We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800 Open access books available
116,000 International authors and editors
120M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Design and Simulation Challenges of a Linac-Based Free Electron Laser in the Presence of Collective Effects

S. Di Mitri
Sincrotrone Trieste
Italy

1. Introduction

The generation of Free Electron Laser (FEL) radiation relies on the extraction of electromagnetic energy from kinetic energy of a relativistic electron beam by propagating it along the axis of a periodic lattice of alternating magnetic dipolar fields, known as undulator. This forces the beam to undulate transversally, thus causing the electrons to emit electromagnetic radiation. The fundamental wavelength emitted is proportional to $\lambda_u/\gamma^2$, where $\lambda_u$ is the undulator period, typically a few centimeters long, and $\gamma$ is the relativistic Lorentz factor of the electrons, which typically reaches several thousand for X-ray emission. The main figures of merit of an FEL are extremely high brilliance, close to full transverse and longitudinal coherence, a bandwidth approaching the Fourier limit and a stable and well characterized temporal structure in the femtosecond time domain. We can identify two general ways to generate X-rays with an FEL. The Self Amplified Spontaneous Emission (SASE) [1–4] relies on the interaction of electrons and photons that are emitted by the electron beam itself. The electron bunching that generates the coherent emission of radiation starts to grow from the natural noise of the initial electron distribution. For this reason, the SASE output radiation is relatively poor in longitudinal coherence. In the High Gain Harmonic Generation (HGHG) scheme [5–11], instead, the initial energy modulation is driven by an external seed laser. It is then transformed into density bunching in a dispersive section inserted in the undulator chain. In this case, the output FEL properties reflect the high longitudinal coherence of the seed laser.

The FEL high brilliance, high intensity and shot-to-shot stability strongly depends on the electron beam source. As an example, an FEL requires a high peak current to increase the number of photons per pulse and reach power saturation at an early stage in the undulator. Magnetic bunch length compression is one way to increase the electron bunch current. It is carried out via ballistic contraction or elongation of the particles path length in a magnetic chicane. The linac located upstream of the magnetic chicane is run off-crest to establish a correlation between the particle longitudinal momentum with respect to the reference particle and the $z$-coordinate along the bunch, i.e. the bunch head has a lower energy than the tail. In the magnetic chicane, due to their lower (higher) rigidity, leading (trailing) particles travel on a shorter (longer) path than trailing (leading) particles. Since all particles of the ultra-relativistic beam travel in practice at the speed of light, the bunch edges approach the centroid position.
and the total bunch length is finally reduced. Unfortunately, magnetic compression is very often enhancing single particle and collective effects that may degrade the electron beam quality. Delivering a high quality electron beam and machine flexibility to serve a broad range of potential applications imposes severe requirements on the final electron beam parameters and the machine design. The primary goal of the machine design is that of preserving the 6-D electron beam emittance, $\epsilon$, at the electron source level. Liouville’s theorem [12] states that the phase space hypervolume enclosing a chosen group of particles is an invariant of the Hamiltonian system as the particles move in phase space, if the number of particles in the volume does not change with time. This volume is the emittance of the particle ensemble. In its more general sense, Liouville’s theorem applies to Hamiltonian systems in which the forces can be derived from a potential that may be time dependent, but must not depend on the particles’ momentum. Thus, the following collective phenomena limit the applicability of the Liouville’s theorem to particles motion in an accelerator: collisions, space charge forces – intended as short-range inter-particle Coulomb interactions –, wake fields, electromagnetic radiation emission and absorption. In order to preserve the initial 6-D volume along the entire electron beam delivery system, all the afore-mentioned effects have to be analytically evaluated and simulated. In particular, we will focus on the particle motion in a single-pass, linac-driven FEL in the presence of the following short-range effects:

i) space charge (SC) forces [13–16];

ii) geometric longitudinal and transverse wake field in the accelerating structures [17–18];

iii) coherent synchrotron radiation (CSR) emission in dispersive systems [19–27];

For a clearer illustration of these topics, we will initially assume the particle motion being uncoupled in the transverse and in the longitudinal phase space. A good transverse coherence of the undulator radiation is ensured by the following limit [28] on the transverse normalized beam emittance: $\gamma\epsilon \leq \gamma \lambda/(4\pi)$, where $\lambda/(4\pi)$ is the minimum phase space area for a diffraction limited photon beam of central wavelength $\lambda$. Typically, $\gamma\epsilon \approx 1$ mm mrad for $\lambda$ in the nm range. Note that the local emittance, referred to as “slice”, can vary significantly along the bunch to give hot-spots where lasing can occur. In fact, in contrast to linear colliders where particle collisions effectively integrate over the entire bunch length, X-ray FELs usually concern only very short fractions of the electron bunch length. The integration length is given by the electron-to-photon longitudinal slippage over the length of the undulator, prior to FEL power saturation. The FEL slippage length is typically in the range 1–30 $\mu$m, a small fraction of the total bunch length. Thus, the electron bunch slice duration can reasonably be defined as a fraction of the FEL slippage length.

The longitudinal emittance or more precisely the energy spread for a given electron bunch duration, has to be small enough to permit the saturation of the FEL intensity within a reasonable undulator length. At saturation of a SASE FEL, $P \approx \rho P_e$, where $P_e$ is the electron beam power and $\rho$, the so-called Pierce parameter [29], is the FEL gain bandwidth expressed in terms of normalized energy. It is seen to be a measure of the efficiency of the interaction, with typical values in the X-ray regime of $10^{-4} \leq \rho \leq 10^{-3}$. The relative energy spread of the electron beam at saturation is $\sigma_\delta \approx \rho$. Thus, if there is an initial electron energy spread approaching the maximum, which occurs at an FEL saturation of $\sigma_\delta \geq \rho$, then the FEL interaction is greatly reduced. For a seeded FEL such as in a HGHG scheme, the total energy spread $\sigma_{\delta,\text{tot}}$ is approximately given by the quadratic sum of the uncorrelated term $\sigma_{\delta,\text{unc}}$, the energy modulation amplitude induced by the seeding laser $\Delta\delta$, and the residual energy chirp
The maximum acceptable deviation from the desired flatness of the longitudinal phase space is limited by $\sigma_{\delta,\text{tot}} \leq \rho$. At the same time, the FEL harmonic cascade is effective only for $\Delta \delta \geq N \sigma_{\delta,\text{un}}$ with $N$ the ratio between the seed wavelength and the harmonic wavelength at which the final undulator is tuned. So, if $N = 10$ to produce, as an example, 20 nm and we want $\sigma_{\delta,\text{tot}} \leq 1 \cdot 10^{-3}$, we require a final energy chirp $\sigma_{\delta,\text{ch}} \leq 10^{-3}$ and a final slice energy spread (here assumed to be as uncorrelated for conservative calculations) $\sigma_{\delta,\text{un}} \leq 150 \text{ keV}$ at the beam energy of 1.5 GeV. Typical electron beam parameters of the fourth generation linac-based FELs (from infrared to X-rays spectral range) are listed in the following: 0.1–1 nC charge, 0.5–2 mm mrad normalized emittance, 0.5–3 kA peak current, 0.05–0.1 % relative energy spread and 1–50 GeV final electron energy.

2. Short-range space charge forces

The electron beam generation from a metallic photo-cathode [30,31] in the $\gamma<200$ regime is dominated by SC forces, which scale like $\gamma^{-2}$ in the laboratory frame, and by image charge forces in the immediate cathode vicinity. The dynamics in these regions is usually simulated with quasi-static 3-D codes like GPT [32] or Astra [33] (the static electric field is calculated in the beam frame and both electric and transverse magnetic fields are included in the laboratory frame). They can model the electron gun azimuthal asymmetries (due to imperfections) and also large beam aspect ratios. Special beam current shaping at the cathode can be included: this is required to ameliorate the linearization of the longitudinal phase space in the presence of high beam charge (> 100 pC) and strong longitudinal geometric wake field, with a further positive impact on the final energy chirp [34]. These codes predict an uncorrelated energy spread out of the RF photo-injector in the range 1–3 keV rms [35–37]. The radiative force related to the variation of the bunch total electromagnetic energy during acceleration has also been recognized as a new source of local energy spread [38]. This new physics can only be studied with codes that correctly calculate the beam fields from the exact solutions of the Maxwell’s equations that is the full retarded potentials. The slice energy spread is of crucial importance for the suppression of the so-called microbunching instability [39–41]. In fact, the velocity spread of the relativistic electrons acts as a low pass filter effect for density and energy modulations generated in the Gun. While the conversion of energy and density modulation amplitudes happens over the SC oscillation wavelength of ~ 1 m, only wavelengths longer than 10’s of mm survive out of the Gun [42].

Even when the electrons reach energies as high as ~50 MeV, SC forces have to be considered in two cases. In the first case, the longitudinal electric field generated by clusters of charges or density modulation along the bunch can still be sufficiently high to induce an energy modulation as the beam travels along the accelerator. Such an energy modulation translates into density modulation when the beam passes through a dispersive region (this might happen in a magnetic bunch length compressor or in a dispersive transfer line), with a consequent degradation of the energy and the current flatness. This dynamics is assumed to be purely longitudinal and it is discussed in detail in Section 6. In the second case, SC forces might be enhanced because of the very high charge density achieved with the bunch length compression. Although 3-D tracking codes can be used to simulate the compression, an analytical estimation of the impact of these forces on the transverse dynamics is still possible. Following [15] we find that the rms transverse envelope equation for a bunched beam in a
linac is:
\[
\sigma'' + \frac{\gamma'}{\gamma} \sigma' + K \sigma = \frac{k_s}{\gamma^3 \sigma} + \frac{\epsilon_{th}}{\gamma^2 \sigma^3}
\]  (1)

Here, the standard deviation of the beam transverse size \( \sigma \) is assumed to be a function of the axial position \( s \) along the linac, \( \gamma' = d\gamma/ds \) is the accelerating gradient, \( K = (eB_0)^2/(2mc\gamma)^2 \) is the focusing gradient of a solenoid of central field \( B_0 \), \( k_s = 1/(2I_A) \), \( I_A = 17kA \) being the Alfvén current, and \( \epsilon_{th} \) is the thermal emittance, which is mainly due to the photoemission process at the cathode surface: it is a Liouville invariant throughout acceleration. We now consider the following invariant envelope solution for the beam size:
\[
\bar{\sigma} = \frac{1}{\gamma'} \sqrt{\frac{k_s}{\gamma(1/4 + \Omega^2)}}
\]  (2)

where \( \Omega = \sqrt{K\gamma'}. \) The laminarity parameter, \( \rho_L \), is defined as the ratio of the “space charge term” driven by \( k_s \) and the “emittance term” driven by \( \epsilon_{th} \) in eq.1, computed with the substitution \( \sigma = \bar{\sigma} \):
\[
\rho_L = \left( \frac{k_s}{\epsilon_{th} \gamma' \sqrt{1/4 + \Omega^2}} \right)^2
\]  (3)

If \( \rho_L \gg 1 \), the particles motion is dominated by SC forces with negligible contribution from the betatron motion. By computing the laminarity parameter as function of the beam parameters along the transport system, we can identify machine areas where \( \rho_L \gg 1 \), which should be investigated more carefully with 3-D codes. As an example, for an electron linac driven by a standing wave photoinjector with no external focusing, \( \Omega^2 = 1/8 \), the energy at which the transition occurs, \( \rho_L = 1 \), can be quite high:
\[
\gamma_{tr} = \sqrt{\frac{2k_s}{3 \epsilon_{th} \gamma'}}
\]  (4)

often corresponding to several hundreds of MeV. Unfortunately, the transition from SC dominated (\( \rho_L \gg 1 \)) to quasi-laminar (\( \rho_L \ll 1 \)) motion cannot be described accurately by this model because, by definition, \( \bar{\sigma} \) is a valid solution of eq.1 only for \( \rho_L \gg 1 \).

We consider the following example: a 400 pC bunch time-compressed to reach 0.5 kA and a 800 pC bunch compressed to reach 1 kA. In both scenarios we assume \( \epsilon_{th} = 0.6 \) mm mrad, \( \gamma' = 39.1 \) (corresponding to 20 MV/m in a S-band linac); one- and two-stage compression is adopted at the energy of, respectively, 300 and 600 MeV. The transition energy computed with eq.4 is 500 MeV for the low charge and 1 GeV for the high charge. We can conclude that the beam dynamics is not SC-dominated in the case of low charge/two-stage compression and only weakly dominated in the high charge/two-stage compression. The one-stage compression is performed at an energy well below the computed thresholds for both charges. Thus, a careful study of the 3-D beam dynamics in the presence of SC forces should be carried out for this option. Figure 1 shows the laminarity parameter, eq.3 with \( \Omega = 0 \), computed on the basis of a particle tracking, performed with the elegant code [43], from the injector end (100 MeV) to the linac end (1.5 GeV), in the configuration of one-stage compression. No external solenoid focusing is considered. The laminarity parameter and the peak current are computed as the average value over the bunch core, which runs over \( \sim 80\% \) of the total bunch length.

www.intechopen.com
Fig. 1. Laminarity parameter and peak current computed from particle tracking. Left: 400 pC compressed by a factor of 6.5. Right: 800 pC compressed by a factor of 10. In both cases, \( \rho_L \gg 1 \) immediately downstream of the first magnetic compression, where the peak current in the bunch core rises to 1 kA. Then \( \rho_L \) falls down with \( \sim \gamma^{-2} \) dependence.

3. Short-range geometric longitudinal wake field

3.1 Analytical model

For a longitudinal charge distribution \( \lambda_z \), the energy loss of a test electron due to the electromagnetic wake of leading electrons is given by the geometric wake potential [17, 18]:

\[
W(z) = -\int_{z}^{\infty} w(z-z')\lambda_z(z')dz'
\]  

where \( w(z-z') \) is the Green’s function, also called “wake function”, that emulates the effect of the wake fields as generated by a single particle. Because of the principle of causality, the wake is zero if the test electron is in front of the wake source. If the beam is much shorter than the characteristic wake field length \( s_0 \) [44] and if the structure length \( L \) is much longer than the catch-up distance \( a^2/(2\sigma_z) \), where \( a \) is the cell iris radius and \( \sigma_z \) is the rms bunch length, then the wake field is said to be, respectively, in the periodic structure and in the steady state regime:

\[
\frac{a^2}{2L} \ll \sigma_z \ll s_0
\]

The characteristic length \( s_0 = (0.41a^{1.8}g^{1.6}/L^{2.4}) \) [44] is function of the cell iris radius \( a \), of the cell inner width \( g \) and of the iris-to-iris distance \( L_c \). In the very special case of periodic structure, steady state regime and very short electron bunches, the wake function assumes a simple form. For the longitudinal component we have in [V/C/m]:

\[
w_L(0^+) = \frac{Z_0c}{\pi \sigma}
\]

Here \( Z_0=377 \) \( \Omega \) is the free vacuum impedance. Typical values are \( s_0=1.5 \) mm, \( s_1=0.5 \) mm, \( \sigma_z=40-100 \) \( \mu \)m and \( a^2/(2L_c)=2-20 \) \( \mu \)m. So, while the steady state approximation is always satisfied, the periodic structure approximation might be not. Nevertheless, it was found that by computing the short-range wake numerically and fitting it with a simple function, one can obtain a result that is valid over a large range of \( z \) (position along the bunch) and over a useful
range of parameters [45]:

\[ w(z) = \frac{Z_0 c}{\pi a^2} e^{-\sqrt{z/s_0}} \left( \frac{V}{C \cdot m} \right) \]  

(8)

Depending on the specific geometry of the accelerating structure, eq. 8 can be modified with additional terms whose dependence on \( z \) is a polynomial. In such cases, the wake amplitude and the polynomial coefficients are determined by fitting procedures (see [46] as an example).

### 3.2 Energy loss

The longitudinal wake potential induces a total energy loss of the electron beam so that the relative energy change at the bunch length coordinate \( \bar{z} \) is [17]:

\[ \delta w(\bar{z}) = -\frac{\epsilon^2 L}{\gamma mc^2} \int_0^{\infty} w(z)n(\bar{z} - z)dz \]  

(9)

where \( n(z - z) \) is the longitudinal particle distribution with normalization \( \int_{-\infty}^{\infty} n(z)dz = N \) (\( N \) is the total number of electrons in the bunch). As an example, for a uniform longitudinal bunch profile, one has \( n(z) = N / (2\sqrt{3}\sigma_z) \) for \( |z| \leq \sqrt{3}\sigma_z \) and \( n(z) = 0 \) for \( |z| > \sqrt{3}\sigma_z \). If the constant wake function in eq.7 is used, then eq.9 yields a linear wake-induced energy change along the bunch coordinate:

\[ \delta w(\bar{z}) = -\frac{2N\epsilon L}{\gamma a^2} \left( 1 + \frac{\bar{z}}{\sqrt{3}\sigma_z} \right) \]  

(10)

where we have used the identity \( Z_0 c\epsilon_0 = 1; \epsilon_c = 2.82 \cdot 10^{-15} \text{ m} \) is the classical electron radius. It is straightforward to calculate the standard deviation of the wake-induced relative energy loss:

\[ \sqrt{\langle \delta w(\bar{z})^2 \rangle} = -\frac{2N\epsilon L}{\sqrt{3} \gamma a^2} \]  

(11)

For a S-band, 1 GeV linac with inner iris radius of 10 mm and bunch charge of 200 pC, the total loss is of the order of a few MeV. Since the uncorrelated energy spread is a few order of magnitudes smaller than this, the energy loss translates into correlated energy spread. In the linear approximation, it could be removed by running off-crest some accelerating structures at the end of the linac in order to compensate this additional energy chirp.

### 3.3 High order energy chirp

Generally, a linear description of the longitudinal beam dynamics is not accurate enough. In fact, a nonlinear energy chirp usually affects the longitudinal phase space. It reduces the effective compression factor, enlarge the FEL spectral output bandwidth (via quadratic component of energy chirp) [47] and create current spikes at the bunch edges during compression (via cubic component of energy chirp) [34, 48], which lead to further detrimental effects on energy spread and emittance due to enhanced CSR field and wake fields. To investigate this nonlinear particle dynamics, we start with the expression for the bunch length

www.intechopen.com
transformation through magnetic compression at 2nd order:

$$z = z_0 + R_{56} \delta + T_{566} \delta^2$$  \hspace{1cm} (12)

\(R_{56}\) \((T_{566})\) is the integral of the first (second) order dispersion function along the chicane, taken with the signed curvature of each dipole. It governs the linear (quadratic) path-length dependence from the particle energy. In the following, we choose a longitudinal coordinate system such that the head of the bunch is at \(z<0\). A chicane has \(R_{56} < 0\) and \(T_{566} > 0\) with this convention. A linac with energy gain \(eV \sin \phi\) (\(V\) and \(\phi\) are the RF accelerating peak voltage and phase, respectively) imparts to the beam the following linear and quadratic energy chirp, 

\[\delta = \frac{MV}{E_{BC}} \approx \delta_{0,u} + h z_0 + h' z_0^2,\]

with:

\[
h = \frac{1}{E_{BC}} \frac{dE}{dz} = \frac{2 \pi}{\lambda_{RF}} \frac{eV \cos \phi}{E_0 + eV \sin \phi}
\]

\[
h' = \frac{1}{2} \frac{dh}{dz} = -\left(\frac{2 \pi}{\lambda_{RF}}\right)^2 \frac{eV \sin \phi}{E_0 + eV \sin \phi}
\]

\(E_0\) and \(E_{BC}\) are the beam mean energy at the entrance and at the exit of the linac, respectively.

The “linear compression factor” is defined as:

\[C = \frac{\sigma_{0,z}}{\sigma_z} \approx \frac{1}{1 + h R_{56}}\]  \hspace{1cm} (14)

In practice, compressions by a factor bigger than \(\sim 3\) are dominated by nonlinear effects such as sinusoidal RF time-curvature (mostly giving a quadratic energy chirp) in the upstream linac and \(T_{566}\). For simple magnetic chicanes with no strong focusing inside, the RF and the path-length effects \(T_{566} \approx -3 R_{56}/2\) always conspire with the same signed 2nd order terms to make the problem worse. By inserting eq.13 into eq.12, we obtain the bunch length transformation at 2nd order:

\[
\sigma_z^2 = (R_{56} \sigma_{0,u})^2 + (1 + h R_{56})^2 \sigma_{20}^2 + (T_{566} \sigma_{0,u}^2)^2 + (h^2 T_{566} + h' R_{56})^2 \sigma_{20}^4 + (2 h T_{566} \sigma_{0,u}^2)^2
\]  \hspace{1cm} (15)

In order to linearize the 2nd order bunch length transformation, the use of a short section of RF decelerating field at a higher harmonic of the linac RF frequency \([49, 50]\) is usually adopted, thereby maintaining the initial temporal bunch profile and avoiding unnecessary amplification of undesired collective effects. The necessary harmonic voltage is \([50]\):

\[
e V_s^2 = \frac{E_{BC1}}{1 + \frac{\lambda_2^2}{\lambda_s^2} \frac{T_{566}}{|T_{566}|} \left(1 - \frac{\sigma_{20}}{\sigma_{0,u}}\right)^2} - E_0
\]  \hspace{1cm} (16)

The square of the harmonic ratio \(n^2 = (\lambda_s/\lambda_2)^2\) in the denominator suggests that higher harmonics are more efficient for 2nd order compensation, decelerating the beam less.
Unlike the quadratic chirp, the cubic energy chirp in a S-band linac is dominated by a contribution from the longitudinal wake potential (this may include both SC in the injector and geometric longitudinal wake field), rather than by higher order terms from the RF curvature. It has three main disrupting consequences: i) it reduces the efficiency of the magnetic compression for the bunch core, since during compression the edges “attract” particles from the core reducing the current in this region; ii) it induces current spikes at the edges that may be dangerous sources of CSR, with a direct impact on the transverse emittance and on the energy distribution; iii) wake field excited by a leading edge spike may cause additional energy spread in the low gap undulator vacuum chambers. The cubic chirp is always negative for a flat-top charge distribution \cite{51}. After the interaction with longitudinal wake fields, its sign is reversed at the entrance of the second compressor, if present, so enhancing the energy-position correlation of the bunch edges with respect to the core. The edges are there over-compressed producing current spikes. On the contrary, a negative cubic chirp at the chicane provides under-compression of the edges. For these reasons the sign of the cubic term is related to the topology of the longitudinal phase space and to the final current profile.

For a given charge and bunch length, the interaction of the cubic chirp coming from the injector with the longitudinal wake field of the succeeding linac cannot be arbitrarily manipulated. However, the user has one more degree of freedom to manage the cubic chirp before reaching the magnetic compressor, that is by setting the harmonic cavity a few degrees away from the usual decelerating crest. Typical voltages of a fourth harmonic (X-band) RF structure adopted for compensating the quadratic energy chirp are in the range 20–40 MeV. The RF phase is usually shifted by a few X-band degrees from the decelerating crest to cancel the cubic energy chirp. Adjustments to the voltage and to the phase have to be studied with a simulator, depending on the cubic energy chirp coming from the injector and on the effective compression factor.

### 3.4 Current shaping

In some cases the knob of off-crest phasing the high harmonic structure to minimize the cubic energy chirp may be weak and a significant increase is needed in the amplitude of the structure voltage. One way to achieve this is to use a density distribution other than the standard parabolic one. This is one of the motivations leading to the technique of current shaping \cite{34}. The basic premise for current shaping is that the output bunch configuration is largely pre-determined by the input bunch configuration and that therefore it is possible to find a unique electron density distribution at the beginning of the linac that produces a distribution at the end of the linac that is flat both in energy and in current. To find this distribution, one needs to reverse the problem, i.e. start at the end of the linac and move backwards towards the beginning of the linac. Eq.17 shows that for a given electron density \(\lambda_z\) and wake function \(w_z\), the electron energy at the end of a section of the linac, defined as \(\delta_j\) (with \(z_j\) being the electron coordinate taken with respect to the bunch center), can be determined using the electron energy \(\delta_i\) and the coordinate \(z_i\) at the beginning of the section:

\[
\delta_j(z_j) = \delta_i(z_i) + eU \cos(kz_j + \phi) - LQ \int_{z_i}^{+\infty} w_z(z_i - z')\lambda_z(z')dz'
\]  

(17)
where $U, \phi, L$ define the RF voltage, phase and length of the linac section, $k$ is the wave number, $e$ is the electron charge and $Q$ is the bunch charge. For a relativistic beam, the electron distribution function $\lambda_z$ does not change during acceleration, i.e. $z_i = z_f$, and, therefore, eq.17 can be used to define $\delta_i(z_i)$ as a function of $\delta_f(z_f = z_i)$. Thus, beginning with a desirable electron distribution at the end of the linac section, one can find the distribution at the beginning of the linac section that will eventually make it. A similar situation arises in a bunch compressor if the CSR energy change is negligible with respect to that induced by the longitudinal wake field in the linac. Then, the electron coordinate at the beginning of the bunch compressor can be found using the electron coordinate at the end of the bunch compressor using eq.12.

The above considerations justify a concept of reverse tracking [34]. LiTrack [52], a 1-D tracking code, can be used to convolve the actual line-charge distribution with the externally calculated longitudinal wake function. A desirable distribution both flat in energy and current is set up at the end of the accelerator. Starting with this distribution and tracking it backward, the nearly linear ramped peak current shown in Figure 2 is obtained at the start of the accelerator. This

![Fig. 2. Reverse tracking. It begins with “flat-flat” distribution at the end of the accelerator (top line) and moves towards beginning of the accelerator (bottom line). Published in [M. Cornacchia, S. Di Mitri, G. Penco and A. A. Zholents, Phys. Rev. Special Topics - Accel. and Beams, 9, 120701 (2006)].](image)

result can be understood if one uses the wake function for an accelerating structure consisting of an array of cells, eq.8, and convolutes it with a linearly ramped current distribution. The wake potential is highly linear and this is why the final distribution is flat in energy.

Producing a linearly ramped electron bunch current at the exit of the injector is somewhat of a challenge because of the strong nonlinearity of the SC fields at low energy. The longitudinal blow-up of the electrons from the cathode to the first accelerating structure poses a limit to the ramping fraction of the bunch that meets the current linearity requirement. A fourth-degree
polynomial distribution was found in [34] to offer the best cancellation of the high orders nonlinear contributions of the SC field, and thus increases the bunch fraction that follows a linear ramp. This cancellation helps preserving the linearity of the fields in the space-charge dominated part of acceleration.

4. Short-range geometric transverse wake field

4.1 Analytical Model

Similarly to the longitudinal case, the transverse wake function of a linac structure can be approximated with an analytical expression [53]:

\[ w^1(z) = \frac{4s_1Z_0c}{\pi a^4} \left[ 1 - \left( 1 + \sqrt{\frac{z}{s_1}} \right) e^{-\sqrt{z/s_1}} \right] \left[ \frac{V}{C \cdot m^2} \right] \]  

(18)

The transverse motion of a relativistic electron in the linac in the presence of the short-range geometric transverse wake field is described by an ordinary 2nd order differential equation in the complete form. The l.h.s. of this equation is the homogeneous equation for the betatron motion in the horizontal or vertical plane; the r.h.s. contains the convolution of the transverse wake function with the local current distribution and is also linearly proportional to the relative displacement of the particle from the axis of the accelerating structure [54, 55]:

\[ \frac{1}{\gamma(\sigma)} \frac{d}{d\sigma} \left[ \gamma(\sigma) \frac{d}{d\sigma} \sigma \gamma(\sigma) \right] + \kappa(\sigma)^2 \sigma(\sigma, \gamma) = 
\]

\[ e(\sigma) \int_{-\infty}^{L} \sigma(\xi - \zeta') F(\zeta') \left[ x(\sigma, \zeta') - d_e(\sigma) \right] d\zeta' \]

(19)

where \( \sigma = s/L \) is the distance from the linac entrance normalized with the total linac length \( L \); \( \zeta = z/l_b \) is the longitudinal bunch coordinate at location \( \sigma \) measured after the arrival of the bunch head, normalized with the full width bunch length; \( F(\zeta) = I(\zeta)/I_{pk} \) is the local current normalized with the maximum peak current along the bunch; \( \kappa = kL \) is the average normalized focusing strength \( k \) integrated along the linac length \( L \); \( \sigma(\xi, \zeta') \) is the transverse wake function normalized with the wake amplitude; \( d_e \) is the transverse offset of the beam with respect to the linac axis. Finally, \( e(\sigma) = e_r(\gamma_n/\gamma(\sigma)) \) is the factor coupling the particle betatron motion (described by the homogeneous form of the previous equation) to the wake field driving term. It is given by [54]:

\[ e_r = \frac{4\pi e_0 w_n(1)I_{pk}L^2}{I_A \gamma(0)} \]  

(20)

where \( I_A=17 \) kA is the Alfven current, \( w_n(1) \) is the wake function normalized to its amplitude and computed for the particle at the bunch tail, \( I_{pk} \) is the peak current. Unlike the monopole nature of the longitudinal wake field pattern, the short-range geometric transverse wake field is excited by electrons traveling off-axis. When the electron bunch travels near the axis of the accelerating structures, the transverse wake field is dominated by the dipole field component. As a result, the bunch tail oscillates with respect to the head forming in the \((z, x)\) and in the \((z, y)\) plane a characteristic “banana shape” [56]. Persistence of the slice oscillations along the linac and their amplification may cause the conversion of the bunch length into the transverse
dimension (beam break up). So, the displaced bunch tail adds a contribution to the projection of the beam size on the transverse plane, eventually increasing the projected emittance.

4.2 Emittance bumps

The FEL power relies on the energy exchange between the electrons and the light beam along the undulator chain; this interaction is made possible when the two beams overlap. Additionally for HGHG FELs, this is mandatory in the first undulator, where the external seeding laser has to superimpose on the electron bunch. The transverse kick induced by the dipole wake potential imposes an upper limit to the bunch length that is based on the single bunch emittance growth. In order to evaluate this limitation, we recall the approximate transverse emittance dilution through an accelerating structure of length $L$,

$$\Delta \epsilon \approx \left( \frac{\pi r_e}{Z_0 c} \right)^2 \frac{N^2 \langle w \rangle^2 L^2 \beta \Delta^2}{2 \gamma f \gamma f \epsilon}$$

(21)

This is predominantly a linear time-correlated emittance growth and can be corrected. The wake field, $\langle w \rangle$, is expressed here as the approximate average transverse wake function over the bunch given by eqs.18, evaluated at the bunch centroid. Typical misalignment tolerances are in the range $\Delta = 10 \sim 100 \mu m$ in order to ensure $\Delta \epsilon / \epsilon \leq 1\%$ per structure. If the electron bunch and the accelerating structure parameters do not completely fit into the approximated eq.6 the machine design and alignment tolerances are made more robust and reliable by particle tracking studies that include the geometric wake functions and all realistic alignment errors. Computer codes like elegant, PLACET, MTRACK and MBTRACK [58–61] adopt the Courant-Snyder variables to calculate the growth of the bunch slice coordinates caused by a random misalignment of various machine components in the presence of the geometric transverse wake fields. The effect of the wake field can therefore be integrated into the machine error budget.

We are going to show that control over the transverse wake field instability can be gained in a reliable way by applying local trajectory bumps, also called “emittance bumps” [62–66]. Special care is here devoted to the incoherent part of the trajectory distortion due to random misalignment of quadrupole magnets (150 $\mu m$ rms), accelerating structures (300 $\mu m$ rms), Beam Position Monitor (BPM) misalignment (150 $\mu m$ rms) and finite resolution (20 $\mu m$) and beam launching error (150 $\mu m$, 10 $\mu rad$). The wake field effect in the presence of coherent betatron motion of the electron bunch is studied with the elegant code. The simulations show that a global trajectory correction provided through a response matrix algorithm is not sufficient to damp the transverse wake field instability; for this reason local trajectory bumps are applied to suppress it. The bumps technique looks for an empirical “golden” trajectory for which all the kicks generated by the transverse wake field compensate each other and the banana shape is finally canceled. In practice, the implementation of the local bumps foresees the characterization of the transverse beam profile as a function of the bunch longitudinal coordinate (banana shape), projected on screens separated by a proper phase advance (to reconstruct the head-tail oscillation). This could be done in a dedicated diagnostic section, downstream of the linac, by means of RF deflectors [67].
To enhance the wake field instability, we have designed the linac with 14 accelerating structures: the first 7 ones have 10 mm iris radius and bring the electron beam from the initial energy of 100 MeV to 600 MeV; the last 7 structures have the smaller iris radius of 5 mm and increase the energy to 1.2 GeV. The 800 pC, 10 ps long bunch is time-compressed twice, by a factor of 5 at 250 MeV and by a factor of 2 at 700 MeV. Thus, the high impedance (smaller iris) structures are traversed by the higher rigidity, shorter bunch. This scheme is expected to minimize the transverse wake field instability that, nevertheless, has still an impressive effect on the projected emittance, with respect to the early part of the linac. For illustration, only one set of errors—randomly chosen over a meaningful sample of error seeds—is shown in Figure 3. Simulations have been carried out with $2 \cdot 10^5$ particles divided into 30 longitudinal slices.

Fig. 3. Left: the projected emittances blow up as the beam enters into the small iris accelerating structures. The trajectory is corrected everywhere to 200 $\mu$m level. Right: suppression of the transverse wake field instability after some trajectory bumps have been done in the last linac section. Published in P. Craievich, S. Di Mitri and A. A. Zholents, Nucl. Instr. and Methods in Phys. Res. A 604 (2009)

4.3 Slice centroid Courant-Snyder amplitude

As a next step, the validity of the trajectory manipulation is checked in the presence of shot-to-shot trajectory jitter. This can be generated by beam launching error jitter, quadrupole magnet mechanical vibration and power supply current ripple, jitter of the residual dispersion induced by misaligned quadrupoles, energy jitter translating into trajectory jitter through residual dispersion. For a first rough estimation of the instability effect, let us reasonably assume that the transverse beam size, in each plane, is covered by (at least) four standard deviations ($\sigma$) of the particle position distribution. In order for the instability to be suppressed, we want the bunch tail do not laterally exceed the head by more than $1\sigma$. In this case, the relative growth of the beam size is 25%. Equivalently, the relative emittance growth we could tolerate is 50%. Notice that if the instability is suppressed at the linac end, then the slice centroid transverse offset and divergence are small. Hence the bunch tends to maintain its shape in the $(t, x)$ and $(t, y)$ plane at any point of the line downstream. On the contrary, if the banana shape is pronounced, the slice optics in the bunch tail is mismatched to the magnetic lattice. Then, the bunch tail performs betatron oscillations around the head axis.
and the banana shape at any point downstream will depend on the Twiss parameters at the point of observation. For this reason, the Courant-Snyder amplitude of the slice centroid is now introduced \[68\] as a parameter to characterize the instability (same applies to the vertical plane):

\[
\epsilon_{SC} = \gamma x x_{cm}^2 + 2\alpha x x_{cm} x_{cm}^\prime + \beta x x_{cm}^2
\]  \hspace{1cm} (22)

\(\epsilon_{SC}\) is a constant of motion in absence of frictional forces such as geometric wake fields and emission of radiation; this is just the case for the beam transport downstream of the linac, where also coherent and incoherent synchrotron radiation is neglected. \(\epsilon_{SC}\) provides a measurement of the amplitude of motion that is independent of betatron phase. Its square root is proportional to the amplitudes of the slice centroid motion \(x_{SC}(s)\) that describes the banana shape. In general, \(x_{SC}\) is the linear superposition of three main contributions: i) the betatron motion, \(x_{SS}\), generated by focusing of misaligned quadrupoles; ii) the trajectory distortion, \(x_{ST}\); iii) the transverse wake field effect, \(x_{SW}\). Notice that \(x_{offset} = x_{SS} + x_{ST}\) is approximately the same for all slices along the bunch. Regarding the instability, only the motion relative to the bunch head is of interest. Thus, we define a new slice centroid amplitude relative to the motion of the bunch head:

\[
\epsilon_{SW,x} = \gamma x (x_{SC} - x_{offset})^2 + 2\alpha x (x_{SC} - x_{offset})(x_{SC}' - x_{offset}') + \beta x (x_{SC}' - x_{offset}')^2
\]  \hspace{1cm} (23)

The effect of the trajectory jitter on the scheme for the suppression of the instability can be evaluated by looking to the shot-to-shot variation of the centroid amplitude \(\epsilon_{SW,x}\) over the bunch duration. In fact, we require that the standard deviation (over all jitter runs) of the slice lateral deviation be less than the rms (over all particles) beam size \(\sigma_x = \sqrt{\beta_x \epsilon_x}\): \(\epsilon_{SW,x} \leq 1\). We manipulate this expression with the following prescriptions. First, \(x_{offset}\) is a constant. Second, the slice Twiss parameters are the same as the projected ones even in case of slice lateral displacement. Third, the slice Twiss parameters remain constant over all jittered runs. Then, we re-define the variable \(\sqrt{\epsilon_{SW,x}} \equiv Q'_x\) and the previous expression becomes an instability threshold given by the ratio between the standard deviation of \(Q'_x\) and the square root of the rms projected (unperturbed) emittance:

\[
\frac{\sigma_{Q_x}}{\sqrt{\epsilon_x}} \leq 1
\]  \hspace{1cm} (24)

When eq.24 is applied to each slice of the bunch, it is possible to predict which portion of the electron bunch can be safely used for the seeded FEL operation even in the presence of trajectory jitter. When the condition 24 is widely satisfied for most of the bunch slices, that is if the machine error budget and jitter tolerances are respected, we do not expect any important effect of the jitter on the FEL performance.

5. Coherent synchrotron radiation

5.1 Analytical model

The effect of synchrotron radiation is here analyzed for a smooth electron density function, when the emission is at wavelengths of the order of the bunch length, \(l_b\), and much longer than the typical wavelength of incoherent emission: \(\lambda_{CSR} \geq l_b \gg \lambda_{incoh}\), where \(\lambda_{incoh} = \ldots\)
(4\pi R/3\gamma^3), \gamma is the relativistic Lorentz factor and R is the bending radius. The coherent emission is characterized by an intensity spectrum that is proportional to the square of the number of particles N times the single particle intensity, unlike the incoherent emission that is simply linear with the number of particles:

\[
\left(\frac{dI}{d\omega}\right)_{tot} = N(N + 1)|F(\omega)|\left(\frac{dI}{d\omega}\right)_{e}
\]

where \(|F(\omega)|\) is the Fourier transform of the longitudinal particle distribution (form factor); it is of the order of 1 for very short bunches. When \(\lambda_{CSR} \approx l_b\), a cooperative scale length of the process can be defined that describes the interaction of electrons and photons during the emission. This is the “slippage length” \([25]\), \(s_L = \frac{R\theta}{2\pi} + \frac{R^3}{24}\), where \(\theta\) is the bending angle. In this case, the CSR emission depends on the details of the charge distribution, of the geometry of the electrons path and it causes a variation of the electron energy along the bunch (energy chirp). Owing to the fact that the energy variation happens in a dispersive region and that different slices of the bunch are subject to a different energy variation, they start betatron oscillating around new, different dispersive orbits during the emission, thus increasing the projection of beam size on the transverse plane. At the end of compression, the bunch will be suffering of an additional (nonlinear) energy chirp and of a projected emittance growth in the bending plane.

The energy variation along the electron bunch can be evaluated by means of the CSR wake potential. In the “steady-state” approximation, \(R/\gamma^3 \ll l_b \leq s_L\), it can be expressed as follows \([25]\):

\[
W_{CSR}^{SS}(z) = -\frac{1}{4\pi\varepsilon_0} \frac{2e}{3^{1/3}R^{2/3}} \int_{-\infty}^{z} \frac{1}{(z - z')^{1/3}} \frac{d\lambda(z')}{dz'} dz'
\]

The energy loss per unit length of the reference particle due to the radiation emission of the entire bunch is then \(dE/dz = NeW_{CSR}^{SS}(z)\). In \([25]\), the authors distinguish different regimes of CSR emission depending on relation between bunch length, bending magnet length and slippage length. So, using eq.26 in the short bunch \((l_b \leq s_L)\), long magnet \((\gamma\theta \gg 1)\) approximation for a Gaussian line-charge distribution, the induced rms relative energy spread \([26]\) is (in S.I. units):

\[
\sigma_{\delta,CSR} = 0.2459 r_e N R \frac{\theta}{R^{2/3}\sigma_e^{4/3} \gamma}
\]

Eq.26 points out that the energy loss is proportional to the first derivative of the longitudinal charge distribution. So, a stronger CSR induced energy loss is expected, for example, from a Gaussian line-charge than from a uniform one with smooth edges. Also, a current spike in the bunch tail could drive a damaging CSR emission.

When the bunch length is much longer than the slippage length, the afore-mentioned steady-state regime provides incorrect results. Transient effects when the bunch enters and leaves the magnet have to be taken into account \([25]\). Moreover, the electron bunch moves inside the vacuum chamber that acts as a waveguide for the radiation. Not all spectral components of the CSR propagate in the waveguide and therefore the actual radiating energy is smaller than in a free space environment. For an estimation of the shielding effect of vacuum
Design and Simulation Challenges of a Linac-Based Free Electron Laser in the Presence of Collective Effects

chamber, the recipe suggested in [27] is used:

$$\frac{\Delta E_{\text{shielded}}}{\Delta E_{\text{free,pace}}} \approx 4.2 \left( \frac{n_{th}}{n_c} \right)^{5/6} \exp \left(- \frac{2n_{th}}{n_c}\right), \ n_{th} > n_c$$

(28)

Here \(n_{th} = \sqrt{2/3} \left( \pi R / \Delta \right)^{3/2}\) is the threshold harmonic number for a propagating radiation, \(\Delta\) is the vacuum chamber total gap, \(n_c = R / \sigma_c\) is the characteristic harmonic number for a Gaussian longitudinal density distribution with the rms value of \(\sigma_c\). The meaning of \(n_c\) is that the spectral component of the radiation with harmonic numbers beyond \(n_c\) is incoherent.

Figure 4 shows the calculated effect of shielding for a vacuum chamber with \(\Delta = 8\) mm. In case of very wide vacuum chambers (inner radius \(\geq 30\) mm), most of the CSR emission is not shielded when a bunch length of the order of 1 ps is considered.

Fig. 4. Suppression of CSR by the vacuum chamber shielding.

5.2 Emittance growth

The energy loss induced by CSR is inversely proportional to the bunch length. Since in a magnetic chicane the bunch length reaches its minimum already in the third magnet, the global CSR effect is dominated by the energy spread induced in the second half of the chicane. Given the CSR induced energy spread \(\sigma_{\delta,CSR}\), the beam matrix formalism [69] can be used to estimate the projected emittance growth induced by CSR in the transverse phase space:

$$\Delta \epsilon \over \epsilon_0 \approx \frac{1}{2} \beta \theta^2 \sigma_{\delta,CSR}^2$$

(29)

Due to the \(\beta\)-dependence of the emittance growth, an optics design with very small betatron function in the bending plane can help to reduce the CSR effect. This is especially true in the second half of the chicane, where the bunch length reaches its minimum. The physical meaning of this is given by recalling that, for any \(\alpha\), a small \(\beta\)-function corresponds to a high beam angular divergence. If this is large enough, the CSR kick is largely dispersed in the particle divergence distribution – the perturbed beam divergence is computed as the squared
sum of the unperturbed beam divergence and the CSR kick, that can therefore be neglected –
and no relevant CSR effect is observed in the bending plane. Typically, a horizontal betatron
function at level of 1 m limits the relative projected emittance growth to below $\sim 10\%$.

This formalism, however, does not take into account the motion in phase space of the bunch
slices that causes such emittance blow up. In fact, the CSR induced projected emittance growth
is the result of the bunch slices misalignment in the transverse phase space. This misalignment
is meant to be a spatial and an angular offset of each slice centroid respect to the others. This
offset is correlated with the $z$-coordinate along the bunch. In principle, the emittance growth
can be completely canceled out if this correlation is removed. The spatial (angular) offset
evaluated at a certain point of the lattice is the product of $\eta_z (\eta_z')$ with the CSR induced energy
change, integrated over the beam path. If a $\pi$ betatron phase advance is built up between two
points of the lattice at which the beam is emitting CSR in identical conditions, then we have
the integral of an odd function over a half-period and its value is zero [70]. Such a scheme
allows the design of even complex beam transport line (arc or dog-leg like) where a relatively
large number of quadrupole magnets is dedicated to build a $-I$ transport matrix between
successive dipole magnets. Large bending angles, usually translating into short transport
lines, are therefore allowed, even in the presence of high charge, short bunches.

5.3 Numerical methods

We introduce here three particle tracking codes that can be used to support the analytical
study of CSR instability. They are elegant [43], IMPACT [71] and CSRTrack3D [72]. The
flexibility of these codes allows the investigation of the compression scheme and CSR effects
independently from the analytical approximation for the magnet length ($\gamma \phi \gg 1$ or $\ll 1$) or
bunch length ($\sigma_z \gg R\phi^3/24$) [25]. Moreover, the codes allow the simulation of an
arbitrary longitudinal current profile since they convolve the CSR wake function with the
actual current profile at the entrance of the magnetic chicane. elegant implements a 1-D
CSR steady-state and transient force approximation for an arbitrary line-charge distribution
as a function of the position in the bunch and in the magnet; the charge distribution is
assumed unchanged at retarded times [26]. The 1-D model ($r_\tau \ll \sigma_z^{2/3} R^{1/3}$, where $R$
is the orbit radius of curvature) does include neither the effects of the transverse distribution
on the CSR fields nor the field variation across the beam. IMPACT computes quasi-static
3-D SC forces in the linac with the exception of CSR which is treated with the same 1-D
algorithm as in elegant. CSRTrack3D treats sub-bunches of variant shape traveling on
nonlinear trajectories in the compressor. Figure 5 shows the slice emittance distribution (in
the bending plane) after that a 800 pC, 10 ps long bunch has been compressed by a factor of
10 in a symmetric magnetic chicane ($R_{56} = -49$ mm) at the energy of 250 MeV [73]. The good
agreement between IMPACT (courtesy of J. Qiang, LBNL) and elegant demonstrates that
SC forces in the range 100–250 MeV, simulated in IMPACT but not in elegant, do not affect
the compression substantially. At the same time, CSRTrack3D (courtesy of K. Sonnad, LBNL)
predicts some slice emittance bumps due to CSR, but not critical.

Assuming that the injector is able to produce a beam whose parameters satisfy the FEL
requirements, the beam transport and manipulation in the main linac should not degrade
the area in the phase space by more than $\sim 20\%$. Simulations indicate that this threshold
can be satisfied for the longitudinal core of the bunch, while it is harder to apply it when
Fig. 5. Codes benchmarking slice emittance of a 800 pC, 10 ps electron bunch time-compressed $C = 10$ in a magnetic chicane, at 250 MeV. Published in S. Di Mitri et al., Nucl. Instr. and Methods in Phys. Res. A 608 (2009).

also the bunch edges are included. These regions are characterized by a lower charge density, therefore they are subjected to a different dynamics at very low energy, where the beam is generated in the presence of important SC forces that strongly depend on the charge density. The different dynamics of the bunch edges with respect to the core leads to a mismatch of the local distribution function (defined in the transverse and in the longitudinal phase space) with respect to the rest of the bunch. Moreover, the finite length of the bunch enhances a nonlinear behaviour of the space charge electric field at the bunch edges that introduces in turn a local nonlinear energy chirp, which leads to local over-compression and optics mismatch. Thus, we expect a stronger effect of the CSR instability in those regions. At the same time, the very ends of the bunch usually contain a smaller number of particles than the bunch core. This implies a bigger uncertainty in the computation of the beam slice parameters due to numerical sampling errors. For all these reasons, particle dynamics in the bunch head and tail is usually studied only with particle tracking codes and the final beam quality is referred to $\sim 80\%$ of the beam population contained in the bunch core. We finally notice that a large slice emittance at the bunch ends is not a limiting factor for a seeded FEL because those portions of the electron bunch are not foreseen to interact with the external seed laser. Also for a SASE FEL, we expect the amplification process would be greatly suppressed in this area.

6. Microbunching instability

6.1 Analytical model

CSR emission is only one aspect of a more complex dynamics called microbunching instability. This is driven by the interaction and reciprocal amplification of the CSR and Longitudinal Space Charge (LSC) field. The latter determines the variation of particles’ longitudinal momentum. When the beam exits the photoinjector, the SC oscillation period is typically of the order of meters and any beam density modulation is practically frozen. Thus, without loss
of generality, the microbunching instability is assumed to start at the photoinjector exit from a pure density modulation caused by shot noise or unwanted modulation in the photoinjector laser temporal profile. Such density modulation amplitudes are of the order of 0.01% in the sub-micron range and reach \( \sim 1\% \) at longer wavelengths [74]. As the beam travels along the linac, the density modulation leads to an energy modulation via the LSC wake. This is equal to the free space-charge wake for the wavelength of interest:

\[
\lambda_m \ll \frac{2\pi d}{\gamma}
\]

\( d \) being the transverse size of the vacuum chamber and \( \gamma \) the Lorentz factor. The expression for the LSC impedance is [24]:

\[
Z(k) = \frac{iZ_0}{\pi kr_b^2} \left[ 1 - \frac{kr_b}{\gamma} K_1 \left( \frac{kr_b}{\gamma} \right) \right]
\]

where \( Z_0 = 377\Omega \) is the free space impedance, \( r_b \) is the radius of the transverse cross section for a uniform distribution and \( K_1 \) is the modified Bessel function of the second kind.

According to the theory developed in [40], the current spectrum is characterized by a bunching factor:

\[
b(k) = \frac{1}{N\epsilon c} \int I(z) e^{-ikz} dz
\]

where \( N \) is the total number of electrons. \( b(k) \) couples with the LSC impedance along a path \( L \) to produce energy modulation of amplitude [40]:

\[
\Delta \gamma(k) \approx -\frac{I_b(k)}{I_A} \int_0^L \frac{4\pi Z(k,s)}{Z_0} ds
\]

where \( I_A = 17\) kA is the Alfvén current. We now consider that the bunch length is compressed in an achromatic magnetic chicane characterized by a momentum compaction \( R_{56,1} \). For a generic initial energy distribution \( V_0(\delta\gamma/\gamma) \) at the entrance of BC1, the resultant density modulation can be expressed through the bunching factor at the compressed wavelength [40]:

\[
b_1(k_1) = \left[ b_0(k_0) - ik_1 R_{56,1} \frac{\Delta \gamma(k_0)}{\gamma} \right] \int d \left( \frac{\delta\gamma}{\gamma} \right) V_0 \left( \frac{\delta\gamma}{\gamma} \right) e^{-ik_1 R_{56,1} \frac{\delta\gamma}{\gamma}}
\]

where \( k_1 = 2\pi/\lambda_1 = k_0/(1 + h R_{56,1}) \) is the wave number of the modulation after compression; it is equal to the initial wave number \( k_0 \) times the linear compression factor \( C = 1/(1 + h R_{56,1}) \), \( h \) being the linear energy chirp. The bunching evolution in a two-stage compression is obtained by iterating the previous expression:

\[
b_2(k_2) = \left\{ b_0(k_0) - ik_1 R_{56,1} \frac{\Delta \gamma(k_0)}{\gamma} \right\} \int d \left( \frac{\delta\gamma}{\gamma} \right) V_0 \left( \frac{\delta\gamma}{\gamma} \right) e^{-ik_1 R_{56,1} \frac{\delta\gamma}{\gamma}} \left( -ik_2 R_{56,2} \frac{\Delta \gamma(k_1)}{\gamma} \right) \times
\]

\[
\times \int d \frac{\delta\gamma}{\gamma} V_1 \left( \frac{\delta\gamma}{\gamma} \right) \exp \left( -ik_2 R_{56,2} \frac{\delta\gamma}{\gamma} \right)
\]

(35)
where the suffix 2 refers to the BC2 element. So, according to eq. 33 the energy modulation amplitude in front of BC1 is:

$$\Delta \gamma(k_0) = \frac{I_0 b_0(k_0)}{I_A} \int_{0}^{BC1} \frac{4\pi Z(k_0, s)}{Z_0} ds$$  \hspace{1cm} (36)$$

while that in front of BC2 is:

$$\Delta \gamma(k_1) = \frac{I_0 b_1(k_1)}{I_A} \int_{BC2}^{BC1} \frac{4\pi Z(k_1, s)}{Z_0} ds$$  \hspace{1cm} (37)$$

The bunching described by eq. 34 assumes a very simple form for an initial Gaussian energy distribution:

$$b_1(k_1) = \left[ b_0(k_0) - ik_1 R_{56,1} \frac{\Delta \gamma(k_0)}{\gamma} \right] \exp \left[ -\frac{1}{2} \left( k_1 R_{56,1} \frac{\sigma_{\gamma,1}}{\gamma} \right)^2 \right]$$  \hspace{1cm} (38)$$

The present analysis is in the linear approximation because it assumes that the microbunching instability starts from a small energy or density modulation, $|ckR_5 \Delta \gamma/\gamma| \ll 1$. The spectral dependence of the microbunching instability gain in the density modulation can be expressed as the ratio of the final over the initial bunching. In the case of magnetic compression, if the initial bunching term can be neglected with respect to the chicane contribution, the instability is said to be in the “high gain regime”, $G(k) \gg 1$. So, the gain in the density modulation after linear compression, due to an upstream energy modulation and for a Gaussian energy distribution, is given by:

$$G(\lambda) = \left| \frac{b_f(\lambda_f)}{b_i(\lambda_i)} \right| = k_f R_{56} \frac{\Delta \gamma}{\gamma} \exp \left( -\frac{1}{2} k_f R_{56} \frac{\sigma_{\gamma,1}}{\gamma} \right)$$  \hspace{1cm} (39)$$

As a numerical example, we assume an initial shot noise with a constant spectral power and calculate the initial bunching according to the formula:

$$|b|^2 = \frac{\sigma_E^2}{\overline{b}^2} = \frac{2\epsilon}{\overline{b}} \Delta \nu$$  \hspace{1cm} (40)$$

where $\Delta \nu$ is the bandwidth. Then, we convolute it with spectral gain function $G(\lambda)$ to obtain:

$$\left( \frac{\sigma_E}{\overline{E}_0} \right)^2 = \frac{2\epsilon c}{\overline{b} \lambda^2} \int G(\lambda)^2 d\lambda \frac{d\lambda}{\lambda^2}$$  \hspace{1cm} (41)$$

Here we used a substitution $\Delta \nu = c\Delta \lambda/\lambda^2$. The slice energy spread in the electron bunch after magnetic compression can be calculated by assuming that the energy spread induced by the microbunching instability will eventually become uncorrelated energy spread. This gives us a large value, $\sigma_E \approx 4$ MeV, which for a 1.5 GeV FEL is one order of magnitude larger than the specification we have mentioned in Section 1.
6.2 Landau damping

The exponential term of eq. 38 shows that the particle longitudinal phase mixing contributes to the suppression of the instability if the initial uncorrelated energy spread \( \sigma_\gamma / \gamma \) is sufficiently larger than the energy modulation amplitude \( \Delta \gamma / \gamma \). In case of non-reversible particle mixing in the longitudinal phase space, this damping mechanism is called energy Landau damping. The “laser heater” was proposed in [16] in order to have an efficient control over the uncorrelated energy spread with the ability to increase it beyond the original small level. The laser heater consists of an undulator located in a magnetic chicane where a laser interacts with the electron beam, causing an energy modulation within the bunch on the scale of the optical wavelength. The corresponding density modulation is negligible and the coherent energy/position correlation is smeared by the particle motion in the chicane.

In order to demonstrate the effect of the laser heater, we compute the spectral gain function for a few different setting of the laser heater and plot them in Figure 6. The parameters in Table 1 have been used for the computation. It is seen here that the larger the energy spread added by the laser heater the more efficient is the suppression of the gain at the high frequency end of the spectra. We also compute the uncorrelated energy spread at the end of the linac as a function of the energy spread added by the laser heater only with the beam and accelerator parameters listed in Table 1. The analytical result is shown in Figure 6. The calculation is simplified by the fact that the interaction between the laser and the electron beam is weak because the required energy spread is small. In this case the changes in laser and beam dimensions along the interaction region can be neglected. Even the slippage effect is negligible because the slippage length is small with respect to the electron and laser pulse length. The heating process is therefore well described by the small gain theory with a single mode [75].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorr. Energy Spread (rms)</td>
<td>2</td>
<td>keV</td>
</tr>
<tr>
<td>Initial Beam Energy</td>
<td>100</td>
<td>MeV</td>
</tr>
<tr>
<td>Beam Energy at BC1</td>
<td>320</td>
<td>MeV</td>
</tr>
<tr>
<td>( R_{50} ) of BC1</td>
<td>-26</td>
<td>mm</td>
</tr>
<tr>
<td>Lin. Compression Factor in BC1</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>Peak Current after BC1</td>
<td>350</td>
<td>A</td>
</tr>
<tr>
<td>Linac Length up to BC1</td>
<td>30</td>
<td>m</td>
</tr>
<tr>
<td>Lin. Compression Factor in BC2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Beam Energy at BC2</td>
<td>600</td>
<td>MeV</td>
</tr>
<tr>
<td>( R_{50} ) of BC2</td>
<td>-16</td>
<td>mm</td>
</tr>
<tr>
<td>Peak Current after BC2</td>
<td>800</td>
<td>A</td>
</tr>
<tr>
<td>Linac Length up to BC2</td>
<td>50</td>
<td>m</td>
</tr>
<tr>
<td>Linac Length after BC2</td>
<td>70</td>
<td>m</td>
</tr>
</tbody>
</table>

Table 1. Parameters used to compute the microbunching instability gain.

As an alternative to the beam heating, energy modulation and transverse emittance excitation induced by CSR can be moderated, in principle, with an appropriate design of the compressor lattice. Although transverse microbunching radiative effects excite emittance directly [76, 77], an indirect emittance excitation via longitudinal-to-transverse coupling typically dominates.
Design and Simulation Challenges of a Linac-Based Free Electron Laser in the Presence of Collective Effects

Fig. 6. Left: spectral gain function for several beam heating levels. Right: Final uncorrelated energy spread vs. energy spread added by the laser heater. For beam heating weaker than that minimum, the instability is not suppressed and the final uncorrelated energy spread grows because of the energy modulation cumulated at the linac end at very short wavelengths. For stronger beam heating, instead, the final uncorrelated energy spread is dominated by that induced by the laser heater. Owing to the (approximate) preservation of the longitudinal emittance during bunch length compression, the final energy spread is linearly proportional to the initial one.

them. This coupling is characterized by the function:

\[ H = \gamma_x \eta_x^2 + 2 \alpha_x \eta_x \eta_x' + \beta_x \eta_x'^2 \]

where \( \gamma_x, \alpha_x \) and \( \beta_x \) are the Twiss functions and \( \eta_x, \eta_x' \) are the dispersion function and its derivative, all in the horizontal bending plane. Using \( H \), we write for the emittance contribution due to CSR:

\[ \Delta \epsilon_x \approx H \delta^2 \]

where \( \delta \) is the spread of the energy losses caused by CSR. It is obvious from eq.43 that the lattice with small \( H \) gives less emittance excitation. Since the strongest CSR is expected in the third and fourth bending magnet of the chicane where the electron bunch is the shortest, we pursue the compressor design with reduced \( H \) in this magnet. Now we would like to give the argument why we may not want to get the smallest possible \( H \). While moving through the chicane bending magnets, the electrons with different amplitudes of the betatron oscillations follow different paths with path lengths described by the following equation:

\[ \delta l = \int_0^s \frac{x(s')}{R} ds' = x_0 \int_0^s \frac{C(s')}{R} ds' + x'_0 \int_0^s \frac{S(s')}{R} ds' \]

Here \( x_0, x'_0 \) are the electron spatial and angular coordinate at the beginning of the chicane and \( C(s), S(s) \) are the cos-like and sin-like trajectory functions. It can be shown that the rms value of \( \Delta l \) taken over the electrons in any given slice of the electron bunch is related to the electron beam emittance through the function \( H \), i.e.:

\[ \Delta l_{\text{rms}} \approx \sqrt{H \epsilon_x} \]
Thus, the lattice with large $H$ spreads slice electrons more apart than the lattice with small $H$ and washes out the microbunching more effectively. In fact, without accounting for this effect, the gain of the microbunching instability would be significantly overestimated. This effect is very similar to the effect of the Landau damping due to the energy spread. Because of the last argument, it is desirable to design the magnetic compressor such as the magnitude of $H$ in the last bend of the chicane can vary at least within a factor of four. It will give some flexibility to maneuver between such tasks as containing the emittance excitation due to CSR that benefits from smaller $H$ and containing energy spread growth due to the microbunching instability that benefits from larger $H$.

6.3 Numerical methods

Simulation of the microbunching instability with particle tracking codes requires a large number of macroparticles. The microbunching amplitude, $b$, due to shot noise in an electron beam with peak current $I_b$ within the bandwidth $\Delta \lambda$ can be estimated:

$$b = \sqrt{\frac{ec}{I_b \Delta \lambda}}$$

For $I_b = 75$ A and $\Delta \lambda = 10 \mu m$ this formula gives $b = 2.52 \cdot 10^{-4}$. Typically, the microbunching due to granularity of the distribution of macro-particles is much larger. For example, we calculate for a 6 ps long electron bunch (fwhm) with $10^6$ macroparticles, $b = 1.3 \cdot 10^{-2}$, which is approximately 50 times larger than the real shot noise.

There are several solutions to overcome the sampling noise problem. Following [73], we mention three of them: i) a smoothed initial particle distribution is taken as start for \textit{elegant} particle tracking code; the particle binning is then filtered during the simulation. Several tens of million particles representing a 0.8 nC, 10 ps long bunch were tracked on parallel computing platforms to resolve the final modulation at wavelengths of 1-10 $\mu m$ [78]; ii) \textsc{Impact} Particle-In-Cell (PIC) code tracked up to 1 billion particles, thus reducing the numerical sampling noise by brute force. The convergence of the final result for the increasing number of macroparticles was demonstrated in [79]; iii) a 2-D direct Vlasov solver code can be used that is much less sensitive to numerical noise than PIC codes. The 4-D emittance smearing effect is simulated by adding a filter, as shown in [80, 81]. The latter technique follows the evolution of the distribution function using Vlasov’s kinetic equation. Ideally, this is absolutely free from computational noise, although some noise can be introduced on which, due to the final size of the grid, the initial distribution function is defined. However, in practice, this noise can be easily kept below the sensitivity level. It has been demonstrated that the tracking codes results and the analytical evaluation converge with small discrepancy when applied to the beam dynamics in a 1.5 GeV linac, in the presence of a moderate two-stage magnetic compression. In the case of comparison of the simulation results with the linear theory, it becomes apparent that a true result will likely be different because of the anticipation that the linear model should fail at the high frequency end of the noise spectra. Nevertheless, even in the analytical case the result gives a correct assessment of the magnitude of the effect. These techniques have been developed and compared for the first time during the design of the FERMI@Elettra FEL [37]. In that case, \textit{elegant} demonstrated that such a linac-based, soft X-ray facility is very sensitive to small initial density modulations and that
the instability enters into the nonlinear regime as the beam is fully compressed in BC2 [78]. The longitudinal phase space becomes folded and sub-harmonics of the density and energy modulation appear. Consequently, the uncorrelated energy spread produced in the injector region has to be increased with a laser heater. For the same case study, IMPACT and the Vlasov solver predicted [80] that a minimum beam heating of 10 and 15 keV rms, respectively, is necessary to suppress the microbunching instability in the one- and two-stage compression scheme. This led to a final slice energy spread of 110 and 180 keV rms, respectively, with a nominal uncertainty of about 15% from code to code. In spite of the results obtained so far, the microbunching instability study still presents some challenges. In spite of the the Vlasov solver agreement with the linear analytical solution of the integral equation for the bunching factor for a compression factor of 3.5, as shown in [82], entrance into the nonlinear regime is predicted by that code when the compression factor reaches 10 [80]. Unfortunately, the analytical treatment of the nonlinear regime remains a work in progress [83] and no nonlinear analytic treatment of the microbunching instability exists at present for codes benchmarking. Second, the initial seed perturbations for the instability are currently not well determined, both in configuration and in velocity space. Moreover, complications from the bunch compression process, which can lead to “cross-talk” amongst different modulation frequencies, make it difficult to extract the frequency-resolved gain curve. Finally, a fully resolved 3-D simulation of microbunching instability can only be accomplished with massive parallel computing resources that are impractical for the machine fine tuning. As mentioned before, only IMPACT implements 3-D SC forces, while elegant and Vlasov solver adopt a 1-D LSC impedance. However, the substantial agreement between the codes suggests that the 3-D SC effect (which is expected to mitigate the microbunching instability) is probably masked by the differences in the computational methods and in the treatment of the numerical noise.

7. Machine configurations and start-to-end simulations

In spite of the specific features that each new FEL source is showing in its conceptual design, flexibility is still a key word for all existing projects, because it allows facility upgrades, new beam physics and back solutions in case of unexpected behaviours. So, if multiple FEL scheme are usually studied for the same source, the driving linac allows different optical and compression schemes for electron beam manipulation. As an example, a moderate compression factor up to 30 in a ≥ 1 GeV linac could be achieved either with a one-stage or a two-stage magnetic compression scheme. However, the two schemes lead to some differences in the final current shaping, transverse emittance and energy distribution, mainly due to a different balance of the strength of collective effects such as geometric wake fields, CSR emission and microbunching instability, as discussed in [84]. The one-stage compression scheme optimizes the suppression of the instability with respect to the two-stage compression for two reasons: firstly, the phase mixing is more effective in BC1 due to the larger $R_{56}$ and to the larger relative energy spread. Secondly, the absence of the high energy compressor does not provide the opportunity to transform the energy modulation accumulated by LSC downstream of BC1 into current modulation. Another positive aspect of the one-stage compression, performed early enough in the linac, is that of minimizing the effect of the transverse wake field, since the induced wake potential is reduced by a shorter bunch length.
The drawbacks are that a short bunch is affected by longitudinal wake field along a longer path than in the two-stage option, where the path to a short final bunch proceeds in two stages. The wake field corrupts the longitudinal phase space by increasing the energy spread, by reducing the average beam energy and by inducing nonlinearities in the energy distribution. We have seen a manipulated current profile has been studied in [34] to overcome this problem. From the point of view of the stability, the two-stage compression has the intrinsic advantage of self-stabilizing the shot-to-shot variation of the total compression factor, \( C \). Let us assume an RF and/or a time jitter makes the beam more (less) compressed in BC1; a shorter bunch then generates stronger (weaker) longitudinal wake field in the succeeding linac so that the energy chirp at BC2 is smaller (bigger). This in turn leads to a weaker (stronger) compression in BC2 that approximately restores the nominal total \( C \).

A specific application of the magnetic compression in order to suppress the microbunching instability was presented in [84]. After removing the linear energy chirp required for the compression at low energy (BC1), an additional and properly tuned \( R_{56} \) transport matrix element (BC2) is able to dilute the initial energy modulation and to suppress the current spikes created by the microbunching instability without affecting the bunch length. In this case the energy and density modulation washing out is more efficiently provided by two magnetic chicanes having \( R_{56} \) of the same sign. In fact, the energy modulation smearing is induced by a complete rotation of the longitudinal phase space; the two chicanes must therefore stretch the particles in the same direction (see Figure 7).

![Fig. 7. Particle distributions of the bunch core upstream (left) and downstream (right) of BC2. An initial modulation amplitude of 1% is introduced at 30 μm wavelength, corresponding to an initial bunching factor of \( 7 \cdot 10^{-2} \). After BC2, the bunching factor calculated for 3 μm wavelength (\( C = 10 \)) shrinks to \( 3 \cdot 10^{-5} \). The final projected normalized horizontal emittance for 60% of the particles in the transverse phase space is 2 mm mrad. Published in S. Di Mitri, M. Cornacchia, S. Spampinati and S. V. Milton, Phys. Rev. Special Topics - Accel. and Beams, 13, 010702 (2010). Thus we see that a multi-stage compression scheme opens different possibilities to the final beam quality. The eventual machine configuration can be chosen depending on the actual FEL requirements in terms of electron beam quality. Once the configuration is fixed, start-to-end, time-dependent simulations are performed to evaluate the global facility performance in the](image-url)
presence of static imperfections and shot-to-shot jitter sources. This is done by chaining SC codes such as GPT and Astra to a linac code such as _elegant_ or LiTrack, then these to FEL codes like Genesis [85] or Ginger [86]. GPT, LiTrack and Genesis have been used in [73] to calculate the sensitivity of the injector, main linac and FEL output, respectively, to the jitter of the photo-cathode emission time, charge, RF voltage and phase, bunch length, emittance and mean energy. First, it was calculated how large the jitter could be in each parameter independently to cause an rms variation of 10% peak current, 0.1% mean energy and 150 fs arrival time. Then, these uncorrelated sensitivities were summed to generate a linac tolerance budget. Jitter analysis of some slice electron beam parameters was also implemented, with an important parameter being the quadratic energy chirp that affects the FEL output bandwidth. Finally, the tolerance budget was used to simulate shot-to-shot variations of the machine parameters and to perform a global jitter study.

8. Notes on the particle-field interaction

In Section 1 we have mentioned the short range SC forces as one of the Coulomb inter-particle interactions that limits the applicability of the Liouville’s theorem to the particle motion. We want to make here a more precise statement. In general, Liouville’s theorem still applies in the 6-D phase space in the limit of very small correlations established by the space charge forces between particles, so that each particle moves in the same way than all the others, in the collective (also named “mean”) field generated by all the others. Quantitatively, this situation is satisfied if the number of particles in the Debye sphere surrounding any particle is large, that is \( \lambda_D \gg n^{-1/3} \), where \( n \) is the density of charged particles in the configuration space and \( \lambda_D \) is the Debye length that is the ratio of the thermal velocity, \( (KT/m)^{1/2} \), to the plasma frequency \( \omega_p = (q^2n/m\epsilon_0)^{1/2}, q \) and \( m \) being the particle charge and mass. Then, a smoothed out potential due to all particles may be calculated from the density distribution in the configuration space and its contribution included in the Hamiltonian system of forces. This procedure leads to the derivation of the Maxwell-Vlasov equation, which self-consistently describes the behaviour of an assembly of charged particles.

By definition, the SC forces describe a Coulomb interaction within a bunch. Their extension to a train of bunches is straightforward. Typically, being the distance between different bunches of the train much larger than the Debye length, each bunch is treated as independent from the others. This is not the case for the geometric wake fields. If the relaxation time of the wake field is shorter than the repetition time of the accelerator, then the electro-magnetic field associated with two succeeding bunches do not interfere and the single bunch wake field is said to be in the short range regime. This is the case already treated in this Chapter. As opposite, the long range wake field is usually present in rings, recirculating linacs and single-pass linacs dealing with a bunch train. In this regime, different bunches “communicate” through the narrow-band (high \( Q \), quality factor) impedances. That is, wake fields deposited in various high-\( Q \) resonant structures can influence the motion of following bunches and can cause the motion to become unstable if the beam currents are too high. To effectively couple the bunch motion, high order modes must have a damping time \( \tau \approx 2Q/\omega \), where \( \omega \) is the mode resonant frequency, longer than the bunch spacing. For modes with \( Q \leq 100 \), this restricts the frequencies to less than 10 GHz. The frequency limit is lower for smaller \( Q \).
9. References

[27] R. Li, C. L. Bohn and J. J. Bisognano, ParticleAccelerator Conference (1997)
[52] K. Bane, P. Emma, PAC’05, FPAT091, Knoxville, Tennessee (2005)
[58] G. Guignard, CERN-SL/91-19 (AP)
[64] G. Guignard et al., Proc. of PAC’97, 8W015, Vancouver, Canada (1997)
[72] M. Dohlus and T. Limberg in Proc. of the FEL Conf. 2004 (M0C0505)
[79] I. Pogorelov et al., in Proc. of ICAP'06, WEPPP02, Chamonix, France (2006)
Free Electron Lasers consists of 10 chapters, which refer to fundamentals and design of various free electron laser systems, from the infrared to the xuv wavelength regimes. In addition to making a comparison with conventional lasers, a couple of special topics concerning near-field and cavity electro-dynamics, compact and table-top arrangements and strong radiation induced exotic states of matter are analyzed as well. The control and diagnostics of such devices and radiation safety issues are also discussed. Free Electron Lasers provides a selection of research results on these special sources of radiation, concerning basic principles, applications and some interesting new ideas of current interest.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
