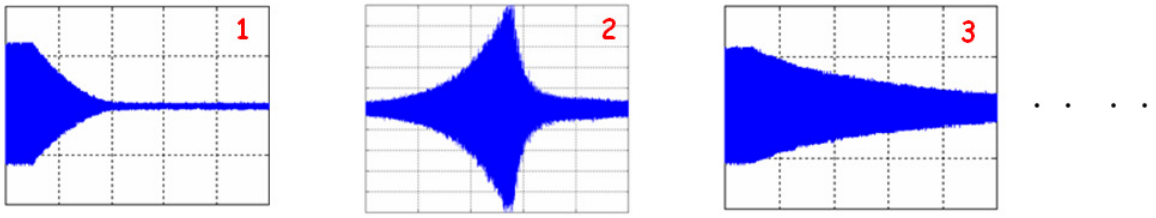


Fig. 76: Examples of transient generation

Growing/damped transients can be analysed by means of 3D graphs. Figure 77 is an example where the amplitude of the oscillation vs. time is plotted for every bunch. The absence of oscillation for some bunches is because the corresponding buckets are not filled with particles.

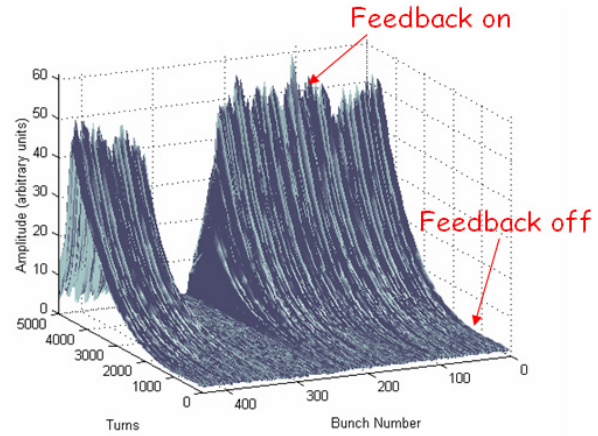


Fig. 77: 3D graph showing the evolution of the oscillation amplitude during a growing/damped transient

Transients can also be analysed by displaying the multi-bunch spectrum vs. time to study the evolution of individual coupled-bunch modes (Fig. 78).

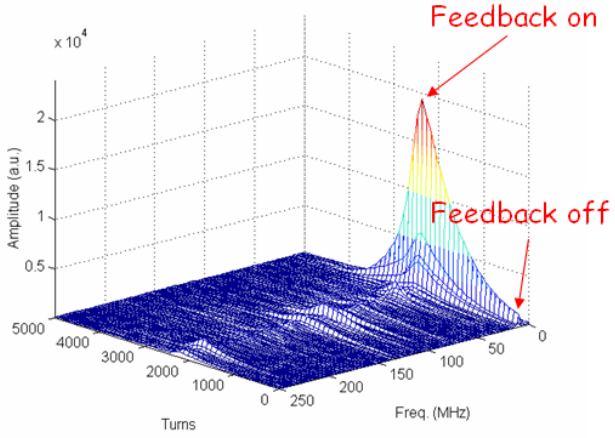


Fig. 78: Evolution of coupled-bunch unstable modes during a growing/damped transient

Transient generation is a helpful tool to tune a feedback system as well as to study coupled-bunch modes and beam dynamics [48–53]. The following list mentions some examples of applications using diagnostic tools and, in particular, transient analysis:

i) measurement of the feedback damping time: can be used to characterize and optimize feedback performance;

ii) measurement of the feedback resistive and reactive response: feedback not perfectly tuned can have a reactive behaviour, namely producing a tune shift when switched on that has to be minimized for optimum performance;

iii) modal analysis: measurement and analysis of the complex eigenvalue of the modes, namely the growth rate (real part of the eigenvalue) and the oscillation frequency (imaginary part of the eigenvalue);

iv) impedance measurement: the analysis of complex eigenvalues and bunch synchronous phases can be used to evaluate the machine impedance;

v) stable modes analysis: coupled-bunch modes below the instability threshold, thus being stable, can be studied to predict their behaviour at higher beam currents;

vi) bunch train studies: the analysis of the behaviour of different bunches along the bunch train can be used to determine the sources of coupled-bunch instabilities;

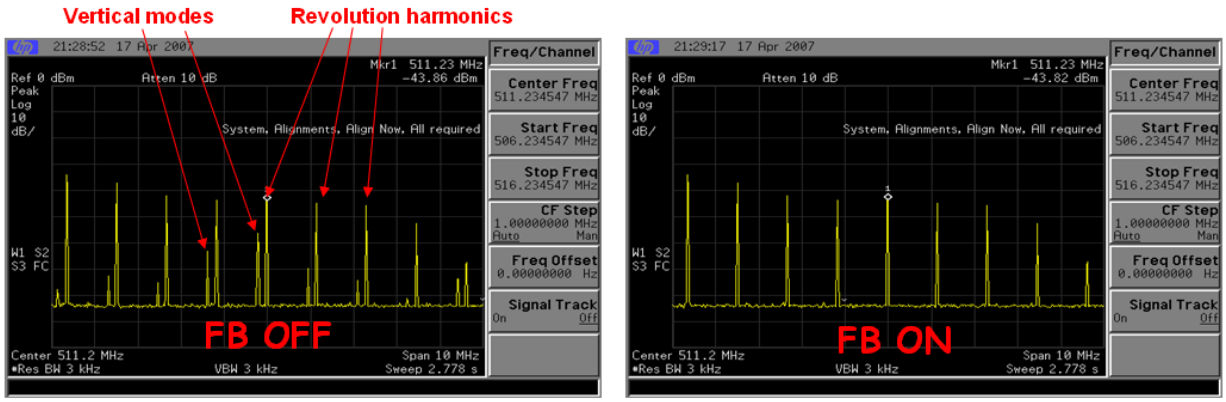
vii) phase space analysis: the analysis of the phase evolution of unstable coupled-bunch modes is used for beam dynamics studies.

7 Observation of feedback effects on beam characteristics

In this section we shall see, with the aid of some pictures, some examples of the main observable effects of feedback action measured using beam instrumentation.

7.1 Beam spectrum

When the feedback is conveniently tuned the sidebands corresponding to an unstable coupled-bunch mode are completely suppressed. Figure 79 shows an example of beam spectrum with feedback off and on.

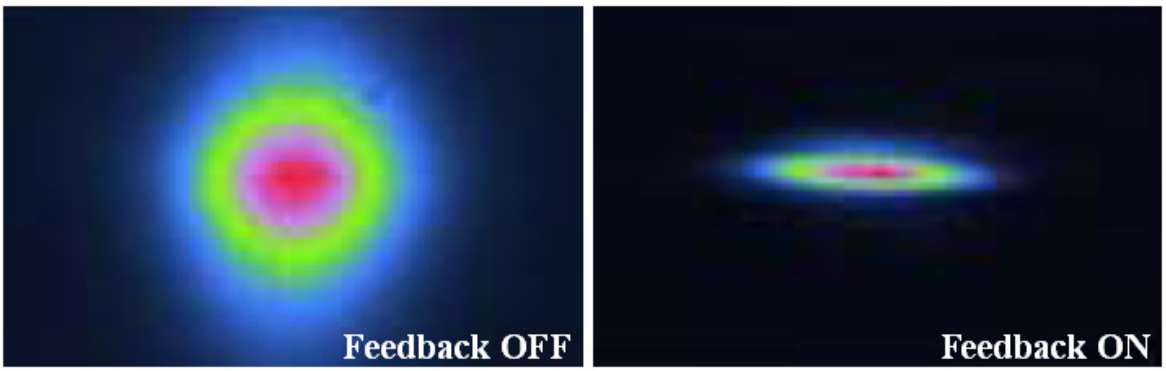


(a) (b)

Fig. 79: Screenshots from a spectrum analyser connected to a stripline pickup. (a) Feedback off; (b) feedback on. The sidebands corresponding to vertical coupled-bunch modes disappear as soon as the transverse feedback is activated.

7.2 Beam transverse profile

Images of an electron beam transverse profile can be produced using the synchrotron radiation generated by a bending magnet. As mentioned in Section 1.4.5, the transverse profile blows up in the presence of a transverse instability (Fig. 80(a)). When activated the feedback squeezes the beam, bringing it to its original size (Fig. 80(b)).

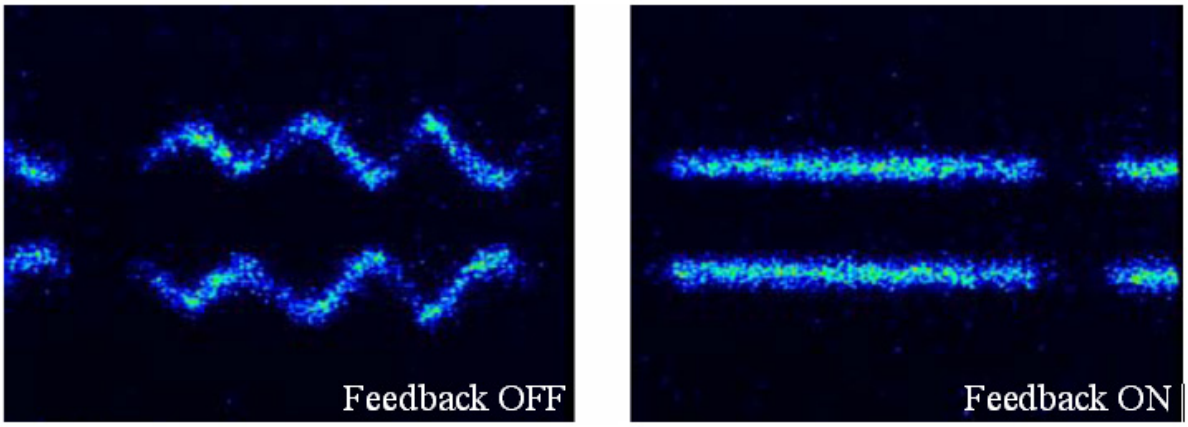


(a) (b)

Fig. 80: Synchrotron radiation monitor images taken at TLS [34] showing: (a) blow-up of the beam size due to a vertical instability; (b) its reduction when the feedback is switched on.

7.3 Streak camera images

Interesting measurements can be carried out by analysing the synchrotron radiation from a bending magnet using a streak camera. In Fig. 81(a) a longitudinal instability is clearly visible. The instability disappears when the feedback is activated (Fig. 81(b)).



(a) (b)

Fig. 81: Images of one machine turn taken with a streak camera in 'dual scan mode' at TLS [34]. The horizontal and vertical time spans are 500 and 1.4 ns, respectively. The longitudinal instabilities observable in (a) are suppressed by the feedback in (b).

7.4 Photon beam spectra

In a synchrotron light source, the ultimate goal of a feedback system is the improvement of the emitted photon beam quality. This can be evaluated by measuring the energy spectrum of the photons produced by an undulator. As we can observe in the example reported in Fig. 82, the amplitude and shape of the generated harmonics are noticeably improved when vertical instabilities are damped by the feedback.

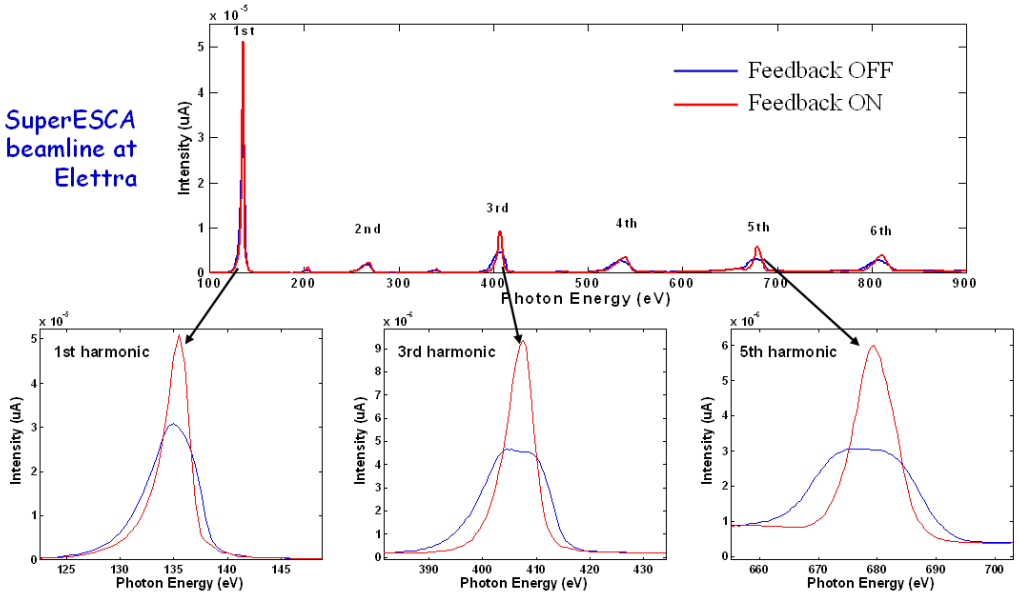


Fig. 82: Energy spectra of the photon beam measured on the SuperESCA beamline at Elettra

8 Conclusions

Feedback systems are indispensable tools for curing coupled-bunch instabilities in storage rings. Technological advances in digital electronics allow the implementation of digital feedback systems where the control algorithms are executed using programmable devices. The theory of digital signal processing is widely used to design and implement digital filters as well as to analyse data acquired by the systems. In fact, thanks to their capability of exciting the beam in different ways and measuring the resulting beam motion, feedback systems are used not only for closed-loop control but also as powerful diagnostic tools. The possibilities offered by these tools are manifold both for the optimization of feedback performance and for beam dynamics studies.

Acknowledgements

I would like to thank Dmitry Teytelman, a member of the SLAC feedback team, for many enlightening discussions and useful suggestions on these topics.

References

[1] M. Svandrlik *et al.,* The cure of multibunch instabilities in Elettra, Proc. PAC'95, Dallas, 1995.

[2] E. Courant *et al., Rev. Sci. Instrum*. **37** (1966) 1579.

[3] M. Zobov *et al.,* Measures to reduce the impedance of parasitic resonant modes in the DAΦNE vacuum chamber, Proc. 14th Advanced IFCA Beam Dynamics Workshop, Frascati, 1997, (Frascati Physics Series Vol. X, 1998), p. 371.

[4] F. Zimmermann *et al.,* First observations of a 'fast beam ion instability' at the ALS, Proc. PAC'97, Vancouver, 1997.

[5] R.A. Bosch *et al.,* Suppression of longitudinal coupled bunch instabilities by a passive higher harmonic cavity, Proc. PAC'93, Washington, 1993.

[6] E. Vogel *et al., Phys. Rev. Spec. Top. – Accel. Beams* **8** (2005) 102801.

[7] G. Lambertson, Control of coupled-bunch instabilities in high-current storage rings, Proc. PAC'91, San Francisco, 1991.

[8] J. Fox *et al.,* Feedback control of coupled-bunch instabilities, Proc. PAC'93, Washington, 1993.

[9] M. Serio *et al.,* Multibunch instabilities and cures, Proc. EPAC'96, Sitges, 1996.

[10] F. Pedersen, *Feedback Systems*, CERN PS/90-49 (AR), 1990.

[11] D Teytelman, Survey of digital feedback systems in high current storage rings, Proc. PAC2003, Portland, 2003.

[12] K. Balewski, Review of feedback systems, Proc. EPAC'98, Stockholm, 1998.

[13] F. Pedersen, Multibunch instabilities, Proc. Frontiers of Particle Beams: Factories with e^+ e^- Rings, Joint US-CERN School on Particle Accelerators, Benalmadena, 1992, pp. 269–292.

[14] F. Pedersen *et al., IEEE Trans. Nucl. Sci.* **NS-24** (1977) 1396.

[15] W. Barry *et al.,* Commissioning of the ALS transverse coupled-bunch feedback system, Proc. PAC'95, Dallas, 1995.

[16] S. Khan *et al.,* BESSY II feedback systems, Proc. PAC'99, New York, 1999.

[17] H.S. Kang *et al.,* Longitudinal and transverse feedback systems for PLS storage ring, Proc. PAC2001, Chicago, 2001.

[18] P. Wesolowsky *et al.,* A vertical multi-bunch feedback system for ANKA, Proc. PAC2005, Knoxville, 2005.

[19] A. Young *et al.,* RF and baseband signal processing in the PEP-II/ALS/DAΦNE longitudinal feedback system, Proc. DIPAC'97, Frascati, 1997.

[20] E. Plouviez *et al.,* Broadband bunch by bunch feedback for the ESRF using a single high resolution and fast sampling FPGA DSP, Proc. EPAC2006, Edinburgh, 2006.

[21] M. Enert *et al.,* Transverse and longitudinal multibunch feedback systems for PETRA, DESY 91–036, 1991.

[22] D. Teytelman *et al.,* Architecture and technology of 500 M sample/s feedback systems for control of coupled-bunch instabilities, Proc. ICALEPCS'99, Trieste, 1999.

[23] J. Byrd *et al.,* Design of the PEP-II transverse coupled-bunch feedback system, Proc. PAC'95, Dallas, 1995.

[24] J. Weber *et al.,* PEP-II transverse feedback electronics upgrade, Proc. PAC2005, Knoxville, 2005.

[25] M. Tobiyama *et al., Phys. Rev. Spec. Top. – Accel. Beams* **3** (2000) 012801.

[26] E. Kikutani *et al.,* Present status of the KEKB bunch feedback systems, Proc. EPAC2002, Paris, 2002.

[27] M Lonza *et al.,* Digital processing electronics for the Elettra transverse multibunch feedback system, Proc. ICALEPCS'99, Trieste, 1999.

[28] D. Bulfone *et al.,* Bunch-by-bunch control of instabilities with the ELETTRA/SLS digital feedback systems, Proc. ICALEPCS 2003, Gyeongju, 2003.

[29] M. Dehler *et al.,* State of the SLS multibunch feedbacks, Proc. APAC2007, Indore, 2007.

[30] J. Sikora *et al.,* Performance of the beam stabilizing feedback systems at CESR, Proc. PAC2001, Chicago, 2001.

[31] J. Randhahn *et al.,* Performance of the new coupled bunch feedback system at HERA-p, Proc. PAC2007, Albuquerque, 2007.

[32] Z.R. Zhou *et al.,* Development of digital transverse bunch-by-bunch feedback system of HLS, Proc. PAC2007, Albuquerque, 2007.

[33] T. Nakamura *et al.,* transverse bunch-by-bunch feedback system for the Spring-8 storage ring, Proc. EPAC2004, Lucerne, 2004.

[34] C.H. Kuo *et al.,* Control of the multi-bunch instabilities at TLS, Proc. APAC2007, Indore, 2007.

[35] W.X. Cheng *et al.,* Single-loop two dimensional transverse feedback for Photon Factory, Proc. EPAC2006, Edinburgh, 2006.

[36] R. Nagaoka *et al.,* Transverse feedback development at SOLEIL, Proc. PAC2007, Albuquerque, 2007.

[37] G. Rehm *et al.,* First tests of the transverse multibunch feedback at Diamond, Proc. DIPAC2007, Venice, 2007.

[38] http://www.i-tech.si/

[39] T. Obina *et al.,* Longitudinal feedback system for the Photon Factory, Proc. PAC2007, Albuquerque, 2007.

[40] http://www.dimtel.com/

[41] M. Dehler *et al.,* Current status of the Elettra/SLS transverse multi-bunch feedback, Proc. EPAC2000, Vienna, 2000.

[42] A. Ghigo *et al.,* Kickers and power amplifiers for the DAΦNE bunch-by-bunch longitudinal feedback system, Proc. EPAC'96, Sitges, 1996.

[43] M. Dehler, Kicker design for the ELETTRA/SLS longitudinal multi-bunch feedback, Proc. EPAC2002, Paris, 2002.

[44] http://www.mathworks.com/

[45] H. Hindi *et al.,* Analysis of DSP-based longitudinal feedback system: Trials at SPEAR and ALS, Proc. PAC'93, Washington, 1993.

[46] H. Hindi *et al.,* A formal approach to the design of multibunch feedback systems: LQG controllers, Proc. EPAC'94, London, 1994.

[47] J. Fox *et al.,* Down sampled signal processing for a B Factory bunch-by-bunch feedback system, Proc. EPAC'92, Berlin, 1992.

[48] S. Prabhakar *et al.,* Calculation of impedance from multibunch synchronous phases: Theory and experimental results (SLAC-PUB-7979, 1998).

[49] D. Teytelmen *et al.,* Frequency resolved measurement of longitudinal impedances using transient beam diagnostics, Proc. PAC2001, Chicago, 2001.

[50] J. Fox *et al.,* Multi-bunch instability diagnostics via digital feedback systems at PEP-II, DAΦNE, ALS and SPEAR, Proc. PAC'99, New York, 1999.

[51] D. Teytelman *et al.,* Observation, control and modal analysis of longitudinal coupled-bunch instabilities in the ALS via a digital feedback system, (SLAC-PUB-7292, LBNL-39329 1996).

[52] S. Khan, Diagnostics of instabilities at BESSY II using the digital longitudinal feedback system, Proc. EPAC2002, Paris, 2002.

[53] A. Drago, Horizontal instability and feedback performance in DAΦNE E+ ring, Proc. EPAC2004, Lucerne, 2004.

Bibliography

H. Winick, *Synchrotron Radiation Sources – A Primer* (World Scientific, Singapore, 1994).

A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).

A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

G.F. Franklin, J.D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems* (Addison-Wesley, Reading, MA, 1994).

S. Prabhakar, New diagnostics and cures for coupled-bunch instabilities (PhD thesis, Stanford University, SLAC-Report-554, 2000).

D. Teytelman, Architectures and algorithms for control and diagnostics of coupled-bunch instabilities in circular accelerators (PhD thesis, Stanford University, SLAC-Report-633), 2003.

Proceedings of the CAS-CERN Accelerator School: Intensity Limitations in Particle Beams, Geneva, Switzerland, 2–11 November 2015, edited by W. Herr, CERN Yellow Reports: School Proceedings, Vol. 3/2017, CERN-2017-006-SP (CERN, Geneva, 2017)

Beam Loss Consequences

F. Cerutti
CERN, Geneva, Switzerland

Abstract
A summary of beam loss types and their effects is presented.

Keywords
Beam–machine interaction; particle shower calculation; radiation effects.

1 Beam losses

The operation of any particle accelerator features some kinds of relevant beam losses. These can take place at devices meant to intercept the beam, such as targets, dumps, stoppers, collimators, and stripping foils. At the design stage, one must consider both regular and accidental impacts. The latter ones are uncontrolled and may be due to magnet failures at injection or extraction, for instance kicker misfiring or electron beam missteering during top-off injection. As a consequence, the impact conditions become extremely severe, in terms of intensity (i.e., number of impacting particles) or brilliance (i.e., intensity per unit area). In this way, a collimator, instead of intercepting the beam halo at a controlled rate, is directly hit by a few bunches. Moreover, any device on the beam path, as the foreseen beam dilution pattern is lost, is subject to too-small spot sizes that are potentially harmful.

One must also routinely deal with diffused losses, throughout the beamline of both linear and circular machines. Linacs are concerned by several loss mechanisms [1]. Lepton rings are especially affected by synchrotron radiation and gas bremsstrahlung, while in hadron rings nuclear reactions between beam particles and residual gas nuclei are an important source of secondary showers, whose amount scales with the gas density and the beam current.

Colliders are exposed to debris regularly produced at the interaction points, proportionally with the delivered luminosity. Typically, its most energetic component, travelling close to the primary beam, escapes the detector and impinges on the accelerator elements, e.g. the final focus quadrupoles.

Furthermore, the beam trajectory may cross unexpected obstacles, represented by dust particles drifting inside the vacuum chamber or by flawed aperture restrictions. The mass thickness of the object and the fraction of current traversing it determine the loss strength.

2 Consequences

Any of the aforementioned beam loss types is the onset of a secondary particle cascade, whose amplitude depends on the primary particle energy. The scenarios span from low-energy beam absorption within the material surface layers, by ionization, to combined hadronic and electromagnetic shower development over hundreds of metres of machine elements, as for instance in the large hadron collider (LHC). In such a broad context, many different physical processes are involved and their microscopic description, as integrated in multipurpose Monte Carlo codes, allows the calculation of relevant macroscopic quantities and, therefore, the evaluation of loss consequences.

Among these, applying to distinct time-scales, one can list the following.

- *Heating.* This is a short-term effect, owing to the total power deposited in the material by the impinging radiation. It calls for cooling measures when needed.
- *Thermal shock.* Remaining in the short-term domain, the material can undergo rupture, depending on the peak power density and on its spatial distribution.

https://doi.org/10.23730/CYRSP-2017-003.515

- *Quenching*. Far below its damage threshold, a superconducting material loses its ability to conduct electricity without resistance as it warms above a critical temperature, owing to the level and profile of energy deposition density induced by the radiation impact.
- *Single-event effects in electronic devices*. These range from bit upsets perturbing device functionality to destructive burn-outs. Their probability of occurrence is proportional to the time integrated high energy hadron fluence. In addition to these stochastic events, the steady accumulation of defects can ultimately lead to device failure, which can be anticipated on the basis of the expected ionizing and non-ionizing dose (with the latter generally quantified in terms of silicon 1 MeV equivalent neutron fluence).
- *Deterioration*. The long-term degradation of critical properties of organic materials (typically insulators) is related to the accumulated peak dose, while for inorganic materials the reference quantities are neutron fluence and displacements per atom in the hottest spots.
- *Oxidation, radiolysis, ozone production*. Determination of the impact of these chemical effects is based on the assessment of energy deposition. Knowing the power absorbed in an air volume, one can, for instance, calculate the resulting ozone concentration, as a function of the ventilation time.
- *Gas production*. A variety of residual nuclei is generated by nuclear reactions and their abundance can be naturally estimated in the simulation of the beam–machine interaction. Among these, one can distinguish (in addition to the radioactive species featured in one of the next items) those leading to gas build-up—typically hydrogen and helium—which in a solid material contributes to embrittlement and swelling.
- *Radiation in public spaces and shielding requirements*. For radiation protection purposes, depending on the aspects to be considered, particle fluence in a given location is transformed into effective dose or ambient dose equivalent (both expressed in sieverts) by means of respective sets of conversion coefficients, which are a function of particle type and energy. The prompt dose equivalent outside a radiation facility, reflecting the applicable radiation level in a public space during normal or accidental operation of the facility, is the quantity to reduce to below acceptable limits through a suitable shielding design.
- *Equipment and air activation, radioactive waste production, access limitations*. Radionuclide generation is responsible for delayed emissions during beam absence that limit access and intervention possibilities as well as equipment handling, including waste disposal. Induced activation is characterized by isotope activities and spatial distribution of residual dose rate after relevant cooling times.
- *Tumour cell destruction*. Particle beams are routinely used in cancer radiotherapy. They are intended to maximize the dose delivery to the tumour mass and to spare as much as possible the healthy tissue. When moving from conventional treatment with photons to hadron therapy, the calculation of biological dose, taking into account the biological effectiveness of the radiation, becomes a key ingredient.

Beside these effects, the radiation development following a beam loss occurrence also allows it to be detected and quantified by suitably located monitors. Ionization chambers, such as the LHC beam loss monitors [2], provide an online observation and play a central role in the machine protection system, triggering beam aborting if the recorded signals exceed predefined thresholds. Their charge collection is proportional to the energy deposition in the sensitive gas volume, which can be simulated and thus represent a compelling benchmark for calculation validation. Another example is given by detectors measuring hadron fluence, thanks to the single-event effect principle previously indicated [3]. Again in the hadron therapy context, treatment monitoring opportunities are offered by the detection of prompt photons or charged particles from nuclear reactions, as well as by positron emission tomography exploiting the tissue activation, namely the production of β^+ emitters.

It should be noted that only the full description of the particle shower initiated by the beam loss can shed light on the actual link between a peripheral monitor signal and the quantity that matters for the accelerator operation or design, such as the peak energy density in a beam intercepting device or in a magnet coil, or even the dose distribution delivered to a patient.

For a more extended discussion of the radiation effects here outlined and of the physical processes behind them, the reader is referred to Ref. [4].

References

[1] M.A. Plum, in Proc. of the 2014 Joint International Accelerator School on Beam Loss and Accelerator Protection, Newport Beach, United States, 5 – 14 November 2014, edited by R. Schmidt, CERN-2016-002 (CERN, Geneva, 2016), pp. 39-62. https://doi.org/10.5170/CERN-2016-002.39

[2] B. Dehning *et al.*, *AIP Conf. Proc.* **648** (2002) 229. https://doi.org/10.1063/1.1524405

[3] G. Spiezia *et al.*, *IEEE Trans. Nucl. Sci.* **61** (2014) 3424. https://doi.org/10.1109/TNS.2014.2365046

[4] N.V. Mokhov and F. Cerutti, in Proc. of the 2014 Joint International Accelerator School on Beam Loss and Accelerator Protection, Newport Beach, United States, 5 – 14 November 2014, edited by R. Schmidt, CERN-2016-002 (CERN, Geneva, 2016), pp. 83-110. https://doi.org/10.5170/CERN-2016-002.83

Ions

R. Nagaoka
Synchrotron SOLEIL, Saint-Aubin, France

Abstract
Ions are generated in a storage ring through collisions of a stored beam with residual gases. Their formation and influence on a stored beam are discussed including the so-called two-beam instability. Several methods for alleviating the ion effects are described.

Keywords
Storage ring; ions; beam dynamics; collective effects; beam instability; electromagnetic interactions.

1 Introduction

Ions are described in this lecture note as a source of disturbance to a beam stored in a storage ring. They are normally created via collisions of a stored beam with residual gases existing in the beam duct. Since some atomic electrons are ripped off through these electromagnetic (EM) interactions, ions are positively charged. They can be collectively repelled or attracted by the stored beam depending upon the sign of electric charge of the stored beam. Ions can induce a number of detrimental effects on the beam such as tune shifts, lifetime reduction and collective beam instability, all leading to a performance degradation of a storage ring.

Historically, ions caused performance limitations in both proton and electron rings. Many studies were carried out by various groups to understand the beam dynamics mechanisms of the encountered phenomena and thereby to explore means to avoid them [1–14]. In light-source storage rings where the stability of a circulating beam is of crucial importance, the serious impact of ions had led some machines, such as DCI, ACO, APS, and KEK-PF, to switch the stored beam from electrons to positrons [15]. As a general trend, a lower beam emittance is achieved in modern storage rings to raise the ring performance in terms of luminosity and brilliance. As a consequence, the trapping of ions in the electro-static potential of a stored beam that caused serious beam instabilities and suffered many rings in the past seems to have become much less of an issue. However, a new direct type of interactions between a stored beam and a beam of ions occurring in a single passage of the former may become a risk for future storage rings due to stronger electro-static field generated by the stored beam. Thus, ions could still be a potentially dangerous source of perturbation for modern and future storage rings.

The present paper is organized as follows. In the next section, we shall go through the basics of beam–ion dynamics, starting with the question of why we have ions in a storage ring. The impact of residual gases and ions on a stored beam, and ion motions due to the EM fields of the stored beam shall be discussed. In Section 3, we shall look at beam instabilities caused by ions. One type is due to trapped ions and the other is the so-called Fast Beam–Ion Instability (FBII). In Section 4, mitigation methods for ion effects shall be reviewed. Some specific examples of observing ion effects shall also be introduced. Conclusions are given in Section 5.

2 Basis of beam–ion dynamics

In this section we shall go through some of the basic physical mechanisms, notions, and studies developed in the past surrounding beam–ion dynamics, starting with asking us why ions are present in

https://doi.org/10.23730/CYRSP-2017-003.519

vacuum chambers. The influence of residual gases and ions on the stored beam, as well as ion motions due to the EM fields of the stored beam and magnets, shall be considered. It must be noted that the descriptions below benefited much from the earlier lecture notes on ions from the CAS series [16, 17], and in particular that of Sakanaka from the KEK-OHO accelerator school series [18].

2.1 Ultra-vacuum and residual gases

In reaching a maximal ring performance, ideally a stored beam should not get any disturbance in circulating along its closed orbit. This means that the beam duct needs be kept under excellent vacuum. With the vacuum technology of today applied to accelerators, we can reach the pressure level of the so-called Ultra-High Vacuum (UHV), which corresponds to lower than 10^{-9} mbar. Nonetheless, the residual gases in UHV could still become a significant source of beam perturbations. The principal mechanism is the collision of residual gases with beam, which specifically means elastic and inelastic EM scattering, causing particle losses and beam lifetime reduction. Through scattering, the residual gases could lose their atomic electrons and be ionized. Generated ions, as an ensemble, could then create an electrostatic potential and act on the beam both incoherently and coherently. We shall see more specifically below what kinds of interactions could take place.

The general trend today with modern storage rings in further raising their performance is to store an increasingly high intensity beam with lower emittance, which is often combined with the use of narrower beam ducts, such as for insertion devices and stronger quadrupole focusing. The lowering of vacuum conductance and the combination of elements mentioned above, bring about a number of reasons for which issues related to vacuum, and therefore to ions, still potentially remain the high concern for modern and future accelerators.

2.2 Collisions with residual gases

Let us consider the case of an electron colliding with residual gases. The collision can be classified into the following three categories (Fig. 1):

i) Møller scattering, which is elastic scattering with an atomic electron;

ii) Rutherford scattering, which is elastic scattering with the EM field of a nucleus;

iii) bremsstrahlung, which is inelastic scattering with the EM field of a nucleus converting a part of the electron's kinetic energy into EM radiation.

The total collision cross-section $\sigma_{\mathrm{Col}}^{\mathrm{Total}}$ of an atom with an atomic number Z_{i} is expressed as a sum of the three:

$$\sigma_{\mathrm{Col}}^{\mathrm{Total}} = Z_{\mathrm{i}} \cdot \sigma_{\mathrm{Møller}} + \sigma_{\mathrm{Rutherford}}(Z_i) + \sigma_{\mathrm{Bremsstrahlung}}(Z_i) \ . \tag{1}$$

Denoting the density of the concerned molecule by d_{m} and the velocity of the stored beam by βc, the collision rate is given by

$$\frac{1}{\tau_{\mathrm{Col}}} = \sigma_{\mathrm{Col}}^{\mathrm{Total}} \cdot d_{\mathrm{m}} \cdot \beta c \ , \tag{2}$$

where d_{m} is related to the partial pressure P_{m} at 20°C via $d_{\mathrm{m}}\,[\mathrm{m}^{-3}] = 2.47 \times 10^{22}\, P_{\mathrm{m}}\,[\mathrm{mbar}]$. In cases where the molecule consists of several atoms and/or there are several species in the residual gases, we take a sum of all contributions,

$$\frac{1}{\tau_{\mathrm{Col}}} = \sum_m \sum_k \frac{1}{(\tau_{\mathrm{Col}})_{mk}} , \tag{3}$$

where the index k is for different atoms in a molecule m and the index m is for molecule species.

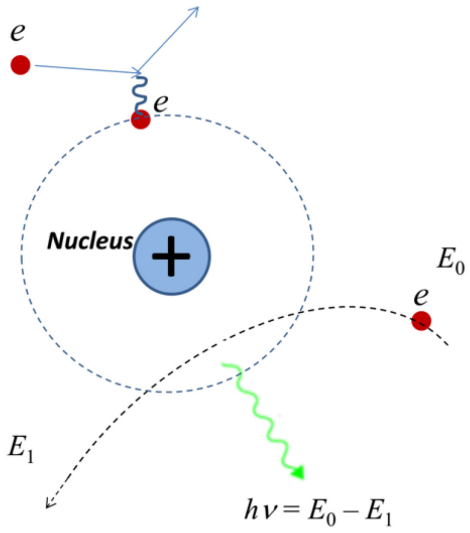


Fig. 1: Processes of collision of an electron with residual gases

2.3 Ionization of residual gases

In a similar way, we can obtain the ionization rate $1/\tau_{\mathrm{ion,m}}$ of a stored particle, or equivalently its ionization time $\tau_{\mathrm{ion,m}}$, by replacing $\sigma_{\mathrm{Col}}^{\mathrm{Total}}$ by the ionization cross-section $\sigma_{\mathrm{ion,m}}$ in the previous formula,

$$\frac{1}{\tau_{\mathrm{ion,m}}} = d_{\mathrm{m}} \cdot \sigma_{\mathrm{ion,m}} \cdot \beta c \ . \tag{4}$$

In general, $\sigma_{\mathrm{ion,m}}$ only depends on the species m and the velocity of the stored particle βc, as follows:

$$\sigma_{\mathrm{ion,m}} = 4\pi \left(\frac{\hbar}{m_{\mathrm{e}} c} \right)^2 \cdot (M^2 \cdot x_1 + C \cdot x_2) \ , \tag{5}$$

where m_{e} is the electron mass, the quantity

$$4\pi \left(\frac{\hbar}{m_{\mathrm{e}} c} \right)^2$$

equals 1.874×10^{-24} [m^2], x_1 and x_2 depend on β as

$$x_1 = \beta^2 \cdot \ln\left(\frac{\beta^2}{1-\beta^2}\right) - 1, \quad x_2 = \beta^2 \ ,$$

and M^2 and C are molecule dependent constants as shown in Table 1.

Table 1. Value of the constants M^2 and C for typical molecules.

Molecule	M^2	C	Z^a	A^b
H_2	0.5	8.1	2	2
N_2	3.7	34.8	14	28
CO	3.7	35.1	14	28
O_2	4.2	38.8	16	32
H_2O	3.2	32.3	10	18
CO_2	5.75	55.9	22	44
C_6H_4	17.5	162.4	46	76

[a]Atomic number; [b]mass number.

2.4 Beam-induced electromagnetic fields and their characteristics

Let us review the static EM field created by a round coasting beam of radius a and current I in a circular chamber of radius b (Fig. 2). Using the relation from Gauss's theorem,

$$\iint \vec{E} \cdot \mathrm{d}\vec{a} \;=\; \int \frac{\rho}{\varepsilon_0} \mathrm{d}V \quad ,$$

where ρ is the charge density and

$$\oint \vec{B} \cdot \mathrm{d}\vec{\ell} \;=\; \mu_0 I \quad ,$$

we get for the radial component of the electric field E_r,

$$E_r \;=\; \begin{cases} \dfrac{\mathrm{e}\lambda}{2\pi\varepsilon_0}\dfrac{r}{a^2} & (0 < r < a) \\ \dfrac{\mathrm{e}\lambda}{2\pi\varepsilon_0}\dfrac{1}{r} & (a < r) \end{cases} \quad , \tag{6}$$

and the azimuthal component of the magnetic field B_φ,

$$B_\varphi \;=\; \begin{cases} \dfrac{\mu_0 I}{2\pi a^2}\cdot r & (0 < r < a) \\ \dfrac{\mu_0 I}{2\pi r} & (a < r) \end{cases} \quad , \tag{7}$$

where $\lambda = I/(\mathrm{e}\beta c)$ is the line density of the electron beam and βc is the speed of the electrons. Since an ion charged to +1e having a longitudinal speed of $\beta_i c$ gets a force from the EM field of $F_r^E = \mathrm{e}E_r$ and $F_\varphi^B = \mathrm{e}\beta_i c B_\varphi$, we find from Eqs. 6 and 7 that for all values of r,

$$\frac{F_r^B}{F_r^E} \;=\; \beta_i\beta \;\cong\; \beta_i \;\ll\; 1 \quad ,$$

as ions are relatively heavy and move slowly. This leads to the fact the magnetic force F_{φ}^{B} of the stored beam can usually be ignored as compared to the electric field F_{r}^{E}.

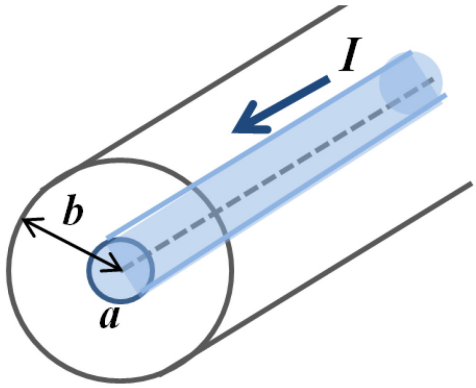


Fig. 2: A coasting beam of radius a and current I in a beam pipe of radius b

2.5 Ion trapping: Coasting beam

Using the static electric field E_r created by the beam (Eq. 6), the electric potential created by a coasting beam can be straightforwardly calculated as given by

$$V(r) = -\int_{b}^{r} E_r \, \mathrm{d}r = \begin{cases} \dfrac{e\lambda}{2\pi\varepsilon_0}\left(\dfrac{r^2}{2a^2} - \dfrac{1}{2} + \ln\dfrac{a}{b}\right) & (0 < r < a) \\ \dfrac{e\lambda}{2\pi\varepsilon_0}\cdot\ln\dfrac{r}{b} & (a < r) \end{cases} . \tag{8}$$

Evaluating the depth of the potential for realistic cases (see Fig. 3), we find that the ions having only a thermal energy in the order of $k_B T$ (~10^{-21} J) cannot escape from the potential, the value of which is of the order of some tens of volt (and therefore an energy of ~10^{-18} J). The other important feature to be noted is that the potential depth increases as the beam emittance decreases and the beam intensity increases. A beam potential calculated for the ISR ring is shown in Fig. 4. Locations of clearing electrodes are indicated by dots [17].

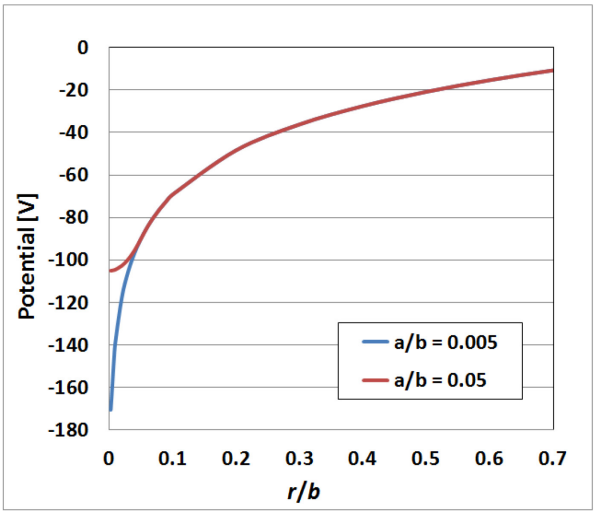


Fig. 3: Static electrostatic potential $V(r)$ created by a beam of radius a. The values for the parameters λ and b were taken from SOLEIL (I = 500 mA and b = 0.0125 m).

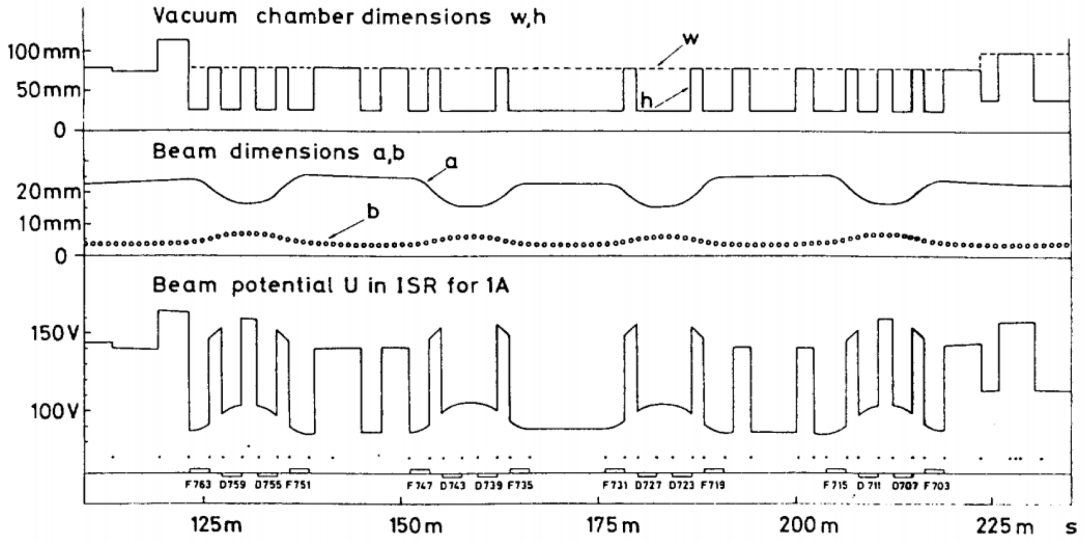


Fig. 4: Evaluation of the beam-induced electrostatic potential in ISR. The locations of clearing electrodes are indicated by dots. Taken from Ref. [17].

2.6 Neutralization

As the trapping of ions progresses, the potential depth decreases due to neutralization of opposing charges, which saturates the trapping process. The degree of neutralization is defined by

$$\eta = \frac{N_i}{N} , \tag{9}$$

where N_i and N are the total number of ions and electrons in the ring, respectively. If all ions are +1e charged, then $0 \leq \eta \leq 1$. For a proton ring, on the contrary, the electrons created by ionization could be trapped. Instead, ions could be repelled by the proton potential and bombard the chamber surface, which in turn induces outgassing. This could lead to a cascading phenomenon called the 'pressure bump'.

2.7 Bassetti–Erskine formula

Analytical expressions for the transverse electric fields E_x and E_y created by an electron bunch having Gaussian distributions in the two transverse planes were derived by Bassetti and Erskine [19]. Since electron beams are usually Gaussian and since, in addition, the transverse ion distributions are often approximated as those of the stored beam, the Bassetti–Erskine formula is frequently used in evaluating the electric forces felt by the two beams. Starting from the Poisson equation,

$$\nabla^2 \phi = \frac{\rho}{\varepsilon_0} , \tag{10}$$

with the transverse charge distribution function ρ given by

$$\rho(x,y) = \frac{Q}{2\pi\sigma_x\sigma_y} \cdot \exp\left[-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right) \right] \tag{11}$$

where Q represents the total charge over the transverse distribution of a bunch, and σ_x and σ_y are the RMS values of the Gaussian distributions, the potential ϕ was solved analytically by assuming $\sigma_x > \sigma_y$ in an integral form. It then follows that the electric field satisfies

$$E_x - \mathrm{i}E_y = -\mathrm{i}\frac{Q}{2\varepsilon_0\sqrt{2\pi(\sigma_x^2-\sigma_y^2)}}\cdot\left\{w(a+\mathrm{i}b) - \mathrm{e}^{[-(a+\mathrm{i}b)^2+(ar+\mathrm{i}b/r)^2]}\cdot w(ar+\mathrm{i}b/r)\right\} \quad (12)$$

with the parameters a, b, and r given by

$$a = \frac{x}{\sqrt{2(\sigma_x^2-\sigma_y^2)}},\quad b = \frac{y}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\text{, and } r = \frac{\sigma_y}{\sigma_x},$$

and $w(z)$ is the complex error function. The two transverse electric field components can consequently be expressed by

$$E_x = \frac{Q}{2\pi\varepsilon_0\sigma_x(\sigma_x+\sigma_y)}\cdot x + \text{higher-order terms}, \quad (13a)$$

$$E_y = \frac{Q}{2\pi\varepsilon_0\sigma_y(\sigma_x+\sigma_y)}\cdot y + \text{higher-order terms} \quad (13b)$$

respectively. Derivations for the case $\sigma_x < \sigma_y$ can similarly be made.

2.8 Ion trapping: Bunched beam

Let us now consider ion trapping with a bunched beam. With bunched beams, ions are only attracted during the passage of a bunch, and drift freely in between two bunches (in places where there are no magnets) (Fig. 5). Transverse motions of an ion thus resemble those of a circulating electron. Their stability (i.e. trapping) can be argued using transfer matrices in the linear approximation. Consider the vertical motion of an (+1e charged) ion in a symmetric beam filling. During the passage of a bunch, Newton's equation for an ion reads

$$M_{\mathrm{ion}}\ddot{y}_{\mathrm{i}} = \mathrm{e}E_y^{\mathrm{e}} = -\mathrm{e}\frac{Q}{2\pi\varepsilon_0\sigma_y(\sigma_x+\sigma_y)}\cdot y_{\mathrm{i}} = -\frac{N}{n_{\mathrm{b}}L_{\mathrm{b}}}\frac{2r_{\mathrm{p}}c^2m_{\mathrm{p}}}{\sigma_y(\sigma_x+\sigma_y)}\cdot y_i, \quad (14)$$

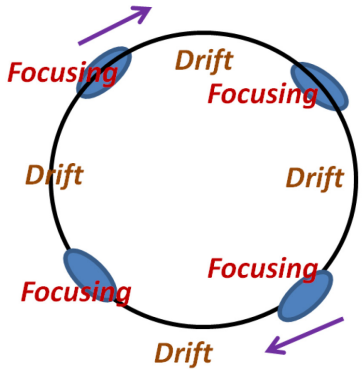


Fig. 5: The focusing system seen by an ion in a storage ring. Circulating electron bunches act as focusing quadrupoles and spacing between them as drifts.

where the double dot signifies a double derivative with respect to the time variable, $M_{\mathrm{ion}} = A \cdot m_{\mathrm{p}}$, where A is the mass number of the ions, N denotes the total number of stored electrons, n_{b} is the number of bunches, r_{p} is the classical proton radius (= $\mathrm{e}^2/4\pi\varepsilon_0 m_{\mathrm{p}}c^2$), L_{b} is the total bunch length, m_{p} is the proton mass, and E_y^{e} represents the vertical electric field created by the electron beam. In the familiar transfer matrix formulation, from the above linear equation of motion we have

$$\begin{pmatrix} y_{\mathrm{i}} \\ \dot{y}_{\mathrm{i}} \end{pmatrix}_{\mathrm{new}} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix} \begin{pmatrix} y_{\mathrm{i}} \\ \dot{y}_{\mathrm{i}} \end{pmatrix}_{\mathrm{old}} , \tag{15}$$

during the passage of the electron beam. The focusing parameter a is given by

$$a = \frac{N}{n_{\mathrm{b}}} \frac{2r_{\mathrm{p}}c}{\beta\sigma_y(\sigma_x + \sigma_y)} \frac{1}{A} \tag{16}$$

and β represents the relativistic factor corresponding to the speed of the stored electrons. The transfer matrix equation in between the passage of two bunches is that of a drift space

$$\begin{pmatrix} y_{\mathrm{i}} \\ \dot{y}_{\mathrm{i}} \end{pmatrix}_{\mathrm{new}} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{\mathrm{i}} \\ \dot{y}_{\mathrm{i}} \end{pmatrix}_{\mathrm{old}} , \tag{17}$$

where τ is given by $2\pi R/n_{\mathrm{b}}\beta c$ (where R is the ring radius).

2.9 Critical mass

The transfer matrix for one period, i.e. from the beginning of a bunch to the next, is therefore

$$M_{\mathrm{period}} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix} . \tag{18}$$

The condition for any linear motions to be bounded $-2 \le Tr(M_{\mathrm{period}}) \le 2$ leads to

$$A \ge A_{\mathrm{c}} \equiv \frac{N}{n_{\mathrm{b}}} \frac{r_{\mathrm{p}}}{n_{\mathrm{b}}} \frac{\pi R}{\beta^2\sigma_y(\sigma_x + \sigma_y)} . \tag{19}$$

The ion mass A_{c} defined by the above relation is called the critical mass. In this model we find that ions whose mass is lower than the critical mass cannot be trapped. It means that, as they are light, the focusing strength of the electron bunches deviates them too far away from the electron trajectory before the next bunch comes to attract them. The evidence of critical mass was experimentally observed in the 1.5 GeV electron–positron collider ADONE, in Frascati, Italy [20].

From Eq. 19, we can extract that ions have less chance of being trapped in a mode where there are few bunches, as A_{c} depends on the number of bunches as inversely square. Another aspect that is important is that the critical mass depends inversely on the product of the two sigmas, σ_x and σ_y, which in turn signifies that A_{c} effectively depends on the horizontal emittance as inversely square. This means that ions are much less likely to be trapped in a low-emittance ring. Again, the reason for this is the focusing of the electron beam being too strong when its transverse dimension is reduced, which kicks the ions away before the next bunch arrives. A numerical comparison showing the above dependence

of the critical mass on the number of bunches and the emittance has been made using the SOLEIL parameters (Fig. 6).

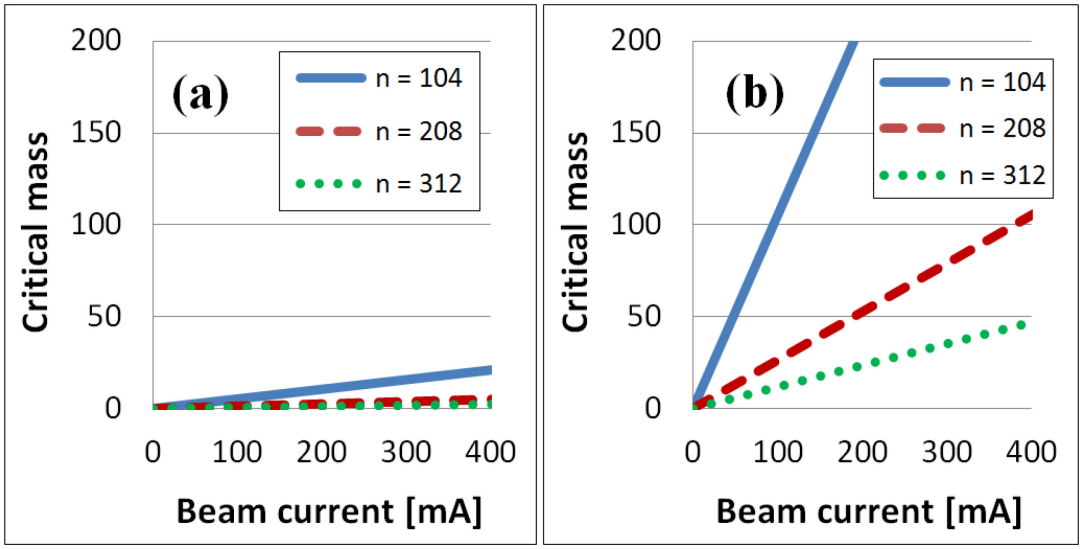


Fig. 6: Critical mass (Eq. 19) as a function of beam current for three symmetric beam fillings with different number of bunches. Machine parameters were taken from SOLEIL. Circumference = 354 m. (a) Horizontal emittance of 4 nm·rad (current value). (b) Horizontal emittance of 0.2 nm·rad (expected upgrade value). In both cases, the vertical emittance is assumed to be 1% of the horizontal value.

2.10 Betatron tune shifts

A cloud of trapped ions generally gives a transverse focusing force to the stored beam, inducing betatron tuneτ shifts $\Delta\nu_{x,y}$. They can be evaluated by the well-known formula

$$\Delta\nu = \frac{1}{4\pi}\oint \beta(s)\cdot\Delta k(s)\,\mathrm{d}s\ , \tag{20}$$

where $\beta(s)$ denotes the beta function, $\Delta k(s)$ represents the quadrupolar errors in a ring, and the integral over s is made around the ring. Assuming that the ions have the same Gaussian distributions as the electrons and are charged to +1e, we can use the Bassetti–Erskine formula (Eq. 13) to obtain the focusing strength $\Delta k_{\mathrm{i}}(s)$ due to ions as follows:

$$(\Delta k_{\mathrm{i}})_{x,y}(s) = \frac{1}{E_0/\mathrm{e}}\frac{\partial (E_{\mathrm{i}})_{x,y}}{\partial x,y} = \frac{1}{E_0/\mathrm{e}}\frac{d_{\mathrm{i}}}{\varepsilon_0}\cdot\frac{\mathrm{e}}{1+\sigma_{x,y}/\sigma_{y,x}}\ , \tag{21}$$

where d_{i} [m^{-3}] represents an ion density.

Similarly, the focusing strength $\Delta k_{\mathrm{SC}}(s)$ due to an electron beam's own space-charge force is given by

$$(\Delta k_{\mathrm{SC}})_{x,y}(s) = \frac{1}{\gamma^2}\frac{1}{E_0/\mathrm{e}}\frac{d_{\mathrm{e}}}{\varepsilon_0}\cdot\frac{\mathrm{e}}{1+\sigma_{x,y}/\sigma_{y,x}}\ , \tag{22}$$

where γ is the relativistic energy factor for the electrons and d_{e} [m^{-3}] is the electron density. Comparing the two tune shifts, we find that, since usually $\frac{d_{\mathrm{e}}}{d_{\mathrm{i}}}\cdot\frac{1}{\gamma^2} \ll 1$, we have $(\Delta\nu_{x,y}^{\mathrm{SC}}) \ll (\Delta\nu_{x,y}^{\mathrm{ions}})$, namely, the

space charge tune shift is generally much smaller for a relativistic electron beam than that induced by ions.

2.11 Ion motions

When the ion mass A is much larger than the critical mass A_C, and if there are no magnetic fields, their drift motions between the passages of two successive bunches can be neglected. The resultant motions therefore become approximately harmonic oscillations,

$$\ddot{u}_i \approx -\omega_{iu}^2 u_i , \tag{23}$$

where

$$\omega_{iu}^2 = \frac{2\lambda r_p c^2}{A} \frac{1}{\sigma_u(\sigma_x + \sigma_y)} \quad (u = x, y) , \tag{24}$$

where $\lambda = I/(e\beta c)$ is the line density of the electrons. Inside a bending magnet where the magnetic field B is non-zero, the equations of ion motion then become

$$\begin{pmatrix} \ddot{s} \\ \ddot{x} \\ \ddot{y} \end{pmatrix} = \begin{pmatrix} 0 \\ -\omega_{ix}^2 \cdot x \\ -\omega_{iy}^2 \cdot y \end{pmatrix} + \omega_c \begin{pmatrix} -\dot{x} \\ \dot{s} \\ 0 \end{pmatrix} , \tag{25}$$

where ω_c is the cyclotron angular frequency given by eB/M_{ion}. The solution to the above equations are known to be off-centred sinusoidal motions for x and s at the frequency $\omega = \sqrt{\omega_{ix}^2 + \omega_c^2}$. In particular, ions drift *longitudinally* at the average speed of

$$< \dot{s} > = \left(\frac{\omega_{ix}}{\omega_c} \right)^2 \left[\omega_c \cdot x(0) + \dot{s}(0) \right] . \tag{26}$$

Ions generally tend to move longitudinally towards a minimum of the potential $V(r)$ of the stored beam. Since $V(0) = e\lambda/(2\pi\varepsilon_0)\cdot[\ln(a/b) - ½]$ (see Eq. (8)), they gather at locations in the ring where a/b is small (i.e. where the stored beam size is small and the chamber aperture is large), which are called the *neutralization spots*. It is considered effective to introduce ion-clearing electrodes at such positions.

2.12 Ion distributions

Many studies assume that ions created by the collision with the circulating beam have the same transverse distributions as the latter, which are usually Gaussian. The above assumption is correct regarding the initial ion distribution when ions are created. However, an equilibrium distribution reached under the beam electric potential turns out to be significantly different from the original Gaussian distribution due to the focusing force [21, 22]. Starting from the original Gaussian distribution with an RMS value given by that of the stored beam σ_e,

$$\rho(y) = \frac{1}{\sqrt{2\pi}\sigma_e} \cdot e^{-\frac{y^2}{2\sigma_e^2}} , \tag{27}$$

we find analytically in the linear regime that the ion distribution deforms to a distribution given by

$$\rho(y) = \frac{1}{\pi\sqrt{2\pi}\sigma_e} \cdot e^{-\frac{y^2}{4\sigma_e^2}} \cdot K_0\left(\frac{y^2}{4\sigma_e^2}\right), \tag{28}$$

where $K_0(z)$ is the zeroth-order modified Bessel function of the second kind (Fig. 7).

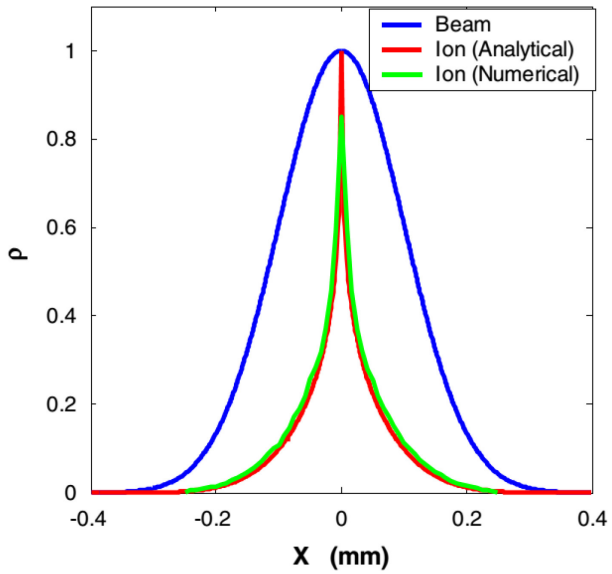


Fig. 7: Transverse distribution of an electron beam (Gaussian) and its corresponding ion distribution. For the ion distribution, a comparison is made between analytical (Eq. 28) and numerical results. Taken from Ref. 22.

Despite this fact, the conventional treatment of assuming a Gaussian distribution for the ions and applying the Bassetti–Erskine formula with the relation $\sigma_i = \sigma_e / \sqrt{2}$ turns out to closely reproduce the electric field created by the ions; a comparison is made in Fig. 8 [22].

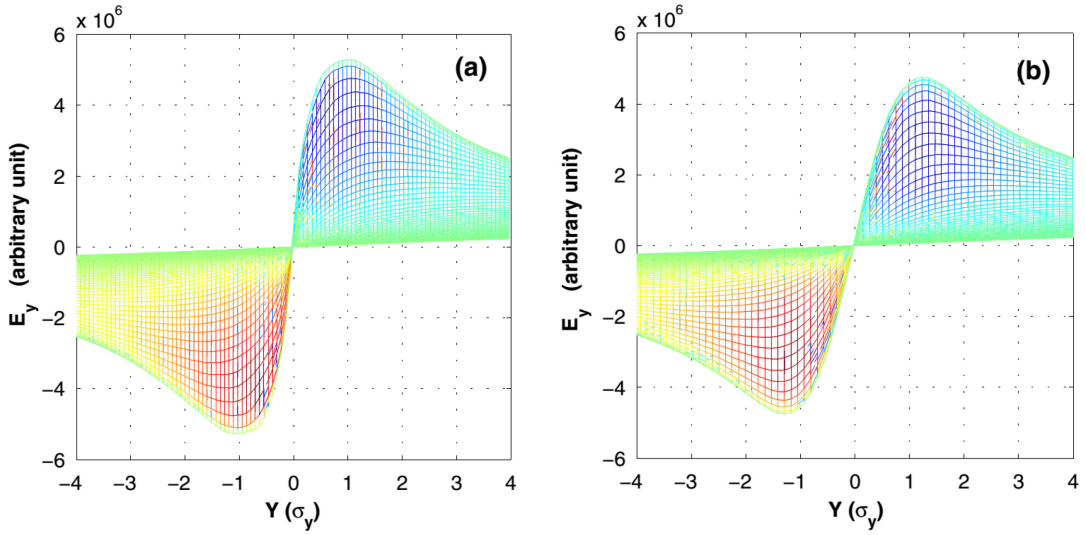


Fig. 8: Vertical electric field E_y of an ion cloud. (a) Obtained numerically via simulation. (b) A Gaussian distributed ion cloud with $\sigma_i = \sigma_e/\sqrt{2}$. The total charge is the same for the two distributions. Different lines in the figures are for different horizontal values of x. The coordinate Y is in the unit of rms beam size. Taken from Ref. [22].

2.13 Lifetime reduction and effective pressure rise due to trapped ions

When ions are trapped, they are populated on the stored beam trajectory whose distributions can be approximated as Gaussian as described previously with $\sigma_{ix} \sim \sigma_x / \sqrt{2}$ and $\sigma_{iy} \sim \sigma_y / \sqrt{2}$. If they are uniformly distributed in a ring, their density may be given by

$$d_{\mathrm{i}} = \frac{2N_{\mathrm{i}}}{\pi\sigma_x\sigma_y L}, \tag{29}$$

where L denotes the ring circumference. Using this localized density around the beam trajectory in the previous beam collision rate formula, we can estimate the lifetime reduction due to trapped ions,

$$\frac{1}{\tau_{\mathrm{ions}}} = \sigma_{\mathrm{Total}} \cdot d_{\mathrm{i}} \cdot \beta c \ . \tag{30}$$

Also, if we apply the relation d_{m} [m^{-3}] = 2.47 × 10^{22} P_{m} [mbar] introduced earlier, we can discuss the effective pressure rise due to trapped ions on the beam trajectory, as given by

$$P_{\mathrm{ions}}\ [\mathrm{mbar}] = \frac{1}{2.47\times10^{22}} \cdot \frac{2\eta N}{\pi\sigma_x\sigma_y L}, \tag{31}$$

where η is the neutralization factor defined by Eq. 9 and N is the total number of stored particles.

3 Two-beam instabilities

In this section, we shall look at beam instabilities that could occur between two beams that co-exist in the same ring, in which one is a circulating beam, and the other an ion beam. In accordance with the history, we shall first go through the case where the ion beam is trapped in the electrostatic potential of the circulating beam, which gave rise to a serious issue in many previous storage rings. We shall then look at the Fast Beam–Ion Instability (FBII), which could take place in a single passage of an electron beam. We attempt to learn about this by going through theoretical, numerical, and experimental studies made by the experts to understand the mechanisms of the instability. Readers interested in two-beam instabilities may also wish to follow the descriptions found in Ref. [23].

3.1 The case of trapped ions

Here we shall follow closely the descriptions made by Sakanaka in Ref. [18], where numerical studies were made to explain the experimental observations in the Photon Factory (PF) ring, running as a synchrotron radiation source at KEK. Indeed, instabilities due to trapped ions were serious issues at PF as well as some of the other light-source rings in the 1990s. With trapped ions, a resonant coupling between the two beams could arise that could lead to an instability and this type of instability was observed in many (second generation) light-source rings. The instability may consequently induce beam pulsation, as shown in Fig. 9, occurring at a frequency of around 20 Hz in the shown example, which is critically detrimental in a light source where the users request perfect stability of the photon beam. As we shall see below, the pulsation is a result of a beam blow-up due to interaction with ions, which is damped via radiation damping during the period when the ions are cleared, which is repeated periodically.

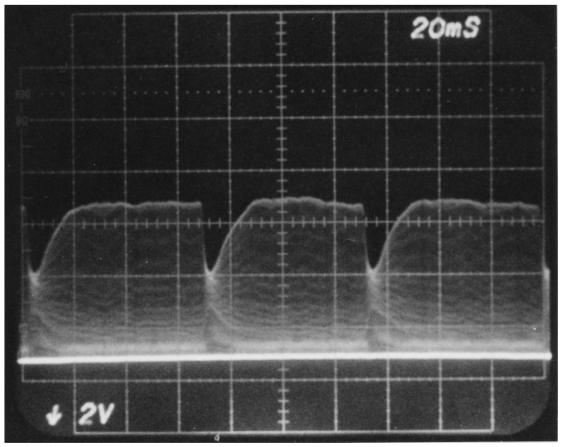


Fig. 9: Vertical beam pulsation observed at Photon Factory, KEK (horizontal unit: 20 ms/div). Taken from Ref. [18].

The beam instability can be described with a simplified model treating the Centre of Mass (CM) oscillations of the two beams [6, 12, 24, 25], by considering vertical oscillations alone, since the two beams interact primarily in this plane due to $\sigma_x > \sigma_y$ as also assumed above in Section 2. The model specifies furthermore that:

- An electron centre of mass y_{cme} oscillating in the ring with the $Q_{\beta y}\omega_0$ (where $Q_{\beta y}$ is the betatron tune) and under a linear force from the ion CM represented by the frequency ω_{e};
- An ion centre of mass y_{cmi} feeling only a linear force from the electron CM represented by the frequency ω_{i};
- Gaussian distributions for the two beams and +1e charge for ions (as in the descriptions above).

The corresponding coupled equations are

$$\begin{aligned} \ddot{y}_{\mathrm{cme}} + Q_{\beta y}^2\omega_0^2\, y_{\mathrm{cme}} &= -\omega_e^2 \cdot (y_{\mathrm{cme}} - y_{\mathrm{cmi}}) , \\ \ddot{y}_{\mathrm{cmi}} &= -\omega_i^2 \cdot (y_{\mathrm{cmi}} - y_{\mathrm{cme}}) , \end{aligned} \tag{32}$$

where

$$\omega_{\mathrm{e}}^2 = \frac{2\lambda_{\mathrm{i}} r_{\mathrm{e}} c^2}{\gamma} \frac{1}{\sigma_y(\sigma_x + \sigma_y)} , \tag{33}$$

$$\omega_{\mathrm{i}}^2 = \frac{2\lambda_{\mathrm{e}} r_{\mathrm{p}} c^2}{A} \frac{1}{\sigma_y(\sigma_x + \sigma_y)} , \tag{34}$$

where

$$r_{\mathrm{e}} = \frac{\mathrm{e}^2}{4\pi\varepsilon_0 m_{\mathrm{e}} c^2} ,$$

is the classical electron radius, and λ_{e} and λ_{i} are the line densities [m^{-3}] of electrons and ions, respectively. Solutions are searched in the form $y_{\mathrm{cme}} = A_{\mathrm{e}} \cdot \exp[\mathrm{i}(n\omega_0 - \omega)t + \mathrm{i}\theta_0]$ and $y_{\mathrm{cmi}} = A_{\mathrm{i}} \cdot \exp(-\mathrm{i}\omega t)$ by introducing a complex frequency ω. Inserting them into Eqs. 32, we get

$$(x^2 - Q_i^2)\cdot\left[(x-n)^2 - Q_y^2 - Q_e^2\right] \;=\; Q_e^2\cdot Q_i^2 \;, \tag{35}$$

where $x = \omega/\omega_0$, $Q_e = \omega_e/\omega_0$, and $Q_i = \omega_i/\omega_0$. If the solution consists of complex numbers, it always appears in the form $a \pm ib$ (where a, b are real), signifying that the two-beam motion is unstable. Numerical studies indicate that instability is likely to appear when n is just above the value of $Q_{\beta y}$.

Let us introduce here a study made at PF by applying the above model [18]. To have an idea of the case concerned, the major ring parameters are listed in Table 2. In Fig. 10 the real solutions at different beam currents are shown as a function of the neutralization factor $\delta = N_i/N$. Instability thresholds as a function of the vertical tune v_y are shown in Fig. 11 for two values of δ for the case of $A = 2$ (H_2^+) and $A = 28$ (CO^+), and are compared with those measured. Comparisons with the experiments shown in the figure indicate that the employed two-centres-of-mass model describes the essential features of the dynamics.

Table 2. Main Photon Factory machine parameters

Parameter	Value
Energy E [GeV]	2.5
Circumference L [m]	187.07
Revolution frequency f_0 [MHz]	1.6026
Momentum compaction α	0.04
Horizontal emittance ε_x [m·rad]	$4.1\,\pi \times 10^{-7}$
Vertical emittance ε_y [m·rad]	$1.2\,\pi \times 10^{-8}$
Horizontal tune $Q_{\beta x}$	5.2–5.5
Vertical tune $Q_{\beta y}$	4.1–4.2
Horizontal damping time $\tau_{\beta x}$ [ms]	9.1
Vertical damping time $\tau_{\beta y}$ [ms]	7.8

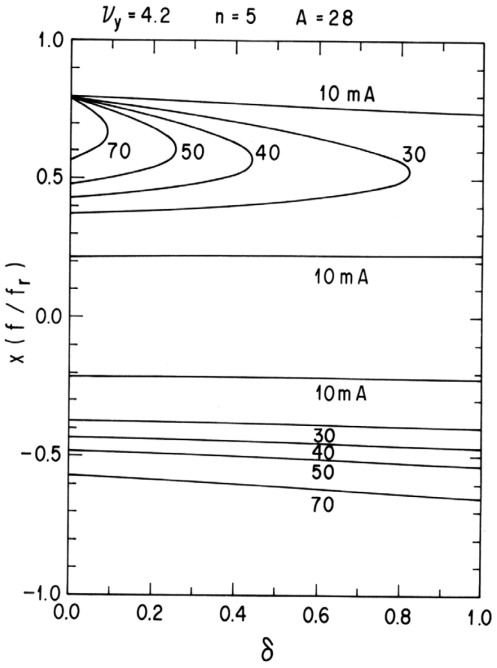


Fig. 10: Real solutions of Eq. (35) at different beam currents as a function of the neutralization factor $\delta = N_i/N$. Taken from Ref. [18].

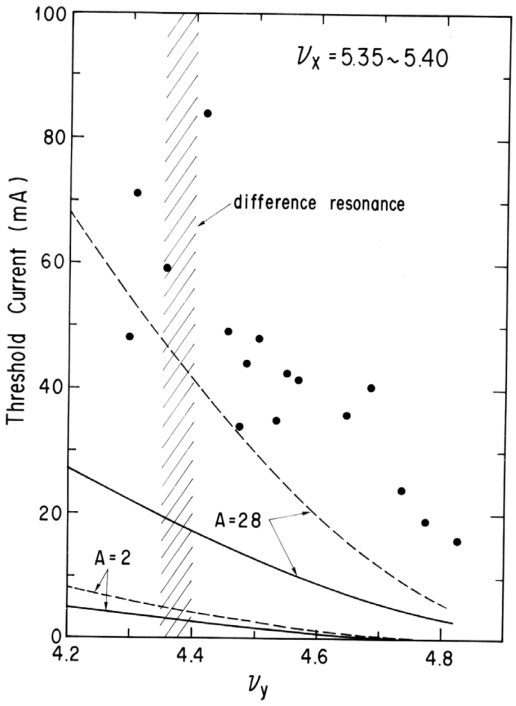


Fig. 11: Instability threshold versus vertical tune obtained from the solution of Eq. (35). Solid lines: $\delta = 1.0$. Dashed lines: $\delta = 0.1$. Black circles: Measured values. $A = 28$: CO^+. $A = 2$: H_2^+. Taken from Ref. [18].

3.2 Fast beam–ion instability

3.2.1 Linear model

We saw that the linear forces between the electron and ion beams represented by ω_e^2 and ω_i^2 depend linearly on the beam intensity and inversely linearly on the product of the transverse beam sizes $\sigma_y(\sigma_x + \sigma_y)$. For modern and future accelerators producing a high intensity and low emittance beam, the 'single-pass' interaction between the two beams may therefore become strong enough to jeopardize performance. In other words, even if ions are not trapped by the electrostatic potential of a stored electron beam, the two streams of beam could interact strongly in a single passage of the circulating beam. Even though the process of interaction for the ion beam is repeated from scratch at every turn, the perturbation upon the electron beam remains and thus could create a kind of resonance, which could eventually render the electron beam unstable. Upon such considerations, a linear theory and simulations were pioneered by Raubenheimer and Zimmermann [26]. This type of two-beam interaction resembles a 'beam breakup in linacs' and does not involve ion trapping. To understand the mechanism and the characteristics of this fast beam–ion instability, let us follow the essential features of the model developed by Raubenheimer and Zimmermann. The transverse coupled linear equations describing such a single-pass interaction of the two beams are given by

$$\frac{\mathrm{d}^2 y_\mathrm{b}(s,z)}{\mathrm{d}s^2} + \omega_\beta^2 \cdot y_\mathrm{b}(s,z) = K \cdot [y_\mathrm{i}(s,t) - y_\mathrm{b}(s,z)] \cdot \int_{-\infty}^{z} \rho(z')\mathrm{d}z' \,, \tag{36}$$

$$\frac{\mathrm{d}^2 \tilde{y}_\mathrm{i}(s,t)}{\mathrm{d}t^2} + \omega_\mathrm{i}^2 \cdot \tilde{y}_\mathrm{i}(s,t) = \omega_\mathrm{i}^2 \cdot y_\mathrm{b}(s,z) \,. \tag{37}$$

Three variables s, z, and t are used to describe the beam motions. A longitudinal position in the ring is specified by s, at which the electron beam may interact with ions. The relative position within the circulating electron beam is specified by z, with the definition $z = 0$ at the centre of the bunch train and extending in between $-z_0$ and $+z_0$, i.e. $-z_0 \le z \le +z_0$ (see Fig. 12). The head of the bunch train is defined as $z = -z_0$. Since the electron beam is assumed to be circulating at the speed of light c, the time variable t can be related to s via $t = (s + z)/c$.

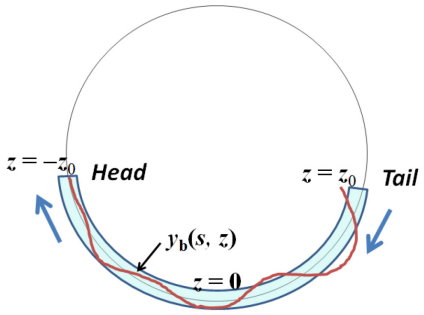


Fig. 12: The model, variables and coordinates assumed in the study of Fast Beam-Ion Instability (FBII) made in Ref. [26].

It must be noted that, in their original paper, the assumed accelerator is not restricted to a ring and could well be a transfer line. Also, the perturbing beam is not limited to ions and could be electrons against a positively charged beam. Reflecting the large difference in the mass between an ion and an electron, however, an ion frequency ω_i generally corresponds to an oscillation period that is much longer than the time spacing between bunches. The interaction between the two beams therefore becomes of a coupled-bunch nature, while in the case of electron clouds, the two-beam instability is usually of a single bunch nature.

In the first equation, $y_b(s,z)$ and $y_i(s,t)$ represent the electron and ion beam centroids, respectively. As in the model in Section 1.1, the non-perturbed motion of an electron beam centroid is a betatron oscillation represented by the oscillation constant ω_β (= $Q_{\beta y}\cdot\omega_0$). Like in the case of ion trapping, its motion is impacted by the electrostatic potential of ions, as represented by the constant K, corresponding to ω_e^2 in Eq. 32, its attractive force depending proportionally on the difference between the two amplitudes. What physically distinguishes this model from the previous one, however, is that, reflecting the single-pass ionization process, the amplitude of the ion perturbation depends explicitly on the number of electrons upstream the concerned electron beam centroid $y_b(s,z)$ at the relative position z, as indicated by the last factor on the right-hand side of Eq. 36. The longitudinal distribution of the electron bunch train (Fig. 12) is denoted by $\rho(z)$, which is normalized to unity.

The second equation (Eq. 37) describes the vertical centroid of a transverse slice of ions $\tilde{y}_i(s,t)$ created at a position s at a certain moment t' ($< -t$) due to the collision of electrons with the residual gases. Reflecting the way they are generated, the initial conditions $\tilde{y}_i(s,\, t' \,|\, t') = y_b(s,z')$ and $\mathrm{d}\tilde{y}_i(s,\, t' \,|\, t')/\mathrm{d}t = 0$ are adopted for a transverse slice. The ion-beam centroid $y_i(s,t)$ that influences the motion of an electron-beam centroid $y_b(s,z)$ consists of all possible ion slices $\tilde{y}_i(s,t \,|\, t')$ with $t' = (s + z')/c$ created at the position s until the time t. This is modelled as a $\rho(z)$-weighted average over z',

$$y_{\mathrm{i}}(s,t) \;=\; \frac{\int_{-\infty}^{z} \mathrm{d}z' + \rho(z')\cdot \tilde{y}_{\mathrm{i}}(s,\; t \mid s+z')}{\int_{-\infty}^{z} \rho(z')\mathrm{d}z'} \;. \tag{38}$$

Having well defined the model, the coupled equations are solved via perturbation expansion in K/ω_β. A great simplification in the mathematical derivation is obtained by assuming a rectangular distribution for $\rho\,(z)$ as shown in Fig. 13, which assumes that there is no variation of the ion frequency ω_{i} along the bunch train. An asymptotic solution is derived in the form

$$y_{\mathrm{b}}(s,z) \;\approx\; \frac{\mathrm{e}^{2\sqrt{\eta}}}{\eta^{1/4}}\;\sin(\omega_{\mathrm{i}} z - \omega_\beta s + \theta - \phi)\;, \tag{39}$$

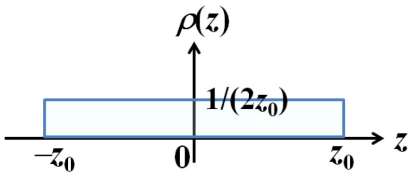


Fig. 13: A rectangular beam density distribution employed in Ref. [26]

where a variable η, given by

$$\eta \;\equiv\; \frac{K\cdot\omega_{\mathrm{i}}\cdot(z+z_0)^2 s}{16\omega_\beta z_0} \tag{40}$$

is assumed to be large ($\eta \;>>\; 1$) and θ and ϕ are constants appearing from the initial phases of oscillations. From the solution obtained above, the asymptotic growth rate evaluated at the tail of a bunch train $z = z_0$ (i.e. where the instability is strongest) is given by

$$\tau_{\mathrm{aymp}}^{-1}(s^{-1}) \;\approx\; \frac{N_{\mathrm{e}}^{3/2} n_{\mathrm{b}}^2}{\gamma}\times\left[5 p_{\mathrm{gas}}\;[\mathrm{Torr}]\;\frac{r_{\mathrm{e}} r_{\mathrm{p}}^{1/2} L_{\mathrm{sep}}^{1/2} c}{\sigma_y^{3/2}(\sigma_x+\sigma_y)^{3/2} A^{1/2}\omega_\beta}\right], \tag{41}$$

where N_{e} is the number of electrons per bunch, n_{b} is the number of bunches, L_{sep} is the longitudinal spacing between bunches, and p_{gas} is the residual gas pressure in Torr. The formula is obtained for a horizontally flat electron beam.

As we see from Eq. (39), the asymptotic growth rate $\tau_{\mathrm{aymp}}^{-1}$ is obtained in the form

$$y_{\mathrm{b}}(s,z_0) \sim \mathrm{e}^{\sqrt{t/\tau_{\mathrm{asmp}}}}$$

and is therefore not an e-folding time. From Eq. (41), we understand that it depends strongly on the number of bunches ($\propto n_{\mathrm{b}}^2$), the number of particles per bunch ($\propto N_{\mathrm{e}}^{3/2}$) and the transverse beam sizes $(\propto\; \sigma_y^{-3/2}\cdot(\sigma_x+\sigma_y)^{-3/2})$. The assumed linear model is supposed to break down when the amplitude of the oscillation $y_{\mathrm{b}}(s,z)$ exceeds the vertical beam size σ_y where the coupling force between the two beam falls off. The growth rate above was evaluated for several existing rings in Ref. [25] (Table 3).

In particular, significantly short growth times result for ALS and the ESRF, i.e. in the light-source rings. However, no clear evidence of FBII was observed for these machines. The following points were therefore discussed as possible explanations.

i) The developed model assumes constant ω_i, whereas these light sources have strongly varying β functions due to adoption of double and triple bend achromat lattices. Namely, the ion frequency ω_i could effectively vary significantly around the ring.

ii) The presence of Landau damping sources such as strong sextupoles and non-zero chromaticity, which these rings generally possess.

iii) Other important nonlinear effects not considered in the linear model.

Table 3. Parameters and the asymptotic growth rates evaluated for several existing rings. Taken from Ref. [26].

Parameter	Accelerator					
	SLC arc	SLC e^+ DR	ALS	HERA e^-	CESR	ESRF
ε_x^N [m]	5×10^{-5}	3×10^{-5}	1.2×10^{-5}	2×10^{-3}	2.7×10^{-3}	7.5×10^{-5}
ε_y^N [m]	5×10^{-6}	3×10^{-6}	2×10^{-7}	1.1×10^{-4}	1.2×10^{-4}	7.5×10^{-6}
n_b	1	1	328	210	7	330
N_b	3.5×10^{10}	4×10^{10}	7×10^9	3.7×10^{10}	4.6×10^{11}	5×10^9
$\beta_{x,y}$ [m]	4	1,3	2.5,4	25	14,13	8,8
$\overline{\beta_y}$ [m]	4	3	4	25	13	8
σ_x [μm]	50	114	101	1000	2000	224
σ_y [μm]	15	62	17	230	400	70
z_0 [σ_z]	1 mm	5.9 mm	100 m	3024 m	335 m	140 m
E [GeV]	46	1.2	1.5	26	5	6
P [Torr]	10^{-5}	10^{-8}	10^{-9}	10^{-9}	5×10^{-9}	2×10^{-9}
Particle species	e^+	e^+	e^-	e^-	e^-	e^-
$\omega_{ion}/2\pi$ [MHz]	4×10^5	5×10^4	25	0.8	0.6	8.3
Single or multi-bunch	Single	Single	Multi-bunch	Multi-bunch	Multi-bunch	Multi-bunch
τ_{asym} [$z \approx z_0$]	1.1 μs	490 μs	2.4 μs	211 μs	3.9 ms	50 μs

3.2.2 Simulation of fast beam–ion instability

In addition to the linear model above, Raubenheimer and Zimmermann developed a simulation code to study numerically the growth of instability as a complementary and more rigorous method. The numerical simulation using macro-particles to represent the two beams (i.e. the strong–strong model) has indeed the large advantage of being able to integrate nonlinear effects, such as those due to finite beam sizes, and the capacity to follow self-consistently and dynamically the evolution of bunch distributions of the two beams. In the developed scheme, the ionization processes via the beam–residual gas collisions were simulated by using a specific ionization cross-section and partial pressure of a gas species to generate ions. The space-charge forces of each of the two beams were then calculated and applied to macro-particles of the opponent beam. The cascading process of ions growing in number due

to the successive arrival of electron bunches at the interaction point is rigorously treated. All ions at the end of each beam passage are discarded, assuming an ion-clearing beam gap from turn to turn. More specifically, the main features of the developed simulation are given below:

- Motions of macro-particles described with coordinates (x, x', y, y', $\delta E/E$);
- Beam bunches are initially Gaussian, longitudinally and transversely;
- Collision with gas takes place at some specified points in a ring (or a linac);
- A beam bunch is divided into ~five slices longitudinally;
- Each macro-particle is free to move in x and y according to the E-fields, but fixed in z;
- Two-dimensional grids (e.g. 25 × 25) with relation to the centre of mass of each slice introduced;
- At each grid:
  - Ions are created according to the specified pressure and collisional cross-section;
  - Ions have zero initial velocity;
  - The E-field of the beam is evaluated with the Bassetti–Erskine formula and applied to the ions;
  - Ion density and ion-induced E-field is calculated and applied to beam macro-particles.

Although the obtained simulation results showed certain dependence of the growth rate on the initial conditions of the beam, the calculated growth rates agreed well with the predicted asymptotic growth rates. Some of the typical simulation results taken from Ref. [26] are shown in Fig. 14.

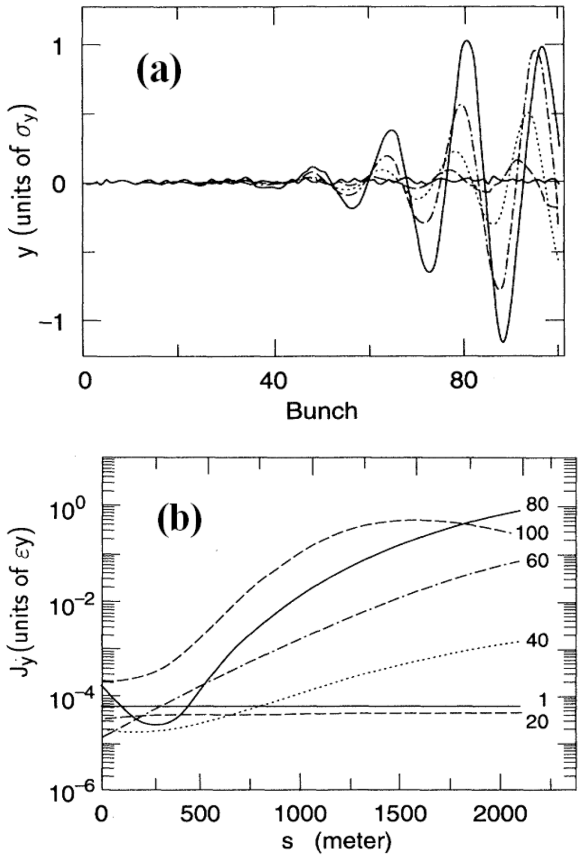


Fig. 14: (a) Growing transverse amplitude of different bunches in a bunch train at several instants. (b) Growth of the transverse action variable J_y versus s (time) for several bunches in a bunch train. Taken from Ref. [26].

Macro-particle simulations of FBII can generally be quite time consuming, especially as the physical process of the collisional ionization is intrinsically sequential and cannot be parallelized. If the bunch distribution of the electron beam can be assumed to not change through its interactions with ions, the beam bunch can be treated as a rigid object (i.e. one macro-particle). Such a model, conventionally called the weak–strong model, can bring about a great simplification and reduction in CPU time [27]. It allows the integration of other physically important ingredients into the simulation, such as transverse bunch-by-bunch feedback and/or the effect of coupling impedance of a ring [28]. The impact of transverse feedback fighting against FBII to stabilize an electron beam is simulated in Fig. 15 using a weak–strong code developed by Xia *et al*. [29].

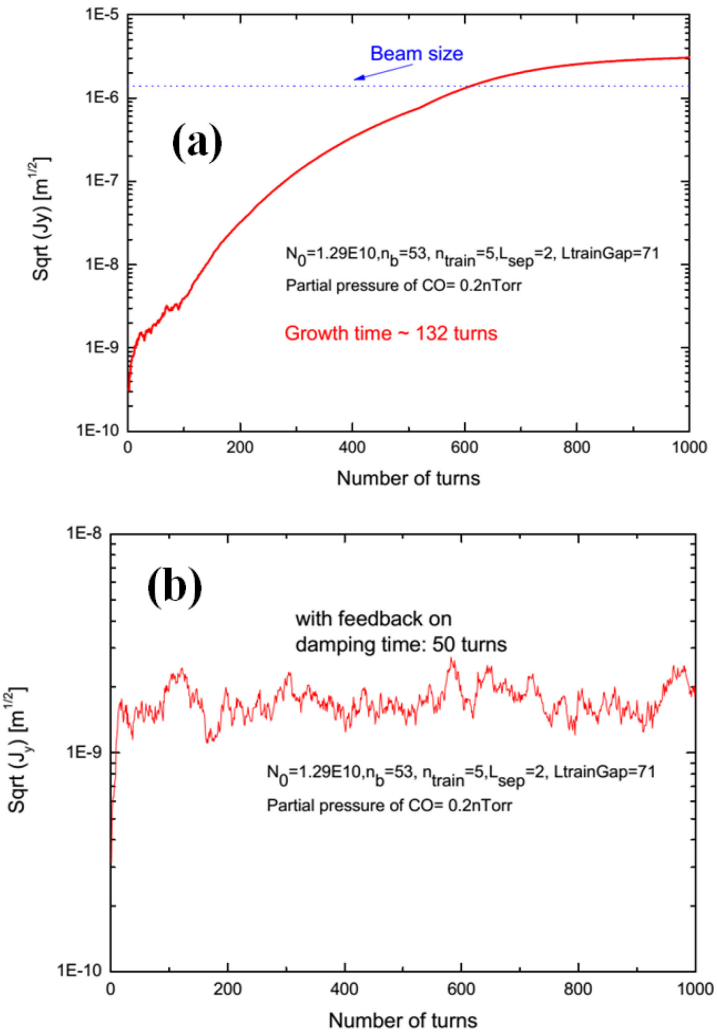


Fig. 15: Simulation of transverse amplitude growth (square root of the transverse action variable J_y) versus time (turn number) without (a) and with (b) bunch-by-bunch transverse feedback. Taken from Ref. [29].

3.2.3 Experimental studies of fast beam–ion instability

The phenomenon of FBII was experimentally demonstrated in ALS [30], PLS [31], and KEK [32] by artificially increasing the vacuum pressure in a ring. Let us follow the major outcomes from the first such attempt made at ALS [30]. As basic conditions, He gas was injected into the ring to increase the vacuum pressure to attain 80×10^{-9} Torr and bunch-by-bunch transverse feedback was switched on to stabilize beam against conventional beam instability such as that due to coupling impedance. A comparison of the vertical beam size was then made to the nominal pressure case as a function of the length of a bunch train, leaving always a large beam gap of more than 80 buckets. A steady increase of

the vertical beam size was observed until the bunch train reached some 15 bunches and then saturated above (Fig. 16). The evolution of the coherent signals exhibited by the beam was followed as a function of its intensity, where the measured peak frequencies turned out to well reproduce what was expected from the theory (Fig. 17). Another interesting observation was the beam current distribution along the bunch train after inserting a vertical scraper to scrape off a vertically blown-up beam (Fig. 18). As expected from the theory, the intensity tended to decrease from the head to the tail of a bunch train. All of these observations well confirmed the characteristics of FBII. Studies made at PLS and KEK gave equally good agreement with theory.

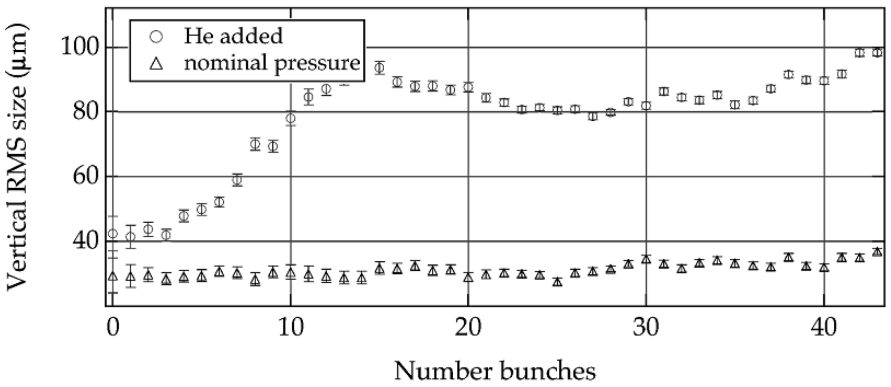


Fig. 16: Measured vertical beam size as a function of the length of a bunch train for two different values of vacuum pressure. Taken from Ref. [30].

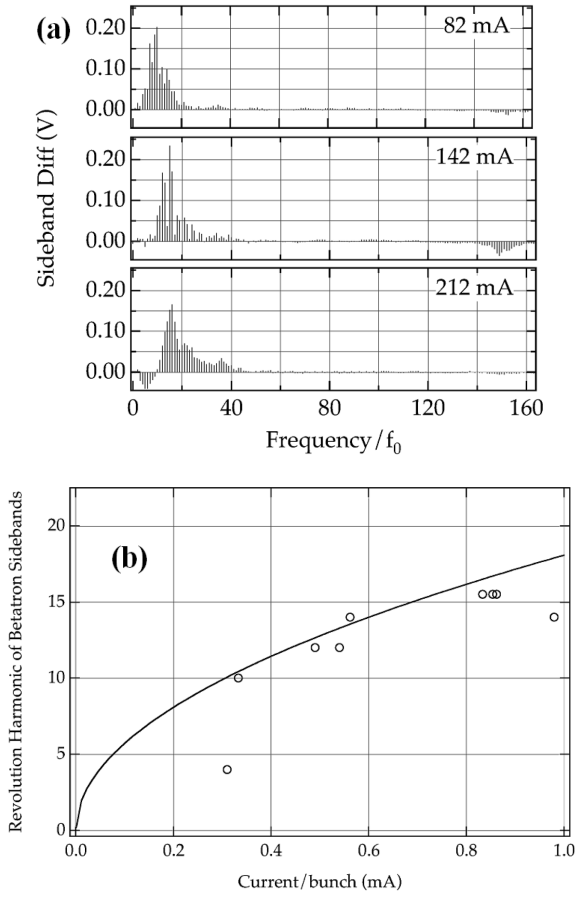


Fig. 17: (a) Measured spectra of coherent vertical beam oscillations at three different beam intensities. (b) Comparison of measured coherent frequencies of beam oscillations with theory. Taken from Ref. [30].

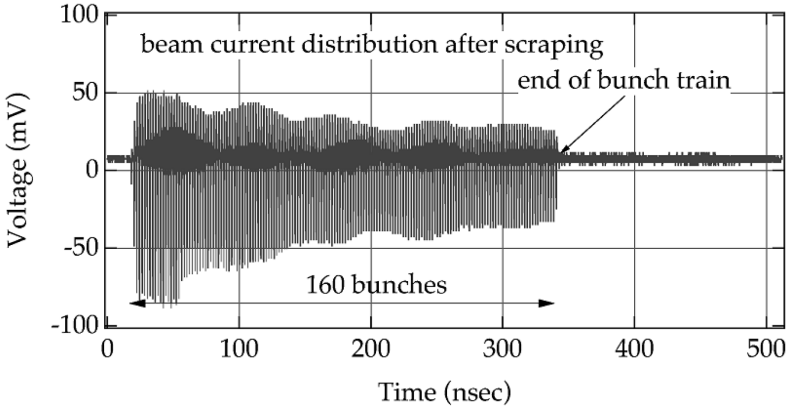


Fig. 18: Measured vertical amplitude along the bunch train after insertion of a vertical scraper. Taken from Ref. [30].

3.2.4 Effect of ion decoherence in growth rate evaluation

In the rest of this section (Subsections 3.2.4 and 3.2.5), let us look at some further theoretical efforts made by several groups to better describe FBII. One that we are going to see here is the influence of ion frequency spread, which exists in reality, but was ignored in the linear model introduced in Subsection 3.2.1. As is known, a spread in the frequency generally helps to reduce the instability growth rate via Landau damping. There are at least three sources of ion frequency spread.

i) Due to the electron beam density variation that may arise from relative transverse displacement of the two beams. As can be seen from Eq. (34), the ion frequency depends on the square root of the electron beam density. So if the transverse extension of an electron beam was finite, as usually represented with a Gaussian distribution, the density of electrons varies according to where an ion probes it.

ii) Amplitude-dependent frequency shift due to the nonlinearity of the static potential of the electron beam.

iii) Electron beam size variations along the ring arising from beta function changes. Again, the dependence of the ion frequency on beta functions can be confirmed from Eq. (34).

To elucidate the effect of ion decoherence analytically, Stupakov *et al*. [33] introduced a distribution function $f(\omega_i)$ on the ion frequency in the previous linear model that we saw in Subsection 1.2.1, and used it to average over all possible transverse ion slices to get the ion beam centroid $y_i(s,t)$ in Eq. (38), with the normalization

$$\int f(\omega_i)\mathrm{d}\omega_i \;=\; 1 \;. \tag{42}$$

Proceeding in an analogous way to solve the coupled linear equations as in the earlier model (cf. Eqs. 36 and 37), we arrive at the following equation for the electron beam centroid $y_b(s,z)$:

$$\frac{\partial^2 y_b(s,z)}{\partial s^2} \;+\; \frac{\omega_\beta^2}{c^2}\cdot y_b(s,z) \;=\; -\frac{K}{2z_0}\int_0^z z'\;\frac{\partial y_b(s,z')}{\partial z'}\;D(z-z')\;\mathrm{d}z' \;, \tag{43}$$

where $D(z-z')$ is named as a decoherence function given by

$$D(z-z') \;=\; \int \mathrm{d}\omega_\mathrm{i} \cdot \cos[\frac{\omega_\mathrm{i}}{c}(z-z')] \cdot f(\omega_\mathrm{i}) \;. \tag{44}$$

The physical picture taken here is that the beam–ion instability develops on a time scale that is much larger than both the betatron and ion oscillation periods ($K \ll \omega_\beta^2,\ \omega_\mathrm{i}^2$), which is true in most cases. The above justifies looking for a solution in the form

$$y_\mathrm{b}(s,z) \;=\; A(s,z) \cdot \mathrm{e}^{-\mathrm{i}\omega_\beta s/c + \mathrm{i}\omega_{\mathrm{i}0} z/c} \;, \tag{45}$$

where $\omega_{\mathrm{i}0}$ is the central ion frequency. In the absence of frequency spread $f(\omega_\mathrm{i}) = \delta(\omega_\mathrm{i} - \omega_{\mathrm{i}0})$, the decoherence function becomes unity and the solution $A(s,z)$ is confirmed to consistently reduce to the asymptotic solution found in Subsection 1.2.1. In Ref. [33], the decoherence functions in the exponential form were explored to be able to derive analytically the solutions $A(s,z)$ for the first two sources of the frequency spread given above. Analytical results were found to be in good agreement with macro-particle simulations, where the treated ion tune spreads caused a reduction of the instability growth rate by roughly a factor of 2.

3.2.5 Wake function description of an ion cloud

The second theoretical development we shall look at is about modelling a fast beam–ion interaction with a transverse dipolar wake function or, equivalently, the machine-coupling impedance that is its Fourier transform, conventionally used to describe the interaction of a beam with its surrounding vacuum chambers. A first such attempt was in fact made in describing a positron coasting beam driven unstable by an electron cloud [34]. An analogous treatment extending the idea to ion clouds interacting with bunched electron beam was then made by two groups [22, 35]. Here let us briefly follow the work of Wang *et al*. in Ref. [22]. There is initially an ion cloud formed by collisions of electrons with residual gases, consisting on N_i ions. If there then comes an electron bunch composed of N_e electrons with its centre of mass deviated by $\Delta y_{\mathrm{e}0}$ from the centre of the ion distribution, assumed to be described with a Gaussian, the ion distribution gets a kick from the electron bunch and starts oscillating coherently. The oscillating ion distribution can in turn give a kick $\Delta\overline{y}_\mathrm{e}'(s)$ to another electron bunch following the first bunch at a distance s behind. If the initial displacement $\Delta y_{\mathrm{e}0}$ is small enough and corresponds to the linear part of the ion distribution, all of this process can be followed analytically using the relations derived in previous sections. In addition, we find that $\Delta\overline{y}_\mathrm{e}'(s)$ is proportional to $\Delta y_{\mathrm{e}0}$ so that the transverse dipolar wake excited by a bunch of electrons is defined by

$$W_y(s) \;\equiv\; \frac{\gamma}{N_\mathrm{e}\varepsilon_0 r_\mathrm{e}} \cdot \frac{\Delta\overline{y}_\mathrm{e}'(s)}{\Delta y_\mathrm{e}} \;[\mathrm{V C^{-1}\,m^{-1}}], \tag{46}$$

which does not depend on $\Delta y_{\mathrm{e}0}$, namely it satisfies the linear response condition. The linearity of the wake function $W_y(s)$ as defined above was numerically confirmed, as shown in Fig. 19, where the wakes excited by various values of $\Delta y_{\mathrm{e}0}$ are seen to converge on the same values.

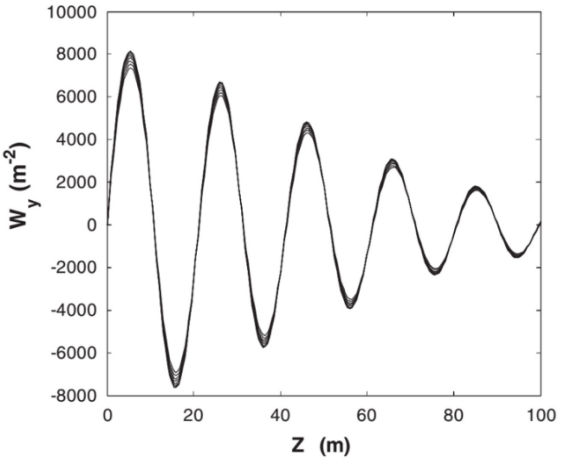


Fig. 19: Simulated wakes of an ion cloud for various Δy_{e0} ranging from one tenth to one sigma with a step of one tenth of a sigma. Taken from Ref. [22].

Also noticed for $W_y(s)$ (Fig. 19) is the damping of the oscillation, which is due to the nonlinearity of the beam–ion space charge force, inducing an ion frequency spread. Wang *et al*. managed to parameterize the wake function in the usual broadband resonator function as

$$W_y(s) \;=\; \hat{W}_y \cdot \exp\{-\omega_i s/(2Qc)\} \cdot \sin\left(\frac{\omega_i s}{c}\right) \tag{47}$$

where

$$\hat{W}_y \;=\; \frac{N_i}{\varepsilon_0}\left(\frac{r_p L_{sep}}{A N_e}\right)^{1/2} \cdot \left[\frac{4}{3}\cdot\frac{1}{\sigma_y(\sigma_x+\sigma_y)}\right]^{3/2} \;[\mathrm{VC^{-1}\,m^{-1}}], \tag{48}$$

and the Q values close to 9 were found to reproduce well the numerically obtained wakes of most of the ions and electron beam sizes (Fig. 19). The impedance function corresponding to $W_y(s)$ in Eq. (47) is therefore given by

$$Z_y(\omega) = \frac{\hat{W}_y}{\omega}\cdot\kappa\cdot\frac{Q}{1 \;+\; \mathrm{i}Q\left(\dfrac{\omega_i}{\omega}\dfrac{1}{\kappa} \;-\; \dfrac{\omega}{\omega_i}\kappa\right)} \;\approx\; \frac{\hat{W}_y}{\omega}\cdot\frac{Q}{1 \;+\; iQ\left(\dfrac{\omega_i}{\omega} \;-\; \dfrac{\omega}{\omega_i}\right)}\,, \tag{49}$$

where $\kappa = \sqrt{1-1/(2Q)^2} \approx 1$.

The advantage of describing the beam–ion dynamics of a wake or an impedance function is to be able to use the same linearized Vlasov equation formalism developed for conventional coupled-bunch instabilities. In Ref. [22], an application was made to study the stabilization effect of beam gaps introduced in a bunch train.

4 Mitigations and observations of ion effects

In the first part of this section, we shall briefly review various mitigation techniques developed and applied in the past against ions. Three examples of experimental studies of beam ion instabilities shall

then be introduced. The first example is about ion trapping at the Photon Factory (KEK), and the two others are on FBII (SPEAR3 and SOLEIL).

4.1 Mitigations of ion effects

4.1.1 Partial beam fillings/multi-bunch trains/bunch gaps

Partial beam fillings have long been known to be one of the most efficient methods for avoiding ion trapping, and have been constantly applied in the storage rings of today [10, 14, 36]. One can intuitively understand that, during a beam gap, ions see no electrostatic potential of the beam and therefore they can drift away. Numerically, this would correspond to not finding a stable linear solution in the transfer matrix approach explained in Section 2.9 in working with an extended one-turn matrix that integrates a drift matrix that represents the beam gap [37].

As already mentioned, FBII may arise even in the presence of a large beam gap since ions could be generated during a single passage of a bunch train and strongly interact with it in a resonant manner from turn to turn. In such a case, cutting a long bunch train into small pieces and introducing small bunch gaps between the short trains generally helps to reduce the FBII growth rate. Such a study was made by Wang *et al* using the wake function formalism explained in Subsection 3.2.5 (Fig. 20) [22]. An exception to this physical picture has been observed at SOLEIL where, contrarily, partial fills enhance FBII. We shall see this phenomenon in more detail in Subsection 4.2.3.

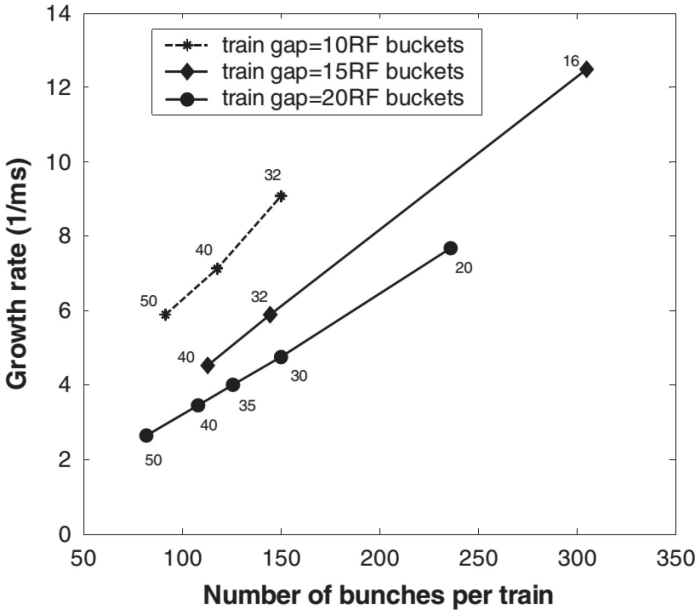


Fig. 20: Calculated growth rate of FBII for different beam fillings, as a function of the number of bunches per bunch train. Three different train gaps of 10, 15, and 20 buckets are considered. The number of bunch trains is also shown in the plot. Taken from Ref. [22].

4.1.2 Ion-clearing electrodes

Introducing a pair of electrodes as shown in Fig. 21 to clear the trapped ions away is also a method known to be effective in reducing ion effects [13, 16, 37]. This worked successfully in several rings such as SRS (Daresbury) [38], ISR (CERN) and Aladdin (University of Wisconsin) [39]. As we saw in Section 2.5, it is efficient to install the electrodes where the beam potential is at a local minimum so that the ions drift and are gathered. To expel ions, the voltage on the electrodes needs be higher than the electrostatic potential of the beam, which for electron beams is typically in the hundreds of volts range (cf. Fig. 3), while for proton beams, it can reach a range of some kilovolts. A too-high voltage may increase the risk of bombarding the accelerated ions on the chamber walls, which in turn could induce

outgassing as observed in proton rings [7, 17]. Ion-clearing electrodes have the disadvantage of increasing the broadband impedance of a ring, decreasing the single-bunch instability thresholds or inducing local machine heating. Optimization of the electrode design should be done in advance to minimize the impedance [40].

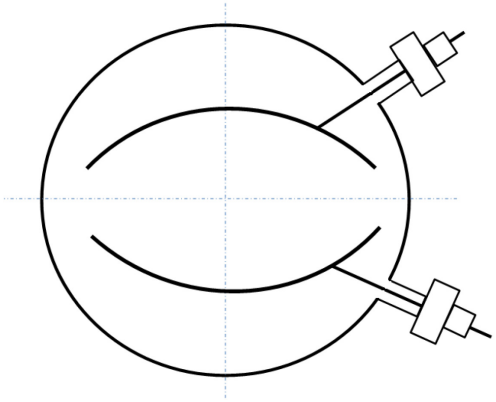


Fig. 21: An illustration of ion clearing electrodes

4.1.3 Positron beam storage

In a ring where a stored electron beam is under the influence of trapped ions, replacing it with a positron beam should resolve the problem of ion trapping, as the positron beam will not attract the ions, which have the same positive charge. In addition, if we recall the discussions of the critical mass in Section 2.9, electrons, which have the opposite charge, are not likely to be trapped by the positron beam since the electron mass is so small. For this reason, a number of lepton storage rings, such as DCI, ACO, SUPERACO, Photon Factory, APS and PETRA-III [15, 41–43], have operated quite successfully with positron beams. However, when the intensity of a stored beam gets high, an electron cloud may be formed via synchrotron radiation generating photo-electrons and lead to electron-cloud instability. Historically, the electron-cloud instability was identified at Photon Factory (KEK) as one of the earliest observations upon changing the stored beam from electrons to positrons to avoid ion trapping [44].

4.1.4 Use of octupoles/chromaticity shifting

Octupole and sextupole magnets can create betatron tune spreads in an electron beam either via an on-momentum amplitude-dependent tune shift or by an off-momentum tune shift with non-zero chromaticity, which can Landau damp ion instability. At Photon Factory (KEK), the combination of the use of octupoles and a partial fill, as explained above, managed to completely suppress ion instability [18]. Since these nonlinear elements simultaneously reduce the dynamic acceptance of a ring in general, however, one needs to evaluate in advance optimal strengths for these elements.

4.1.5 Radio-frequency knockout

Shaking a beam that is under the influence of trapped ions with an external RF field may have the effect of chasing the ions away. Such attempts were made at UVSOR and Photon Factory where the beams were shaken with frequencies in the megahertz range [12, 13]. However, shaking a stored beam may not be an optimal solution in light sources in which excellent beam stability is usually required.

4.1.6 Transverse bunch-by-bunch feedback

With the performance of feedback systems available on the market today, this would be the best method to stabilize a beam against ion instability, whether it is due to ion trapping or FBII. As long as the two-beam instability concerns interaction between the centres of mass as we have described in Section 3, and that the instability growth rate does not exceed the feedback damping time, which appears to be true

in most cases, transverse bunch-by-bunch feedback should be able to stabilize each bunch separately. Experience gained at ESRF and SOLEIL confirm this feature [45, 46]. In both rings, the ion frequencies are typically in the range of a few tens of megahertz, which is low compared to the feedback bandwidth of 176 MHz for these machines. The beam oscillations can therefore be corrected relatively easily. An example is shown in Fig. 22. For future accelerators, however, higher feedback performance may be required to fight against FBII with extremely short growth times. Again at SOLEIL, an exceptional case of feedback turning out to be destructive under extreme situations has been observed, which will be described in more detail in Subsection 4.2.3.

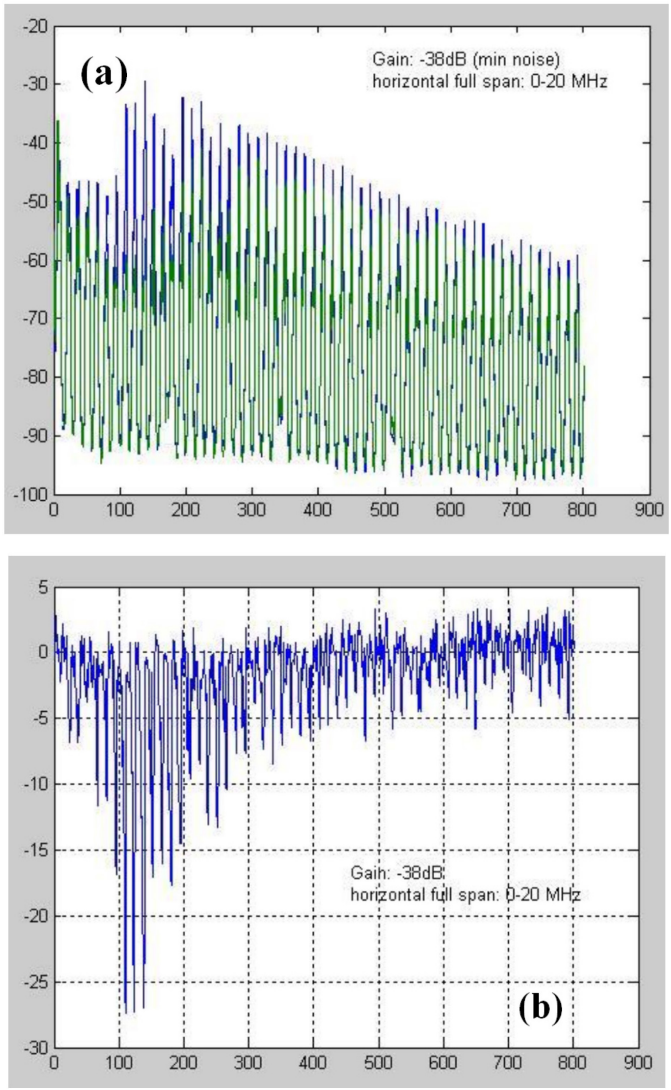


Fig. 22: Vertical beam spectra with (green) and without (blue) transverse bunch-by-bunch feedback in the uniform filling measured at the ESRF. (a) Total amplitude. (b) Difference in amplitude between the two cases. Horizontal axis spans 0-20 MHz. Taken from Ref. [45].

4.1.7 Reduced vertical beam size (by more than a factor of two)

Through theoretical studies of FBII, an interesting idea has emerged that if an electron beam can only blow-up vertically by roughly a factor of two and never get lost due to saturation of the two-beam interaction, one may reduce the vertical beam size by a factor of two in advance by taking blow-up into account [47]. More studies may be needed to fully certify the absence of residual beam blow-up in the saturation regime before employing such a scheme.

4.1.8 Enhancing vertical beta function variations

As we saw in Subsection 3.2.1, the absence of FBII in many modern light sources suggests sources of stabilization in these machines, among which the strong variations of beta function are suspected as they induce ion frequency spreads. Going further along this direction, we may actively enhance vertical beta function variation to stabilize the beam via Landau damping. Again, more numerical studies may be required to certify the damping mechanism quantitatively.

4.2 Observations of ion effects

4.2.1 Experimental characterizations of ion trapping at Photon Factory

A unique and interesting set of measurement have been made at Photon Factory (KEK) that elucidate in more detail the dynamics of the two-beam instability induced by trapped ions. We have already seen in Fig. 9 (Section 2.1) that the instability gives rise to a pulsation of the stored beam with a frequency in a range of some tens of hertz, which is, in fact, particularly disturbing to synchrotron light users. First, the dependence of the pulsation frequency on the vacuum pressure was investigated by controlling the number of Distributed Ion Pumps (DIPs) activated (Fig. 23). A clear trend can be seen in which the frequency elevates as the vacuum pressure gets higher [48].

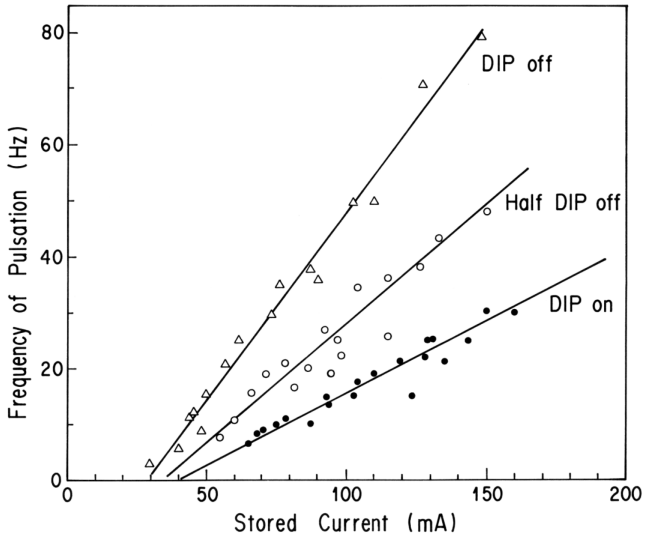


Fig. 23: Dependence of the measured frequencies of the vertical beam pulsation on beam intensity and vacuum pressure. Taken from Ref. [18].

To have a deeper insight into the cycle of pulsation, the bremsstrahlung count rate was followed for the different conditions of the vacuum pressure above [49, 50]. We note here that when there are trapped ions, circulating electrons collide with the nuclei of ions and produce γ rays. Relative changes of the count rate in time would reflect the variation of ions at a local point in the ring, as a change in the vertical size of an electron beam itself would not alter the collision rate with residual gases. The time evolution of the count rate was measured under two different conditions (Fig. 24): both the vacuum pressure and the beam current are lower in Fig. 24(a) than in Fig. 24(b). In Fig. 24(a) we can observe two slopes for the count rate, namely one that corresponds to when the blown-up electron beam is radiation damping and attracting more ions; and another where a slower increase of ions after the electron beam has converged to its original size. The second process is supposed to continue until an instability threshold in the ion density is reached. In Fig. 24(b), on the other hand, we cannot distinguish the two slopes, and the instability threshold is seemingly reached rapidly during the process of radiation damping.

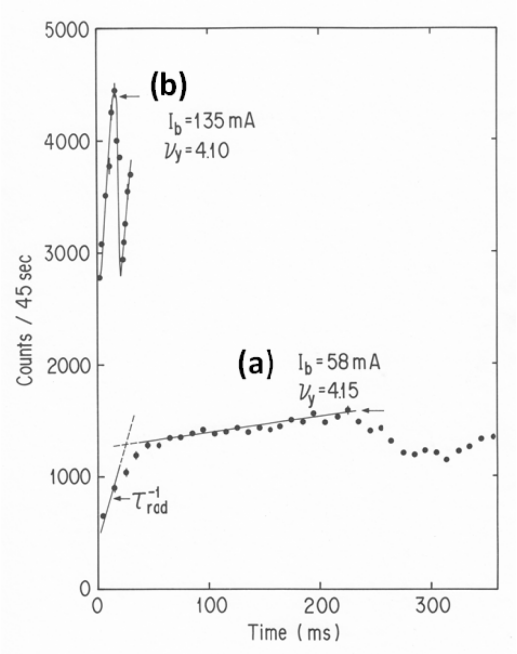


Fig. 24: Time evolution of the Bremsstrahlung count rate measured at two different beam intensities. See text for the different experimental conditions employed for (a) and (b). Taken from Ref. [18].

The bremsstrahlung count rate was compared between the uniform and a (2/3-like) partial filling under the same vacuum condition and the beam current (Fig. 25). The clearly higher rate observed for the uniform filling should signify a larger number of ions trapped in this filling mode.

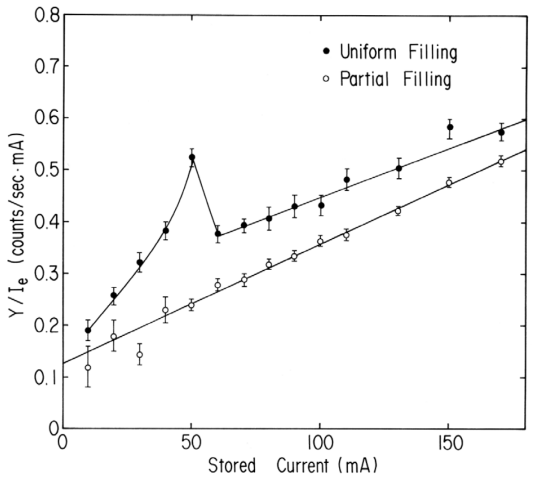


Fig. 25: Measured Bremsstrahlung count rate versus beam intensity for two different beam fillings. Taken from Ref. [18].

4.2.2 Experimental characterizations of fast beam–ion instability at SPEAR3

A systematic characterization of FBII was made at SPEAR3 (SLAC) by measuring the vertical betatron sideband signals over the multi-bunch frequency band under different beam and machine conditions [51]. In their work, these experimental results are also compared with those expected from theory using the wake formalism explained in Subsection 3.2.5. Three such sets of measurement are cited here. The first one concerns the dependence on the vertical beam size, which is adjusted with skew quadrupoles. Without them, the beam size is about 2.3 times larger. As we expect from theory (cf. Eqs. (32–34)), the two-beam interaction gets significantly enhanced as the beam size is reduced (Fig. 26).

The dependence on the bunch filling pattern is shown in Fig. 27. In all three cases shown, the total number of bunches (= 280), the bunch gap (= 15 buckets) and the total beam current (= 500 mA) are kept equal. The comparison clearly indicates the advantage of filling a beam in many short trains of bunches in fighting against FBII. The third measurement concerns the dependence on the vertical chromaticity, which is increased from 2 to 7 (Fig. 28). Again, as expected the chromaticity helps suppression of FBII, presumably through an increased tune spread of the electron beam, though it is clearly correlated with lifetime reductions as indicated in the figure.

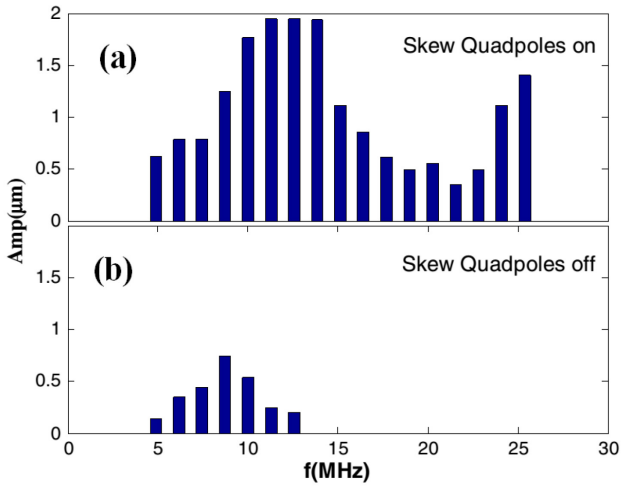


Fig. 26: Measured vertical beam sizes of a single bunch-train (280 bunches) at 192 mA with (a) and without (b) skew quadrupoles. When skew quads were switched off (b), the vertical beam size became roughly 2~3 times larger. Taken from Ref. [51].

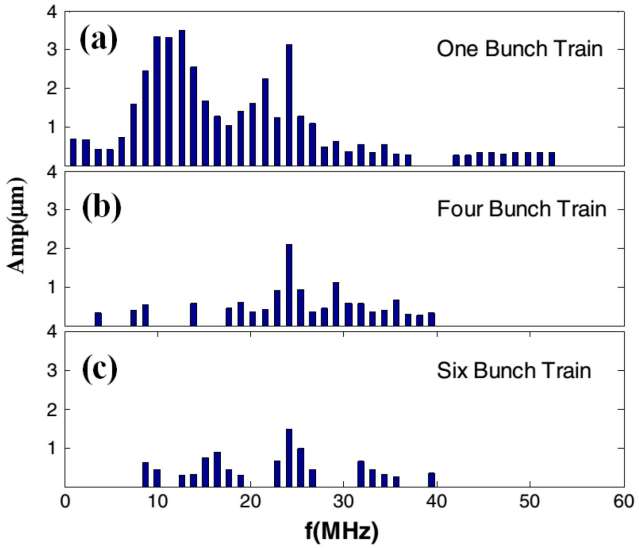


Fig. 27: Measured vertical amplitudes of a stored beam for different fillings. In all cases, there are 280 bunches and the total current is 500 mA. The bunch train gap is 15 buckets (32 ns). Taken from Ref. [51].

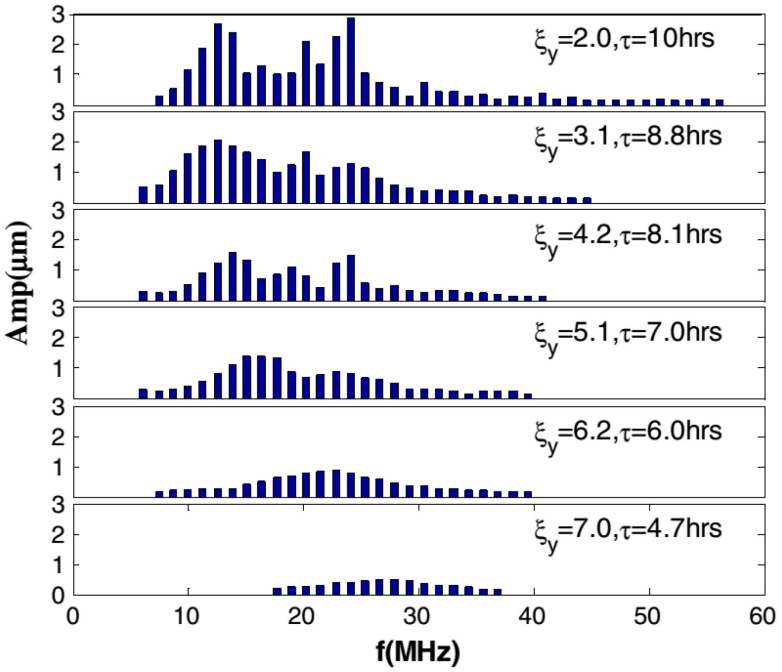


Fig. 28: Measured vertical amplitudes of a stored beam versus vertical chromaticity. Single bunch-train (280 bunches) at 500 mA. Horizontal chromaticity is kept at 2. Taken from Ref. [51].

4.2.3 Beam losses due to a combined effect of fast beam–ion instability driven by beam-induced outgassing, resistive-wall instability, and transverse feedback at SOLEIL

We shall describe below the beam losses, which are often total losses, encountered at SOLEIL at high beam current. After a series of experimental and numerical investigations, these losses were identified as being due to FBII that induces a somewhat complicated combined effect involving Resistive-Wall instability (RW) and transverse feedback [46, 52]. A noteworthy associated feature is that the FBII does not arise from the ordinary vacuum pressure, but occurs uniquely due to localized outgassing of vacuum chambers that are heated by the circulating beam via wake fields.

The fact that multi-bunch operation at SOLEIL is under the influence of a mixture of resistive-wall and ion instabilities was known since the time of commissioning through the analysis of data available from the bunch-by-bunch transverse feedback diagnostics. Transverse feedback was switched off over a short period of time (usually around 1 ms) to let the beam blow up, and the bunch-by-bunch data were acquired over this period of time. The results typically showed that at relatively low current, the beam is under the influence of RW, as seen from the amplitude and phase relations in a bunch train (Figs. 29(a) and 29(b)). However, above a certain current, which is roughly 100 mA in Fig. 29, there is a transition to FBII. In particular, both the bunch-to-bunch betatron phase variations of ~0.9° and ~40° measured (Fig. 29(b)) are in good agreement with what expected from RW and FBII instabilities, respectively. The measured growth rate versus current, averaged over bunches, follows well the curve expected from the RW instability, but with larger error bars at high current, suggesting the nature of mixture of the two instabilities (Fig. 30).

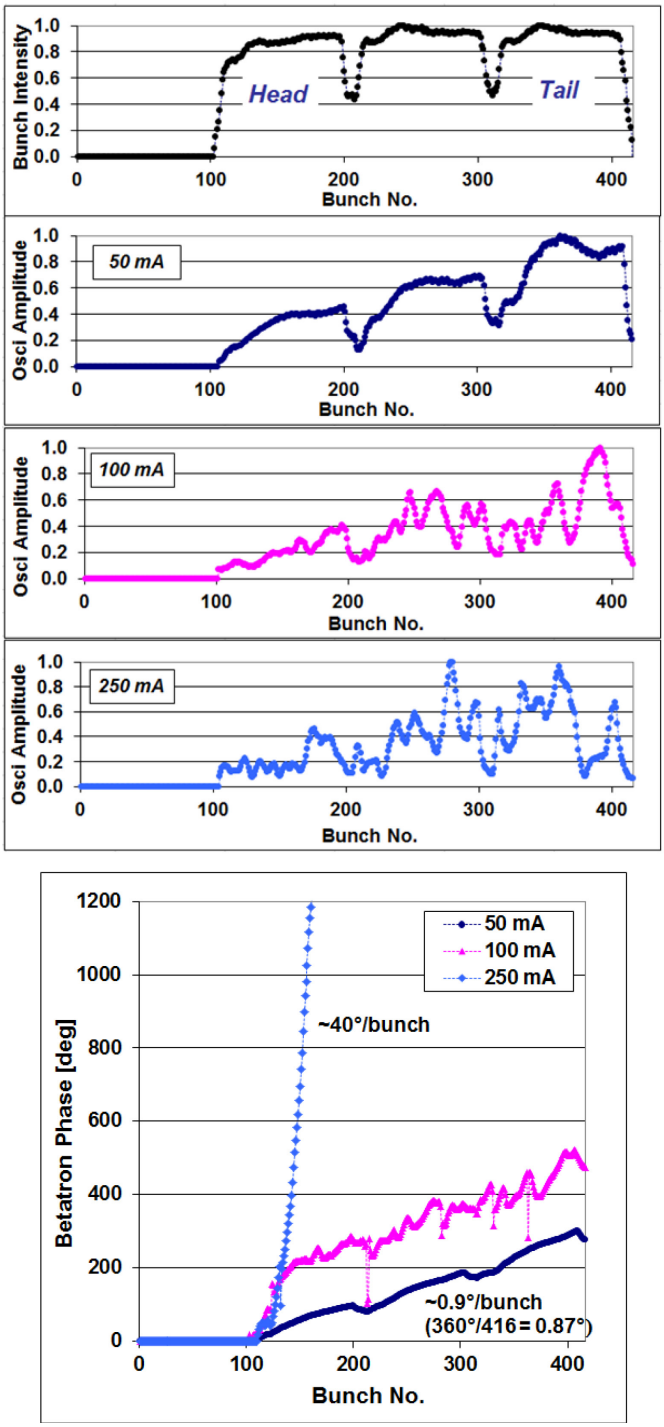


Fig. 29: (a) Vertical amplitude versus bunches in ¾ filling measured as a function of beam current. (b) Measured relative betatron phase with respect to adjacent bunches in a bunch train in correspondence to the cases in (a). Taken from Ref. [53].

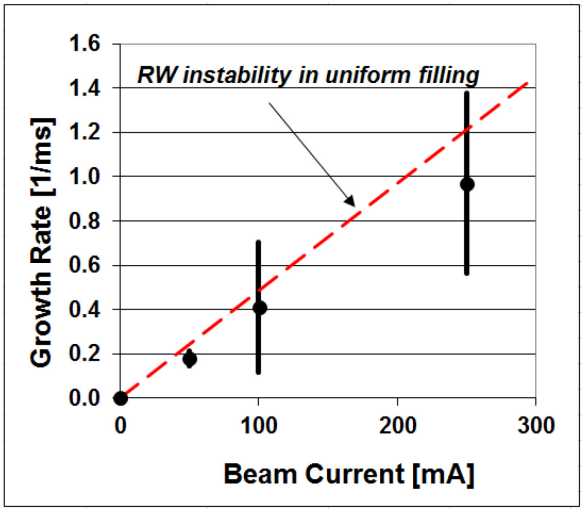


Fig. 30: Growth rate of vertical instability as a function of beam current measured corresponding to the cases in Figs. 29 [53].

Although vacuum conditioning with beam over the years helped to reduce the relative contributions of FBII at a given beam current, as confirmed by re-performing the measurement as described above, FBII still persistently exists at SOLEIL at the nominal current of 500 mA, after nearly 10 years of operation. As already mentioned it often causes beam losses, which strangely happen some ten minutes after ramping the current to its final value. During this period, the beam is diagnosed as being stable. To avoid the beam losses, a number of different beam fillings were tried under the assumption that beam gaps are effective against FBII. However, the experimental results indicated that, to the contrary, the uniform filling gives the most stable beam. This, along with the finding that reducing the RF voltage greatly helps increasing the beam stability, led one to realize that the source of FBII is the beam-induced heating of vacuum chambers via longitudinal wake fields, which in turn triggers outgassing. Thus, keeping the bunch current low and the bunch length long to avoid heating becomes of primary importance. Also, the beam losses were understood to be due to the machine interlock that trips the RF upon detecting a rapid drop of beam current to protect the RF system.

It remained to be understood why the beam current suddenly drops so as to trigger a machine interlock. Once again, the bunch-by-bunch diagnostics, used for a post mortem, helped to get a closer look into what happens to the beam in the last moment before it gets lost (Fig. 31). In Fig. 31(a) where the averaged amplitude of the beam is plotted against time, we see that the beam blows up exponentially before it is lost. This suggests that the sudden current drop is due to the blown-up beam being scraped off either by the vacuum chamber or by the dynamic acceptance. Analyzing the phase relation between adjacent bunches as done in Fig. 29(b) above, we can identify chronologically the appearance of the following three regimes up to the explosion: Fig. 31(b) no phase correlation (i.e. absence of coherence); Fig. 31(c) ion regime; and Fig. 31(d) RW regime. We see in Fig. 29(a) that for some reason transverse feedback fails to keep the beam stable when it is in the ion regime.

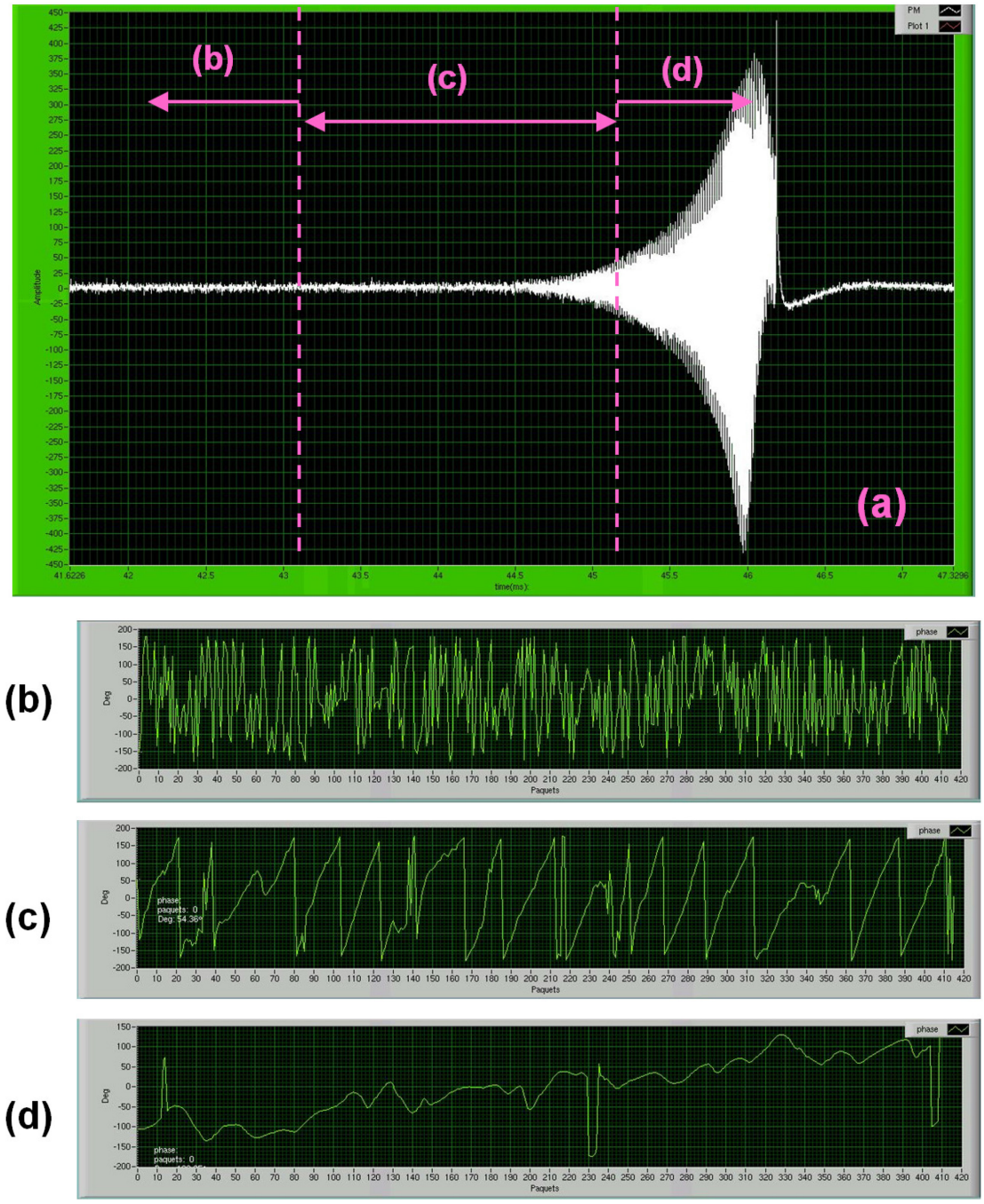


Fig. 31: (a) Measured evolution of the vertical oscillation amplitude average over bunches in ¾ filling versus time. The beam blows up and gets lost at around 46 ms after the beginning of the measurement. Measured relative betatron phase with respect to adjacent bunches along the bunch train identifies three different regimes differentiated in time as indicated in Fig. 31(a): (b) no phase correlation; (c) ion regime; (d) RW regime.

The last observation implies that during some ten minutes of 'silence', the gas density, and therefore the number of ions created at each turn, steadily increase up to the point that the FBII growth rate exceeds the limit of feedback, since the underlying machine heating continues. What is not obvious, however, is the reason why the beam continues to blow up in the RW regime when ions would probably be gone due to large beam oscillations. Namely, we need to understand the failure of feedback in the third RW regime. A possible explanation is that feedback, with its filter and gain used, is not reacting well enough against the fast dynamical change of the beam from FBII to RW. More details are found in

Ref. [53]. A simulation study that includes the effects of RW, FBII, and feedback reproduces the beam behaviour in a similar manner to that observed, supporting the above conjecture (Fig. 32).

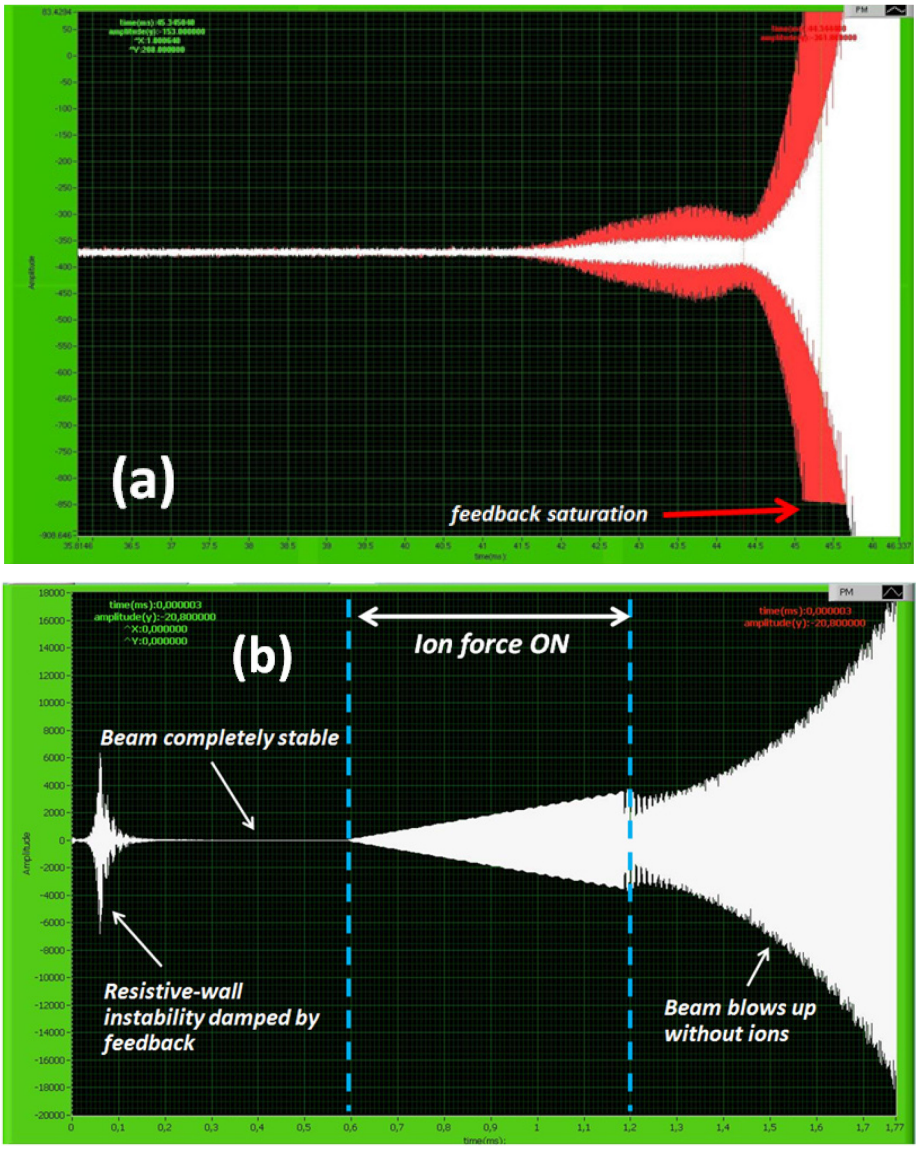


Fig. 32: (a) Measured evolution of the vertical oscillation amplitude (in white) average over bunches in ¾ filling versus time. The beam blows up and gets lost at around 46 ms after the beginning of the measurement. Amplitude of kicks given by bunch-by-bunch feedback is superimposed in red. (b) Simulation of FBII including the resistive-wall (RW) instability and bunch-by-bunch feedback. The beam blows up after the ions have disappeared in the regime feedback works against RW instability. Taken from Ref. [53].

5 Conclusions

Due to the general difficulties of measurement and frequent non-reproducibility of vacuum conditions, beam instabilities arising from ions in the beam duct are often not straightforward to understand, as compared to other collective instabilities. However, the theoretical, numerical, and experimental studies made so far, creating a solid basis for beam–ion physics, allow fairly good explanations and qualitative and quantitative predictions. There are still a number of nonlinear beam–ion dynamics involving Landau damping and stabilization that are not adequately understood, and would require further efforts.

Owing presumably to lower beam emittances in modern storage rings, ion trapping does not seem to be a big issue anymore. However, FBII could jeopardize the performance of future low emittance and high beam intensity accelerators, as its growth rate would get larger. For light sources, in particular, the effort of reaching an ultra-low emittance tends to render the vacuum chamber aperture smaller and smaller. The vacuum issues and hence ion-induced beam instabilities would likely remain important, especially in a combined manner with other effects, as already encountered at SOLEIL. Continuation of beam–ion studies would therefore be of great importance in raising the performance of future accelerators.

Acknowledgements

The author thanks Christian Herbeaux, Nicolas Béchu, Nicolas Baron, the SOLEIL vacuum experts, as well as his colleagues in the accelerator physics group at SOLEIL for useful discussions in preparing this presentation. He thanks Amor Nadji, the Division Head of SOLEIL, for his support of this work. He expresses his special thanks to Shogo Sakanaka at KEK for his helpful remarks, providing the original figures from his studies used in this paper, and for his lecture notes created for the KEK Accelerator School OHO86 (Ref. [18]), from which the present paper much profited.

All figures used in the present paper that are cited from other published papers and notes have their origins explicitly referenced and their copyright permissions properly acquired prior to the publication of this lecture note.

References

[1] R. Jolivot, Anneaux de Stockage. Le piégeage des ions dans ACO et leur balayage, Rapport technique 75-63/RJ-FB, Ecole Normale Supérieur – Faculté des Sciences, Orsay, Laboratoire de l'Accélérateur Linéaire (1963).

[2] B. Angerth, Review of studies on beam neutralization in storage rings, AR/Int. SG/65-1 (1965).

[3] E. Fischer, Clearing fields for the ISR, ISR-VAC/66-15 (1966).

[4] D. Poteaux, Piégeage des ions dans un anneau e^+e^-, Rapport technique 29-69 DP/LN Laboratoire de l'Accélérateur Linéaire, Orsay (1969).

[5] R.D. Kohaupt, Mechanismus der ionenabsangung im Electron Positron Speicherring Doris, Interner Bericht DESY H1-71/2 (1971).

[6] D.G. Koshkarev and P.R. Zenkevich, *Part. Accel.* **3** (1972) 1.

[7] R. Calder, E. Fischer, O. Gröbner and E. Jones, Vacuum conditions for proton storage rings, CERN/ISR-VA/74-26 (1974).

[8] Technical Note – ISR Vacuum Group, The behaviour of ions in presence of a bunched antiproton beam in the ISR, ISR-VA/EF-sm (1978).

[9] Y. Baconnier and G. Brianti, The stability of ions in bunched beam machines, CERN/SPS/80-2 (DI) (1980).

[10] Y. Yamazaki, M. Kihara and H. Kobayakawa, Partially filled multi-bunch mode operation of the Photon Factory electron storage ring and cure of the vertical instability, KEK 83-17 (1983), KEK internal report.

[11] A. Poncet, Ion clearing in EPA, PS/ML/Note 83-1, CERN internal report.

[12] Y. Kamiya, M. Izawa, T. Katsura, M. Kihara, H. Kobayakawa and S. Shibata, Vertical instability caused by ion-trapping in KEK-PF storage ring, Proc. 5th Symp. Acc. Sci. Tech. p. 292. , KEK, Tsukuba, Japan (1984).

[13] T. Kasuga, H. Yonehara, T. Kinoshita and M. Hasumoto, *Jpn. J. Appl. Phys*. **24** (1985) 1212, http://dx.doi.org/10.1143/jjap.24.1212

[14] M.Q. Barton, *Nucl. Instrum. Meth.* **A243** (1986) 278, http://dx.doi.org/10.1016/0168-9002(86)90961-7

[15] H. Kobayakawa, *Part. Accel*. **33** (1990) 81.

[16] G. Brianti, in Proceedings of the CAS-CERN Accelerator School: Antiprotons for Colliding-Beam Facilities, Geneva, Switzerland, 11-21 October 1983, edited by P. Bryant and W.S. Newman, CERN-1984-015 (CERN, Geneva, 1984), pp. 369-383, http://dx.doi.org/10.5170/CERN-1984-015.369

[17] Y. Baconnier, in Proceedings of the CAS-CERN Accelerator School: General Accelerator Physics, v.1, Gif-sur-Yvette, France, 3-14 September 1984, edited by P. Bryant and S. Turner, CERN-1985-019 (CERN, Geneva, 1985), pp. 267-300, http://dx.doi.org/10.5170/CERN-1985-019-V-1.267

[18] S. Sakanaka, Phenomena of ion trapping in storage rings, KEK accelerator school lecture note [in Japanese], OHO'86, KEK, Tsukuba, Japan (1986).

[19] M. Bassetti and G.A. Erskine, Closed expression for the electrical field of a two-dimensional Gaussian charge, CERN-ISR-TH/80-06 (1980).

[20] M.E. Biagini, S. Guiducci, M. Preger, M. Serio and S. Tazzari, Observation of ion trapping at ADONE, Proc. 11th Int. Conf. on High Energy Accelerators, p. 687, CERN, Geneva, Switzerland (1980).

[21] P.F. Tavares, Transverse distribution of ions trapped in an electron beam, CERN-PS/92-55 (LP) (1992).

[22] L. Wang, Y. Cai, T.O. Raubenheimer and H. Fukuma, *Phys. Rev. STAB*. **14** (2011) 084401, http://dx.doi.org/10.1103/physrevstab.14.084401

[23] K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, 2005), p. 609, http://dx.doi.org/10.1142/5835, 5 Toh Tuck Link, Singapore 596224

[24] E. Keil and B. Zotter, Landau damping of coupled electron–proton oscillations, CERN-ISR-TH-71-58 (1971).

[25] Coherent instability due to electrons in a coasting proton beam, edited by H. G. Hereward, CERN-1971-015 (CERN, Geneva, 1971). http://dx.doi.org/10.5170/CERN-1971-015

[26] T.O. Raubenheimer and F. Zimmermann, *Phys. Rev.* **E52** (1995) 5487, http://dx.doi.org/10.1103/physreve.52.5487

[27] K. Ohmi, *Phys. Rev*. **E55** (1997) 7550, http://dx.doi.org/10.1103/physreve.55.7550

[28] G. Skripka, R. Nagaoka, M. Klein, F. Cullinan and P.F. Tavares, *Nucl. Instrum. Meth.* **A806** (2016) 221, http://dx.doi.org/10.1016/j.nima.2015.10.029

[29] G. Xia, K. Ohmi and E. Elsen, Simulation study of fast ion instability in the ILC damping ring and PETRA III, unpublished; a revised version by G. Xia and E. Elsen is published in *Nucl. Instrum. Meth* **A593** (2008) 183, http://dx.doi.org/10.1016/j.nima.2008.05.054

[30] J. Byrd, A. Chao, S. Heifets, M. Minty, T.O. Raubenheimer, J. Seeman, G. Stupakov, J. Thomson and F. Zimmermann, *Phys. Rev. Lett*. **79** (1997) 79, http://dx.doi.org/10.1103/physrevlett.79.79

[31] M. Kwon, J.Y. Huang, T.-Y. Lee, I.S. Ko, Y.H. Chin, H. Fukuma, M. Isawa, K. Ohmi and M. Tobiyama, *Phys. Rev*. **E57** (1998) 6016, http://dx.doi.org/10.1103/physreve.57.6016

[32] H. Fukuma, Y. Chin, S. Kato, E. Kikutani, S. Kurokawa, S. Matsumoto, K. Ohmi, Y. Suetsugu, M. Tobiyama, K. Yokoya and X.L. Zhang, Experimental observations of the ion-related coupled bunch instability in a bunch train in TRISTAN AR, PAC97, Vancouver, 1997, p. 1596, http://accelconf.web.cern.ch/accelconf/pac97/papers/pdf/2V009.PDF.

[33] G.V. Stupakov, T.O. Raubenheimer and F. Zimmermann, *Phys. Rev*. **E52** (1995) 5499, http://dx.doi.org/10.1103/physreve.52.5499

[34] K. Ohmi, F. Zimmermann and E. Perevedentsev, *Phys. Rev*. **E65** (2001) 016502, http://dx.doi.org/10.1103/physreve.65.016502

[35] E.S. Kim and K. Ohmi, *Jpn. J. Appl. Phys*. **48** (2009) 086501, http://dx.doi.org/10.1143/jjap.48.086501

[36] S. Sakanaka, The stability of ions in partially filled mode operation in the electron storage ring, KEK Preprint 86-17 (1986).

[37] C.J. Bocchetta and A. Wrulich, *Nucl. Instrum. Meth.* **A278** (1989) 807, http://dx.doi.org/10.1016/0168-9002(89)91205-9

[38] J.A. Clarke, D.M. Dykes, S.F. Hill, E.A. Hughes, M.W. Poole, P.D. Quinn, S.L. Smith, V.P. Suller and L.A. Welbourne, Source size variation and ion effects in the SRS at Daresbury, PAC1993, Washington DC, p. 3594,

http://accelconf.web.cern.ch/AccelConf/p93/PDF/PAC1993_3594.PDF.

[39] B. Schwarzschild, *Phys. Today* March (1986) 19. http://dx.doi.org/10.1063/1.2815138

[40] M. Zobov, A. Battisti, A. Clozza, L. Lollo, C. Milardi, B. Spataro, A. Stella and C. Vaccarezza, *J. Instr.* **2** (2007) P08002, http://dx.doi.org/10.1088/1748-0221/2/08/p08002

[41] H. Zyngier, J.-C. Besson, M. Bordessoule, P. Brunelle, P. Juan, M.-P. Level, P.-C. Marin, P. Nghiem and E.M. Sommer, SUPER-ACO status report, EPAC 1990, Nice, France, p. 469, http://accelconf.web.cern.ch/AccelConf/e90/PDF/EPAC1990_0469.PDF.

[42] K.C. Harkay and R.A. Rosenberg, *Phys. Rev. STAB* **6** (2003) 034402, http://dx.doi.org/10.1103/physrevstab.6.034402

[43] R. Wanzenberg, Observations of electron cloud phenomena at PETRA III, ECLOUD'12 proceedings, Isola d'Elba, 2012, p. 89, http://cds.cern.ch/record/1668184/files/p89.pdf.

[44] M. Izawa, Y. Sato and T. Toyomasu, *Phys. Rev. Lett*. **74** (1995) 5044, http://dx.doi.org/10.1103/physrevlett.74.5044

[45] E. Plouviez, Ph. Arnoux, F. Epaud, J.-M. Koch, G. Naylor and F. Uberto, Bunch by bunch transverse feedback development at the ESRF, EPAC08, Genoa, 2008, p. 3297, http://accelconf.web.cern.ch/AccelConf/e08/papers/thpc132.pdf.

[46] R. Nagaoka, Observation and analysis of fast beam–ion instabilities at SOLEIL, International Workshop on Linear Colliders 2010, Accelerator Working Group 2, CERN, 2010.

[47] K. Oide, cited as private communication in Ref. [26].

[48] M. Kihara *et al*., Results on accelerator studies of the Photon Factory storage ring, KEK 83-5, KEK, Tsukuba, Japan (1983).

[49] M. Kobayashi, K. Huke, S. Ban and H. Hirayama, Observations of Bremsstrahlung caused by ion trapping, Proc. 5th Symp. Acc. Sci. Tech. p. 148, KEK, Tsukuba, Japan (1984).

[50] H. Kobayakawa, K. Huke, M. Izawa, Y. Kamiya, M. Kihara, M. Kobayashi and S. Sakanaka, Observation of ion trapping phenomenon with Bremsstrahlung, KEK Preprint 85-80 (1986).

[51] L. Wang, J. Safranek, Y. Cai, J. Corbett, R. O. Hettel, T. O. Raubenheimer, J. Schmerge and J. Sebek, *Phys. Rev. STAB* **16** (2013) 104402, http://dx.doi.org/10.1103/physrevstab.16.104402

[52] R. Nagaoka, L. Cassinari, M. Diop, J.M. Filhol, M.P. Level, M. Labat, P. Marchand and R. Sreedharan, Study of ion-induced instabilities and transverse feedback performance at SOLEIL, IPAC2011, San Sebastián, Spain, September 201, p. 712, http://accelconf.web.cern.ch/AccelConf/IPAC2011/papers/mops049.pdf.

[53] R. Nagaoka, R. Sreedharan and L. Cassinari, Fast beam-ion instability arising from local outgassing, presented at TWIICE workshop, Synchrotron SOLEIL, Gif-sur-Yvette, France, 2014, http://indico.cern.ch/event/277919/contributions/626920/attachments/506917/699839/05-RNs_talk_TWIICE.pdf.

Participants

ALBRIGHT, S.	CERN, Geneva, CH
ANDRIANOV, A.	Budker Institute of Nuclear Physics, Novosibirsk, RU
APPEL, S.	GSI, Darmstadt, DE
ASTRELINA, K.	Budker Institute of Nuclear Physics, Novosibirsk, RU
BAGLIN, V.	CERN, Geneva, CH
BAHAMONDE, C.	CERN, Geneva, CH
BARANOV, G.	Budker Institute of Nuclear Physics, Novosibirsk, RU
BARMINOVA, H.	National Research Nuclear University Mephi, Moscow, RU
BARRANCO, J.	CERN, Geneva, CH
BAUDRENGHIEN, P.	CERN, Geneva, CH
BIANCACCI, N.	CERN, Geneva, CH
BOZZOLAN, M.	CERN, Geneva, CH
BREGLIOZZI, G.	CERN, Geneva, CH
CHAI, W.	Institute of Modern Physics, Lanzhou, CN
CLAESSENS, C.	Iap University of Frankfurt, Frankfurt, DE
COSTA PINTO, P.	CERN, Geneva, CH
CULLINAN, F.	Synchrotron Soleil, Gif-sur-Yvette, FR
DANISI, A.	CERN, Geneva, CH
DOROHOV, V.	Budker Institute of Nuclear Physics, Novosibirsk, RU
EIDAM, L.	TEMF, Darmstadt, DE
ESTEBAN MUELLER, J.F.	CERN, Geneva, CH
FRASER, M.	CERN, Geneva, CH
GARCIA ORTEGA, P.	CERN, Geneva, CH
GEITHNER, O.	GSI, Darmstadt, DE
GERARDIN, F.	CEA Irfu and CEA Saclay, Gif-sur-Yvette, FR
GORZAWSKI, A.	CERN, Geneva, CH
GRIGORYAN, A.	CANDLE, Yerevan, AM
GRUDIEV, A.	CERN, Geneva, CH
HOLZ, M.	Uppsala University, Uppsala, SE
HOSTETTLER, M.	CERN, Geneva, CH
JONES, B.	Rutherford Appleton Laboratory, Chilton, UK
KARIUKINA, K.	Budker Institute of Nuclear Physics, Novosibirsk, RU
KARPOV, I.	Institut Fuer Theorie Elektromagnetischer Felder, Darmstadt, DE
KERSEVAN, R.	CERN, Geneva, CH
KESTING, F.	University of Frankfurt, Frankfurt, DE
KWEE, R.	Royal Holloway University of London, Egham, UK
LARI, L.	European Spallation Source, Lund, SE
LASHEEN, A.	CERN, Geneva, CH
LASHEEN, A.	CERN, Geneva, CH
LIU, J.	Institute of Modern Physics, Lanzhou, CN
NAVARRO QUIRANTE, J.L.	CERN, Geneva, CH
NIKIFOROV, D.	Budker Institute of Nuclear Physics, Novosibirsk, RU
NUIRY, F.	CERN, Geneva, CH
PALM, M.	CERN, Geneva, CH
PAPOTTI, G.	CERN, Geneva, CH
PASSARELLI, A.	Tu Darmstadt, Darmstadt, DE
ROGOVSKY, Y.	Budker Institute of Nuclear Physics, Novosibirsk, RU
ROMANO, A.	Tu Darmstadt, Darmstadt, DE

SALVANT, B.	CERN, Geneva, CH
SCHMIDT, J.	CERN, Geneva, CH
SCHOOFS, P.	CERN, Geneva, CH
SHAKER, H.	CERN, Geneva, CH
SHVARTS, D.	Budker Institute of Nuclear Physics, Novosibirsk, RU
SKRIPKA, G.	Lund University, Lundt, SE
STACHYRA, K.	CERN, Geneva, CH
STEM, W.	GSI, Darmstadt, DE
STOEL, L.	CERN, Geneva, CH
TAMBASCO, C.	CERN, Geneva, CH
TAN, J.	CERN, Geneva, CH
TIMKO, H.	CERN, Geneva, CH
TSINGANIS, A.	CERN, Geneva, CH
VENTURA, L.	CERN, Geneva, CH
XU, H.	PSI, Villigen, CH
YUAN, Y.	TEMF and TU Darmstadt, Darmstadt, DE
ZHANG, Y.	Institute of Modern Physics, Chinese Academy, Lanzhou, CN
ZHAO, H.	Institute of Modern Physics, Lanzhou, CN
ZHU, G.	Institute of Modern Physics, Chinese Academy, Lanzhou, CN