BEAM EMITTANCE PRESERVATION IN LINEAR COLLIDERS

G. Guignard
CERN, 1211 Geneva 23, Switzerland

Abstract

The preservation of transverse emittances is critical in linear colliders as perfomance requires small emittances, a large emittance ratio, and small energy spread, in spite of strong wakefields. The damaging effects of these fields are related to the misalignments of the quadrupoles, the accelerating cavities, and the position monitors, and they are present in both single bunch and multibunch modes. In single bunch mode, emittance dilution is reduced by forcing a coherent motion of the bunch, correcting the trajectory, and preventing chromatic effects. The first remedy tends to achieve the same betatron period across the bunch either by imposing a given energy spread or by using RF quadrupoles. The other two are based on step by step trajectory corrections, more global and sophisticated algorithms, non-dispersive bumps, cavity displacements, and separate scaling with energy of the lattice-cell length and the quadrupole strength. In multibunch mode, beam breakup can be controlled either by the use of high-frequency kickers or through an attenuation of the fields generated in the trail of each bunch. As a rule, the latter is possible by damping and detuning the most dangerous dipole components. In both modes, corrections must not upset the energy-spread minimization, and depend on the choice of parameters, linac layout, and positioning tolerances.

Introduction

When the beams collide head-on at the interaction point of a linear collider, the luminosity obtained is directly proportional to the square of the bunch population N_b, to the repetition rate f_{rep}, to the enhancement factors H_x, H_y due to beam disruption, and to the number of bunches k_b, but inversely porportional to the product of the beam sizes $\sigma_x\sigma_y = R\sigma_y^2$ (if R is the beam aspect ratio σ_x/σ_y), i.e.

$$L = \frac{k_b\, N_b^2\, f_{rep}\, H_x H_y}{4\,\pi\, R\, \sigma_y^2} \quad . \tag{1}$$

This equation immediately shows the interest of having a low transverse emittance in order to achieve small beam sizes. The limit here comes from the damping rings presumably used to generate bright beams and can be set between 10^{-6} and 2×10^{-6} rad · m for the total normalized emittance. The actual values considered range from 1.5×10^{-6} to ~ 2×10^{-5} rad · m for the different Linear Colliders (LC) discussed [1] (i.e. NLC at SLAC, JLC at KEK, S-band based SBLC at DESY, the superconducting LC in the TESLA Collaboration, VLEPP in Protvino, and CLIC at CERN).

Given the total emittance, the aspect ratio can be selected, though not necessarily to optimize the luminosity but to reduce all the beam–beam backgrounds, and to improve the resolution on the centre-of-mass energy of the collision. These effects can be characterized by the beamstrahlung parameter Υ (fractional critical photon energy), the fractional energy loss due to radiation δ_B that is a function of the enhancement factor H_Υ for quantum effects and the detector occupancy Ω (number of 2γ events per bunch crossing), which all depend on the parameters entering Eq. (1), in particular on R

$$\Upsilon \propto \frac{\gamma N_b H_y}{\sigma_y\sigma_z}\,\frac{1}{(1+R)} \qquad H_\Upsilon = \left(\frac{1}{1+1.33\,\Upsilon^{2/3}}\right)^2$$
$$\delta_B \propto \frac{\gamma N_b^2 H_y^2}{\sigma_z\sigma_y^2}\,\frac{H_\Upsilon}{(1+R)^2} \qquad \Omega \propto \frac{R\,N_b^2}{f_{rep}(1+R)^2} \quad . \tag{2}$$

Consequently, high R-values lessen the key parameters of Eqs. (2), as required for the experiments. Therefore, all studies of linear colliders are based on flat beams with aspect ratios between 6.5 and 300, depending on the choice of the other parameters.

At constant σ_y, increasing R induces a reduction of L because of (1), although L can be kept the same by choosing different conditions and including the pinch effect. Possible ways to nevertheless boost the luminosity consist of acting on f_{rep}, N_b, and k_b. However, increasing f_{rep} has limitations because of the RF power it would require, and pushing N_b may be restricted by the emittance dilution effects (see below). Therefore, the possibility of functioning in a multibunch mode ($k_b > 1$) is often considered. In this mode, the effective emittance which combines the single bunch emittance with the phase-space walk of all the bunches must be well controlled to preserve the collider performance.

Hence, preserving the single bunch emittance as well as the effective emittance is critical, mainly in the vertical plane since R is larger than 1. And this has to be done in the presence of many sources of imperfections. Considering quasi-static (or slowly drifting) imperfections, one can mention the misalignments of the linac components, possible tilts of these components, wakefields induced by the interaction of the beam with the cavities, and the energy-dispersion of the optics. In this case, one assumes that drifts are slow enough for corrections to be applied, and a variety of correction means have been studied in both single-bunch and multibunch modes. They are reviewed in this paper and illustrated with simulation results. Turning to higher-frequency varying imperfections (jitters), they can be associated with seismic vibrations, technically-induced oscillations (e.g. water cooling), and power supply ripples. For this, fast enough corrections are essentially impossible so that one mainly relies on tight tolerances.

Single Bunch Mode

The presence of strong dipole wakefields implies large kicks originating from misalignments of the cavities and off-centred trajectories. In a single bunch, short-range wakes generated by the leading particles act dowstream, and distort the bunch. To counteract this effect, the well-known technique consists of obtaining a coherent motion by imposing the same oscillation period (or focusing strength k^2) to all particles of the bunch and the condition, called autophasing by its author [2], can be written

$$k^2(z,s) = k_0^2 + \frac{r_0}{\gamma}\int_{-\infty}^{z} \rho\, W_T^{\delta}(z^* - z)dz^* , \qquad (3)$$

where the point-charge wake W_T^{δ} is integrated over the bunch of density $\rho(z)$. All the linear collider proposals (except TESLA where the wakefields are particularly weak) strive to satisfy Eq. (3) or its linearized version (known as BNS damping [3]),

$$\frac{\partial k^2}{\partial z} = \frac{N_b r_0}{\gamma}\frac{\partial W_T}{\partial z} . \qquad (4)$$

Differences arise, however, by the way the variation of k^2 is achieved: a) using the external magnetic focusing, the change can be obtained via an imposed energy spread δ_{BNS} since $k^2 \propto 1/p$ if p is the momentum, and b) the spread in k^2 is created by generating part of the transverse focusing directly from RF fields oscillating at the frequency of the accelerating fields, in so-called microwave quadrupoles [4]. In CLIC, where δ_{BNS} would be large and in conflict with the minimization of the bunch energy spread required for the final focus system, the second solution is proposed.

With the condition of smooth focusing and the equivalence between k^2 and $1/\beta^2$, conditions (3) and (4) indicate that keeping a constant stability margin along the linac implies scaling the β-function with $\sqrt{\gamma}$. With a constrained phase advance, this in turn means that both the focal distance f and the separation L_c of the lattice quadrupoles be scaled with $\sqrt{\gamma}$. However, studies of discrete focusing lattice show that β–function and chromaticity $d\mu/d\delta$ (μ = phase advance, δ = energy deviation) are independent attributes with different sensitivities to a given phase μ. Taking this into account, the new practical form of the BNS damping criterion can be written [5]

$$\frac{1}{2}\frac{\langle \beta \rangle}{\langle d^2\mu / d\delta ds \rangle}\;\frac{e^2 N_b \langle W_T \rangle_{beam}}{\gamma \langle \delta \rangle_{beam}} \cong 1 , \qquad (5)$$

and the $\sqrt{\gamma}$-scaling of β is set aside. In order to balance better the effects of energy dispersion and strong wakefields, and to achieve approximately the phase-advance chromaticity demanded by (5), CLIC introduced a different scaling with energy that involves a variation of μ along the linac. This is done by independent and different scalings of L_c and f, according to Ref. [6]

$$\frac{L_c(s)}{L_0} = \left(\frac{\gamma(s)}{\gamma_0}\right)^{\alpha_a} \quad \frac{f(s)}{f_0} = \left(\frac{\gamma(s)}{\gamma_0}\right)^{1-\alpha_q} . \qquad (6)$$

With $\alpha_a = 0.3$ and $\alpha_q = 0.6$, this gave a 33% gain on the vertical emittance blow-up with respect to standard $\sqrt{\gamma}$ scaling. Figure 1 gives the corresponding Twiss functions plotted against s.

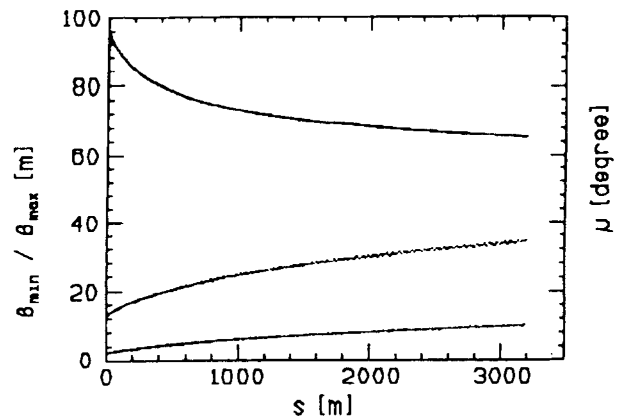

Fig. 1 CLIC Twiss functions with the scaling of Eq. (6).

Although autophasing and appropriate scaling are indispensible in most linear colliders, it is vital to maintain small beam excursions in the elements of the linac. The position tolerances to prevent emittance dilution are therefore very tight. To keep them as reasonable as possible, corrections must be implemented, and those associated with the 'static' case are briefly reviewed:

1) The simplest trajectory correction relies on the individual quadrupoles, displaced transversely to centre the beam in the dowstream beam position monitors (BPM). In a one-to-one scheme, each BPM is close to one quadrupole that is moved to centre the beam in the next immediate BPM. In a one-to-few scheme [6], several BPMs are located between two quadrupoles and each quadrupole is moved to minimize all deviations measured in these BPMs. Iterating this (fast) correction centres the beam in the cavities, and brings the quadrupoles, on the beam path if the BPMs are associated with cavities rather than quadrupoles as in CLIC (strong wakes). This method reduces the orbit by 2 or 3 orders of magnitude and gives alignment tolerances of ~ 5 μm for an acceptable growth of $\gamma\varepsilon_y$, i.e ~ 15% in the NLC, and a factor ≤ 4 in CLIC, for instance.

2) Powerful methods of trajectory correction have been developed at SLAC [7] for compensating the dispersion whilst correcting the orbit and minimizing the wakefield dilutions caused by the corrected trajectory. They are based on minimizing the original orbit, and the differences between 3 or 5 orbits that are created by changing quadrupoles a small amount (10% say, since an error analysis shows that the effectiveness of the technique does not increase very much for larger values [8]). These methods, known under the names of dipersion-free (DF) and Wake-free (WF) corrections, are relatively fast (solution of linear systems involving several correctors, plus iterations), and correct the misalignments at betatron frequency. With BPMs having a resolution better than 1 μm, these methods helped to relax the alignment tolerances to 70 μm in the NLC, and a more modest 10 μm in CLIC (wakefields 20 times stronger). A recent idea [9] that applies to

linacs with strong wakefields consists of trying to compensate in addition an orbit difference created by varying both the bunch population N_b and the bunch length σ_z (the difference between wakes and no wakes). First simulations in the NLC with only cavity misalignments indicate a promising reduction of the emittance dilution with this method.

3) Provided that the dilutions due to energy spread and chromaticity do not induce significant filamentation, non-local corrections of the emittance at diagnostic stations are applicable, assuming the linac is broken into 3 to 6 sectors (to implement the energy scaling [10], for instance). One possibility consists of moving a limited number of cavities (e.g. 19 pairs in the SLC simulations). Another way relies on non-dispersive (ND) bumps [7] (over several oscillations in order to limit their amplitude) with excursions in the structures (Fig. 2) that are optimized by trial and error in order to restrain the emittance growth at the monitors. In the SBLC, ND bumps reduced the blow-up from a factor of 2 to 16% for cavity misalignments of 100 and 60 μm, respectively [11]. In the NLC with 70 μm misalignments, the blow-up went down from a factor of ~ 10 to 10%, while in TESLA (low wakes) the relative emittance growth [12] was reduced by a factor of 5 to 6 (Fig. 3).

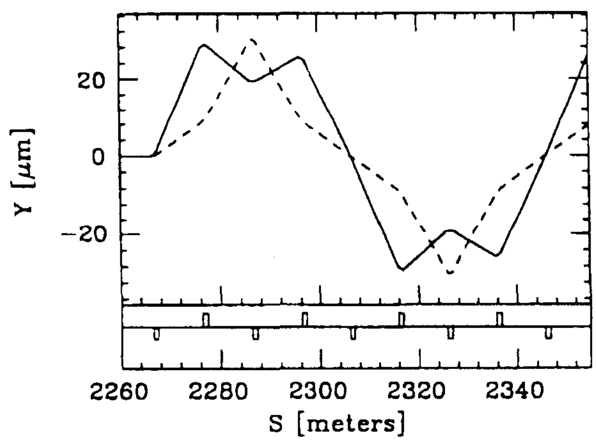

Fig. 2 ND bump (solid line) with a betatron oscillation (dotted) in the NLC.

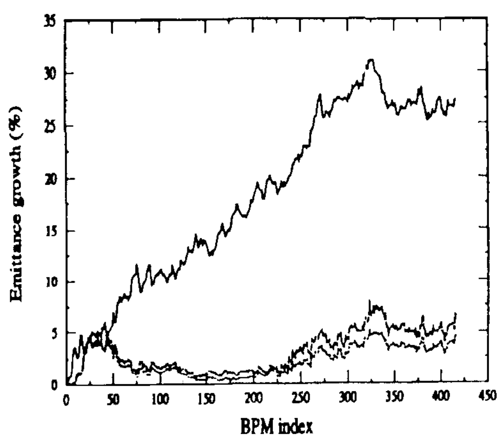

Fig. 3 TESLA emittance growth w/o and with (dotted) ND bumps.

With flat beams, betatron coupling plays an important role in the vertical beam-size growth by exchange of transverse 'energy' (square of the amplitudes). When the lattice is identical in the two planes ($\mu_x = \mu_y$), the effect is enhanced by resonance. Therefore, different phase advance should be selected in each plane using two families of quadrupoles with different focal distances. This was tried in CLIC [10] with ($\mu_x - \mu_y$) = 10°, which then allows an r.m.s. tilt of 1 μrad for the quadrupoles, whilst keeping the emittance target values (1.8 $\times 10^{-6} \times 2\ 10^{-7}$ rad · m). The β-beating in each linac sector (Fig. 4) shows coupling. To reduce it and relax the tolerance, one can envisage coupling correction stations using quadrupole rotations and emittance diagnostics, frequent enough to restrict the filamentation caused by betatron mismatch.

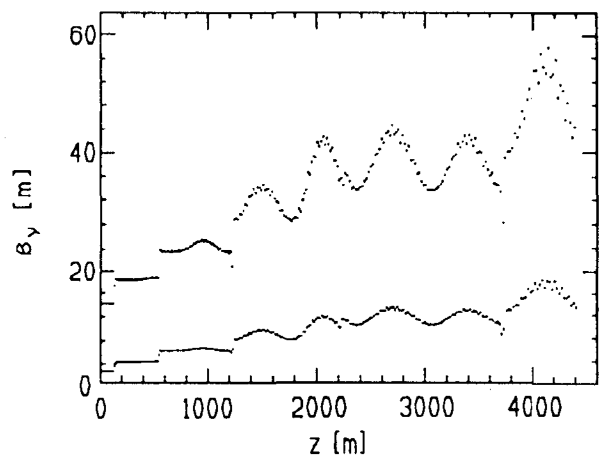

Fig. 4 Betatron beating due to residual coupling in CLIC.

Turning to jitters, i.e. rapid variation in strengths and positions of the active elements, the total emittance is the result of single-bunch dilution and displacement in the phase space of the centre-of-gravity of the bunch as its coordinates change from pulse to pulse. This 'walk' of the bunch ellipse induces a situation where e^+e^- bunches do not collide exactly head-on anymore, which reduces correspondingly the luminosity. The total effective emittance averaged over many pulses can be called 'large' emittance, whilst the term 'small' emittance is reserved for the single bunch [13]. Figure 5 illustrates the transition from the 'small' to the 'large' emittance (30% bigger) in CLIC [13] with random jitters in quadrupole positions of 50 nm (r.m.s.). Limiting the emittance growth to 10%, the tolerances should be ~ 30 nm in CLIC. This value agrees with the 24 nm specified for VLEPP, and the 14 nm considered in the NLC. They are all low enough to suggest the study of active damping of the vibrations (using for instance an accelerometer feeding back a magnet correction coil).

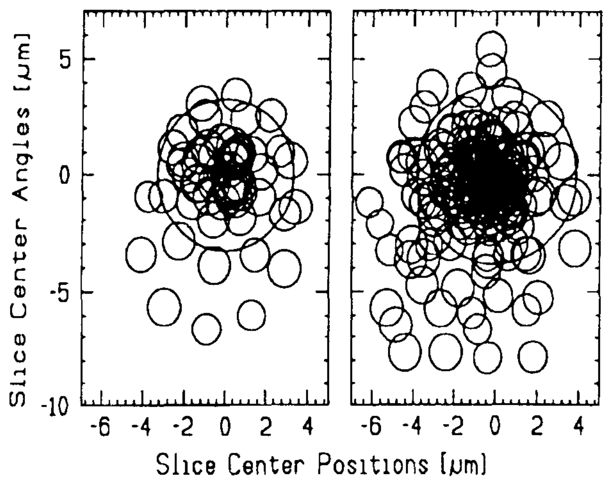

Fig. 5 'Small '(left) and 'Large' (right) emittance with a 50 nm quadrupole jitter in CLIC.

Multibunch Mode

Multibunch mode is a possible strategy aiming at a better luminosity-to-power ratio. However, accelerating a bunch train that extends over one filling time or more, may provoke instabilities and bunch-to-bunch energy variations. Both effects can eventually be responsible for emittance growth and beam breakup.

The longitudinal wakefields are at the origin of inter-bunch beam loading, and the actual RF pulse influences the energy spectrum. Consequent energy variations may induce bunch offsets via the dispersion (and control of the energy is important for the physics experiment). To limit bunch-to-bunch energy spreads, compensation schemes are needed [14]:

1) The most promising scheme is matched filling, i.e. adjustment of the injection timing of the bunch train with respect to the RF pulse and the appropriate choice of bunch spacing. Hence, sufficient extra energy in the RF fill between bunches copes with the energy lost in accelerating the preceding bunches;

2) Staggered timing involves delaying subsets of klystrons so that some accelerating sections are partially filled during the build-up of the beam-loading voltage to its steady-state value;

3) Modulations of the RF input are phase adjustments or small klystron variations when the bunch train passes through cavity sections that compensate the 'sag' one gets in the middle of the train with the matched filling method.

The first scheme applies in principle to all designs, whilst the other two, based on klystron delays and variations, are not appropriate for a two-beam scheme (like CLIC). The last one seems to be best for trains not longer than half of the filling time. In the NLC, the resulting fractional energy deviation for a 90-bunch train is essentially confined between ± 0.001 (Fig. 6) in a long-pulse prefilling compensation scheme [14].

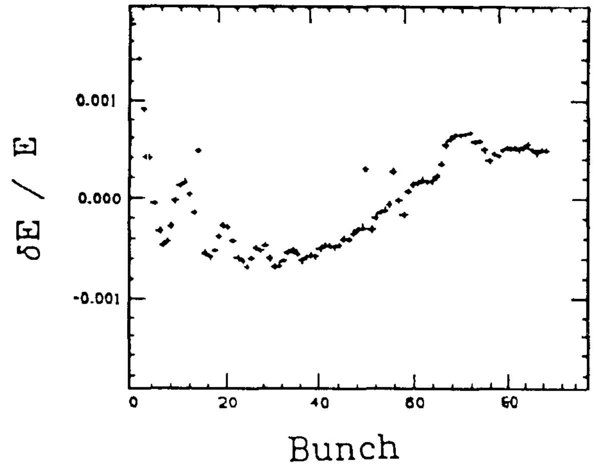

Fig. 6 Energy deviations of bunches in the NLC with the pre-filling compensation scheme.

The transverse wakefields (mainly long-range dipole modes) are directly responsible for emittance growth caused by cumulative beam breakup and transverse bunch offsets in the linac. Transverse modulation is carried from accelerating section to accelerating section, through the beam, and blow-up manifests itself as an amplitude growth from head to tail of the bunch train. Different remedies are possible in order to keep this growth small enough with respect to the single bunch emittance:

1) Damped structures are modified disk-loaded wave guides in which the power of the undesirable wakefield modes is coupled out through radial slots in the disks or azimuthal waveguides, thus permitting the external quality factor Q of these modes to be lowered. This was studied in SLAC and KEK [15], as well as for the lower frequency designs like the SBLC. In the JCL cavities (11.4 GHz) the Q-values of the first synchronous dipole modes have to be of the order of 10 to 100 [16], so that the emittance growth is suppressed to within 10%, with cavity misalignments of 10 μm. For the SBLC (3 GHz), alignment tolerances are about 80 μm if most modes are damped to Q = 2000 in combination with mutual detuning of dipole mode frequencies [11]. Recent investigations for the NLC and wakefield measurements with the test facility ASSET of SLAC [17] reactivated the idea to also combine detuning with such a mode attenuation through wave guides running parallel to the cavity cells and coupled to them [18]. In practice, limitations may come from the low Q-values required, and the large number (at high frequency) of cells involved.

2) Staggered tuning is a variation in the cell dimensions resulting in a cell-to-cell spread (a few per cent) of the dipole mode frequencies. These modes are split into N_f frequency components, whose distribution can be varied. The best one is a truncated Gaussian, giving strong initial roll-off of the wakes,

and low recoherence within the length of the bunch train. This method is considered for all linear colliders using more than one bunch (except TESLA again). Detailed studies [19] were made in particular for the NLC, and a single-particle wake function was calculated for a mean dipole mode frequency of ~ 15 GHz and a fractional spread of ~ 10% per structure. Using four different structure types with interleaved frequencies and repeatedly cycled in the linac, leads to a greater suppression of the wakefield at longer distances [19] (Fig. 7). However, at this level of attenuation, higher dipole bands begin to contribute significantly to the transverse wakes. To separate the critical bands and improve detuning, it has been proposed [20] to vary also the iris thickness from 1.67 to 2.45 mm (keeping constant the fundamental frequency) so that this extra contribution remains within the residual effect of the first band. These techniques are very complex and imply tight tolerances in the construction of the accelerating sections. The ASSET test bench [17] provides an elegant way of measuring the actual wakefields of prototypes of such sections. The first results [21] confirm the expected attenuation, but do not fully corroborate the predicted recoherence.

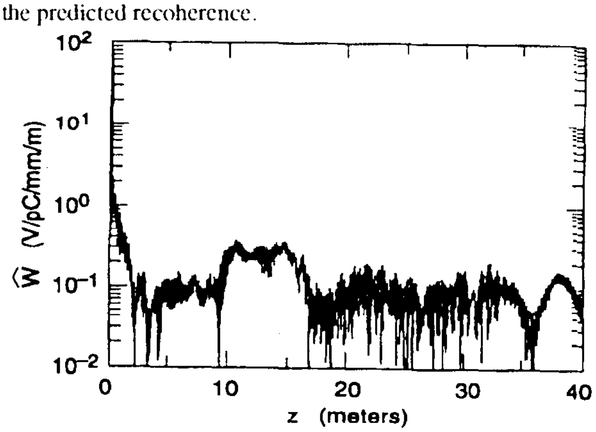

Fig. 7 Envelope of wake function for four structure types in the NLC, with interleaved frequencies.

3) Fast kickers can be used for the realignment of multiple bunches that are scattered by the long range wakefields. Using fast kicker pairs, the kick amplitude imparted to each bunch of the train is adjusted to annul the measurement of a downstream BPM that is phase-shifted by 90°. The NLC [9] needs ≈ 360 MHz kickers for bunch-by-bunch correction, and position monitors with a precision of 0.5 μm at 500 GeV (i.e. $\sigma_y/4$, obtained by averaging over many pulses); the desired feedback should react within seconds and requires several pairs of alignment stations (say 3 to 10). Simulations [22] show (picture unfortunately not available to the author) that with 10 kicker pairs (at 250 MHz) and a cavity misalignment of 50 μm, the emittance smear is reduced to ~ 10 or 20% of the single bunch emittance (from an initial factor of 5). A similar scheme has been tried for TESLA [12]. Although in this case the total dilution is dominated by single-bunch effects and bunch realignment is not indispensable, the multibunch scattering is nevertheless drastically reduced by this technique with 3 stations and 500 μm cavity misalignment (Fig. 8), which illustrates its potentiality.

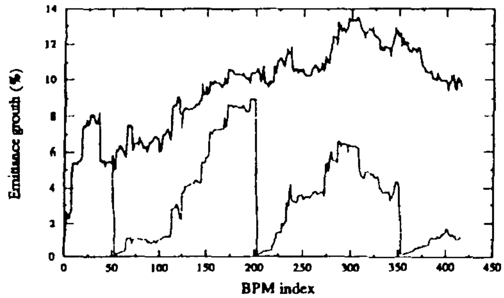

Fig. 8 Multibunch emittance growth before and after (dotted) fast kicker correction in TESLA.

Conclusions

The existence of many methods of correction applying to static perturbations of a single bunch raises the confidence of preserving its transverse emittances, with alignment tolerances between 10 and 100 μm. Nevertheless, high-resolution alignment (correction) systems, accurate position monitors, and several emittance measurement stations are required. The control of jitter effects is however more critical, and mostly relies on tight tolerances (of the order of 10 to 30 nm). In multibunch mode and with static conditions, means of correction exist in principle, but their implementation appears to be difficult and technically complex. Research and development are still needed in these areas to be convinced of the feasibility of the schemes proposed.

References

[1] Proc. of ECFA Workshop LC93, SLAC, 1993.
[2] V. Balakin, Inst. Nucl. Phys., Novosibirsk, Preprint 88-100 (1988).
[3] V. Balakin, A. Novokhatsky, V. Smirnov, Proc. 12th Int. Conf. on High Energy Accelerators, Fermilab, 1983.
[4] W. Schnell, I. Wilson, Proc. Part. Acc. Conf., San Francisco, 1991.
[5] W. Spence, ECFA Workshop E93, KEK, Japan, 1993.
[6] G. Guignard, Proc. Part. Acc. Conf., Washington, 1993.
[7] T. Raubenheimer, SLAC Report 387 (1991).
[8] C. Adolphsen, T. Raubenheimer, Proc. Part. Acc. Conf., Washington, 1993.
[9] T. Raubenheimer, Proc. ECFA Workshop LC93, SLAC, 1993.
[10] G. Parisi, G. Guignard, Proc. EPAC 94, London, 1994.
[11] M. Drevlak, R. Wanzenberg, Proc. EPAC 94, London, 1994.
[12] A. Mosnier, A. Zakharian, Proc. EPAC 94, London, 1994.
[13] G. Guignard, Proc. EPAC 94, London, 1994.
[14] K. Thompson, R. Ruth, Proc. Part. Acc. Conf., Washington, 1993.
[15] H. Deruyter et al., and T. Higo et al., Proc. LINAC 90, Albuquerque, USA, 1990.
[16] T. Taniuchi et al., KEK Preprint 93-81, 1993.
[17] C. Adolphsen, ECFA Workshop LC92, Garmisch-Partenkirchen, 1992.
[18] R. Miller, Private communication.
[19] K. Thompson et al., SLAC-PUB-6032, 1993.
[20] K. Bane et al., SLAC-PUB-6581, 1994.
[21] C. Adolphsen, Private communication.
[22] K. Kubo, Proc. ECFA Workshop LC93, SLAC, 1993.