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 freeelectron 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 ydirections, respectively. Brightness is measured in amps per square metre. Typical values of high brightness beams range between 10^{14} and 10^{16} A/m². 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



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 $100 \,\mu\text{m}$ 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

$$\varepsilon_{\rm rms}^2 = \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2 \,, \tag{3}$$

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

$$\varepsilon_{\rm rms}^2 = C \left\langle x^2 \right\rangle \left\langle x^{2n} \right\rangle - \left\langle x^{n+1} \right\rangle^2 \,. \tag{4}$$

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

$$\varepsilon_{\rm n}^2 = \left\langle x^2 \right\rangle \left\langle \beta^2 \gamma^2 x'^2 \right\rangle - \left\langle x \beta \gamma x' \right\rangle \,. \tag{5}$$

It has been demonstrated in Ref. [19] that this formula can be rewritten as

$$\varepsilon_{\rm n}^2 = \langle \gamma \rangle^2 \, \sigma_{\varepsilon}^2 \, \langle x^2 \rangle \, \langle x^{\prime 2} \rangle + \langle \beta \gamma \rangle^2 \left(\langle x^2 \rangle \, \langle x^{\prime 2} \rangle - \langle xx^{\prime} \rangle^2 \right) \,, \tag{6}$$

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

$$\sigma_x'' = \frac{\varepsilon_n^2}{\gamma^2 \sigma_x^3} + \frac{I}{\gamma^3 I_0 \left(\sigma_x + \sigma_y\right)} , \qquad (7)$$

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-tonoise 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

$$R_0 = \frac{I\sigma_x^2}{2\gamma I_0\varepsilon_n^2} \,. \tag{8}$$

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 $50 - 100 \,\mu\text{m}$.

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

$$\sigma_x = \sqrt{\left(L \cdot \sigma'_x\right)^2 + \frac{d^2}{12}} \,. \tag{9}$$

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 plasmaaccelerated 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

$$\varepsilon_{\rm err} = \frac{2}{\pi} \left(x_{\rm max} \Delta x' + x'_{\rm max} \Delta x \right) , \qquad (10)$$

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.

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = \varepsilon = \gamma_0 x_0^2 + 2\alpha_0 x_0 x_0' + \beta_0 x_0'^2 , \qquad (11)$$

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:

$$M(x, x_0) = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} .$$
(12)

It has been demonstrated [27] that, using Eq. (12) in Eq. (11), the transport matrix for the Twiss parameters is

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & S'C + SC' & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix}.$$
 (13)

Let us define a so-called σ matrix:

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \varepsilon \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix}.$$
 (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

$$\sigma_{11}x^2 + 2\sigma_{12}xx' + \sigma_{22}x'^2 = 1.$$
(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

$$\sigma_{11}^1 = C^2 \sigma_{11}^0 + 2SC \sigma_{12}^0 + S^2 \sigma_{22}^0 .$$
⁽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

$$y(z) = y_0 + \left(\sqrt{\beta\beta_0}\sin\Delta\right)y_0' \pm \frac{qV_0}{pc}kz\left(\sqrt{\beta\beta_0}\sin\Delta\right)$$
(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

$$\left\langle \left(y - \langle y \rangle\right)^2 \right\rangle^{\pm} = \left\langle y_0^2 \right\rangle + \beta \beta_0 \sin^2 \Delta \left\langle y_0'^2 \right\rangle - \left\langle y_0 \right\rangle^2 + \beta \beta_0 \sin^2 \Delta \left(\frac{qV_0}{pc}k\right)^2 \left\langle z^2 \right\rangle + -\beta \beta_0 \sin^2 \Delta \left\langle y_0' \right\rangle^2 - \beta \beta_0 \left(\frac{qV_0}{pc}k\right)^2 \sin^2 \Delta \left\langle z \right\rangle^2 + + 2\sqrt{\beta \beta_0} \sin \Delta \left\langle y_0 y_0' \right\rangle \pm 2\sqrt{\beta \beta_0} \sin \Delta \left(\frac{qV_0}{pc}k\right) \left\langle y_0 z \right\rangle - 2\sqrt{\beta \beta_0} \sin \Delta \left\langle y_0 \right\rangle \left\langle y_0' \right\rangle + \mp 2\sqrt{\beta \beta_0} \sin \Delta \left(\frac{qV_0}{pc}k\right) \left\langle y_0 \right\rangle \left\langle z \right\rangle \pm 2\beta \beta_0 \sin^2 \Delta \left(\frac{qV_0}{pc}k\right) \left\langle y_0' z \right\rangle + \mp 2\beta \beta_0 \sin^2 \Delta \left(\frac{qV_0}{pc}k\right) \left\langle y_0' \right\rangle \left\langle z \right\rangle$$

$$(20)$$

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 $\varphi=0$ and $\varphi=\pi$ is required in order to cancel them. Finally the dimension measured is

$$\sigma = \sqrt{\sigma_0^2 + \sigma_z^2} \tag{22}$$

where $\sigma_z^2 = \beta \beta_0 \sin^2 \Delta \left(rac{q V_0}{p c} k
ight)^2 \left\langle z^2
ight
angle$

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-ofmass 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 timedispersive 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 *j*th particle, \vec{r} being the position vector of the *j*th electron relative to its bunch centre. We can express the electric field generated by the *j*th 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

$$I_{\text{tot}}(\omega) = I_{\text{sp}}(\omega) \sum_{j,k} E_j E_k^* , \qquad (23)$$

where $I_{sp}(\omega)$ 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

$$I_{\text{tot}}(\omega) = I_{\text{sp}}(\omega) \left[N + N \left(N - 1 \right) F(\omega) \right] , \qquad (24)$$

where $I_{\rm sp}(\omega)$ 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 $I_{\rm sp}(\omega)$, 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

$$F(\omega) = \left| \int_{-\infty}^{\infty} \rho(z) \mathrm{e}^{\mathrm{i}\frac{\omega}{c}z} \mathrm{d}z \right|^2 \,.$$
(25)

The inverse transformation gives the searched value of $\rho(z)$

$$\rho(z) = \frac{1}{\pi c} \int_{0}^{\infty} \sqrt{F(\omega)} \cos\left(\frac{\omega}{c}z\right).$$
(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,

$$I(\omega) \propto \int_{-\infty}^{\infty} I(\delta) \cos\left(\frac{\omega\delta}{c}\right) \mathrm{d}\delta$$
, (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 μ m. The radiation is dispersed and then sent to 128 lead zirconate titanate pyroelectric elements with 100 μ m 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 $300 \mu m$, 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 timedependent 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; P₁,P₂, 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 $T_{\rm lim} \approx 2.6\sqrt{T_0T_c}$ [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 ≈ 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 freeelectron 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.