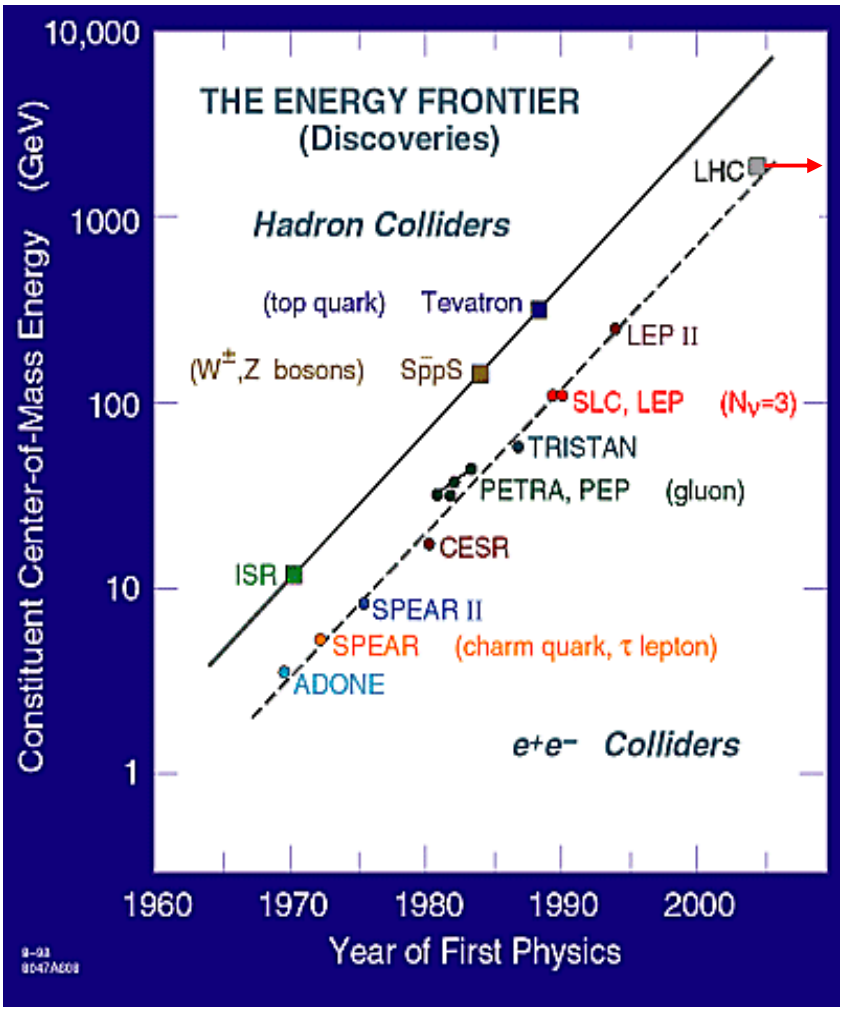
Lecture 1:
Introduction and Overview

Nick Walker, DESY
USPAS, June 16.

The Energy Frontier



500 GeV LC
●
time early!

Moving off the line!

Why a Linear Collider

One can reasonably ask the question "why not just build a bigger storage ring? Why do we need a linear machine that – on the face of it – seems much more complex?" The concept of a linear collider is not particularly new, and was in fact first proposed by M. Tigner in 1965 in a paper entitled "A Possible Apparatus for Electron Clashing-Experiments[1]":

[1] Nuovo Cimento **37** (1965) 1228

"While the storage ring concept for providing clashing-beam experiments is very elegant in concept it seems worth-while at the present juncture to investigate other methods which, while less elegant and superficially more complex may prove more tractable."

So even over thirty years ago it was clear (to some) that the storage ring concept was limited in its energy reach. The reason is due to synchrotron radiation effects and the cost scaling of such facilities.

The LEP collider operated until very recently at CERN in Switzerland is generally considered to be the last energy frontier electron-positron storage ring collider. For a given centre of mass energy, the circumference and cost of the machine is (optimally) defined by the energy loss per turn of the electrons and positrons due to synchrotron radiation. The average power radiated by a single electron (or positron) of energy E for a ring with an average bend field B is

$$P_g = \frac{e^2 c^2}{2\pi} C_g E^2 \left\langle B^2 \right\rangle, \tag{1}$$

where e is the electronic charge, c the velocity of light, and $C_g \approx 8.85\times10^{-5}$ GeV^{-3}m^{-1}. Taking into account the revolution time ($2\pi r / c$, where r is the average ring radius) and the fact that $B = E/(ecr)$, we arrive at

$$\Delta E_{turn} \propto \frac{E^4}{r}. \tag{2}$$

The energy lost per turn has to be replaced by the RF system of the ring, which is a major cost factor for the collider. From equation (2) we see that the cost of the RF system (proportional to the required voltage) is proportional to the fourth power of the energy, and inversely proportional to the radius (circumference) of the ring. The other costs of the collider (vacuum system, magnets, tunnel etc.) scale more or less linearly with the circumference. Thus we have:

- RF costs: $\$_{RF} \propto E^4 / r$
- linear costs: $\$_{lin} \propto r$
- optimum when $\$_{RF} = \$_{lin}$
- optimum cost $(\$_{RF} + \$_{lin}) \propto E^2$

As an example of the cost and circumference scaling, let us take the LEP-II machine as our base, and look at two hypothetical machines: a *super*-LEP with a center of mass energy of 500 GeV, and a *hyper*-LEP at 2 TeV. Table 1 shows the relevant parameters.

		LEP-II	Super-LEP	Hyper-LEP
E_{cm}	GeV	180	500	2000
L	km	27	200	3200
ΔE_{turn}	GeV	1.5	12	240
$\$_{tot}$	10^9 SF	2	15	240

Table 1: Cost and circumference scaling based on the LEP-II storage ring.

It should be clear that – even at 500 GeV centre of mass – a storage ring with a circumference 200 km and a cost of 15 billion SF is unrealistic, and alternatives must be sought.

The main problem is of course that E^4 dependence of the radiated power as we force our electron and positron beams to go around the ring. The answer would seem to be not to bend the beams, but keep them in a straight line, which is exactly the concept of a *linear* collider. Unfortunately, in abandoning our storage ring concept in favour of a linear collider, we immediately arrive at some problems:

- because we cannot store the beams as in a storage ring, a linear collider is a one-pass device where the beams must be accelerated (effectively from rest) to the required energy on each pulse of the machine;
- since we cannot take advantage of the stored beam to slowly ramp the energy up, we must provide several kilometres of linac (RF structures) to achieve the energy in a single-pass.

Hence a linear collider swaps dipole magnets for RF. The typical linac technologies being discussed are on the order of 10 km long, and contain many thousands of RF structures driven by hundreds to thousands of klystrons. Given that we started this discussion by trying to keep the RF costs at an optimum, we might well raise our eyebrows at the linear collider proposals and ask if this is really more cost effective. Indeed, the majority of the linear collider R&D over the last fifteen years has been focused on providing a cost effective technology for the linac. As of writing, the cost advantage over a super-LEP at 500 GeV is estimated at about a factor of two to three, although the tunnel length is considerably less at 20-30 km. Bearing in mind that – unlike the storage ring – the cost scaling for a linear collider is linear with centre of mass energy (you just build more linac), it is very clear that a linear collider is the *only* option for an e^+e^- machine above the TeV range.

			SLC	FLC
centre of mass energy	E_{cm}	GeV	100	500-1000
beam power	P_{beam}	MW	0.04	5-20
vertical beam size at the interaction point	s_y^*	nm	500 (50[2])	1-5
relative beamstrahlung energy loss	dE_{BS}/E	%	0.03	~3-10

[2] 50 nm was the focused beam size achieved at the Final Focus Test Beam (FFTB) at SLAC.

Luminosity	L	10^{34} cm^{-2}s^{-1}	0.0003	~3

Table 2: key parameters for the SLC compared the proposed machines

Current state of linear collider R&D

The Stanford Linear Collider (SLC) was operated at SLAC from 1986 until 1996 and is generally considered a 'proof of principle' of the linear collider concept. The SLC was not, however, a true linear collider since it used a single linac to accelerate both electron and positron bunches, which were separated at the linac exit and then brought into head-on collision by the so-called arcs[3]. However, many of the problems facing the next generation machine were first encountered and solved at the SLC. Table 2 gives a comparison of some key SLC parameters compared to the proposed machines. Clearly what we are currently proposing is major step up from a 'proof of principle' machine, especially the four orders of magnitude in luminosity, which – next to achieving the centre of mass energy – is the most ambitious and challenging goal of linear collider designs. The beam powers are also extremely high and must be dealt with cautiously. Last (but by no means least), achieving and colliding nanometer beams at the interaction point (IP) has been the subject of many man years of on-going R&D.

During and since the SLC there has been much active research on the next generation of linear collider, with a centre of mass energy goal in the range of 0.5-1 TeV (and beyond). To support the R&D activities, several test facilities have been operated at the key laboratories involved in linear collider R&D (CERN, DESY, KEK and SLAC); table 3 summaries these facilities. Together they represent over fifteen years on R&D towards the realisation of a linear collider.

SLC	SLAC, USA	1988-1998	Z factory, proof of LC principle.
Final Focus Test Beam (FFTB)	SLAC, USA	1992-1997	demonstrated required demagnification, achieving 50nm vertical beam size
Next Linear Collider Test Accelerator (NLCTA)	SLAC, USA	1997-	11.4 GHz linac technology test accelerator.
SBAND test facility	DESY, Germany	1994-1998	2.8 GHz S-band test accelerator
TESLA test facility	DESY, Germany	1994-	superconducting 1.3 GHz linac technology test facility and linac. Also includes SASE FEL
Accelerator Test Facility (ATF)	KEK, Japan	1997-	injector and damping ring test facility

[3] the SLC was designed to operate at the Z-pole, and hence the beam energy was about 47 GeV, sufficiently low to allow relatively strong bending magnets.

CLIC Test Facilities (I,II,III)	CERM, CH	1994-	two-beam 30 GHz accelerator test facilities

Table 3: Linear collider R&D test facilities

The Luminosity Issue

There are two critical and key parameters for a high-energy physics collider experiment: centre of mass energy and luminosity. Centre of mass energy is clearly important for the discovery potential of the experiment (new physics), and is generally the first parameter discussed when proposing a new machine (we always tend to refer to being at the 'energy frontier' with these machines). However, the energy reach of the machine is useless unless we can simultaneously deliver a high enough luminosity. The cross-section for physics events generally reduces with the square of the centre of mass energy; consequently we arrive at the following rule-of-thumb for a collider:

$$L \propto E_{cm}^2 . \tag{3}$$

Scaling the achieved LEP luminosity gives us a required luminosity for a 500 GeV centre of mass machine of approximately 10^{34} $cm^{-2}s^{-1}$. Achieving these high luminosities actually dictates many of the fundamental parameters for a linear collider. In the following we will derive a luminosity scaling law and in so doing introduce the fundamental linear collider parameters.

For any colliding beams facility, the luminosity is given by

$$L = \frac{f_c N_b^2}{A}, \tag{4}$$

where f_c is the (mean) bunch collision frequency, N_b are the number of particles per bunch (assumed equal in both beams) and A is the (effective) overlap area of collision at the IP. For a linear collider, trains of n_b bunches are collided at a given repetition frequency f_{rep}, so that $f_c = n_b f_{rep}$. Assuming bunches with Gaussian distributions in both horizontal (x) and vertical (y) planes, we can express equation (4) as

$$L = \frac{n_b N_b^2 f_{rep}}{4p s_x^* s_y^*} H_D . \tag{5}$$

where $s_{x,y}^*$ are the RMS transverse beam sizes at the IP (assumed equal for both beams), and H_D is the *pinch enhancement factor*, which allows for the self-focusing (pinch) of the intense beams during collisions (it has the typical value of ~2). Equation (5) is the usual linear collider luminosity formula quoted in the literature. We will now use it as the basis of our scaling law by introducing the important linear collider parameters.

Centre of mass energy (E_{cm}) and beam power (P_{beam})

The average beam power is given by

$$2P_{beam} = n_b N_b E_{cm} f_{rep} \tag{6}$$

Combining equation (6) with equation (5) yields

$$L = \frac{1}{2p}\left(\frac{P_{beam}}{E_{cm}}\right)\left(\frac{N_b}{s_x^*}\right)\left(\frac{1}{s_y^*}\right)H_D\,. \tag{7}$$

(The reason for the grouping of the remaining parameters will be made clear shortly.) From equation (7) we can see that a higher luminosity can always be achieved by 'cranking up the power'. However higher beam powers are generally constrained by both the available electrical power (including environmental considerations), and the need to deal with the high beam powers within the machine (machine protection and final beam dump issues). Some typical numbers are given below:

N_b	10^{10}
n_b	100
E_{cm}	500 GeV
f_{rep}	100 Hz

$\Big\}\ 2P_{beam} = 8$ MW

This power has to be supplied continuously in order to accelerate each bunch train from rest. The wall-plug power required by the RF is generally much higher than the beam power due to the (in-)efficiency of the RF power sources. We generally write

$$P_{beam} = h_{RF\to beam} P_{RF} \tag{8}$$

where $h_{RF\to beam}$ is the RF to beam power conversion efficiency, which is typically in the range of 20-60% depending on the choice of linac technology. There is a further loss of efficiency in converting the *AC* (or wall-plug) power to RF power (efficiencies in the various components: modulators, klystrons, etc.) . Again these efficiencies tend to be in the 28-40% range. Combining the two efficiencies we arrive at an *AC* to beam power conversion efficiency (which we will refer to simple as h) of 6-24%: hence we need a wall-plug power of >100 MW just to accelerate the beams to the required energy.

Before leaving the beam-power issue, let us introduce the *AC* power (P_{AC}) and the overall efficiency into our luminosity equation (7):

$$L \propto h P_{AC}\left(\frac{1}{E_{cm}}\right)\left(\frac{N_b}{s_x^*}\right)\left(\frac{1}{s_y^*}\right)H_D \tag{9}$$

Intense beams at the IP

In the last section we discussed the power needed to accelerate the beams to the required centre of mass energy. A consequence of the one-pass nature of a 'straight machine' is that must generally have a much lower collision rate than in a storage ring. In LEP – with a 27 km circumference and 4 bunches – the collision rate (per experiment) was 44 kHz; this already represents a factor of ~400 loss in luminosity for a 100 Hz linear collider compared to LEP. Having a large (>100) number of bunches per bunch train gains most of this loss back (at the expense of a high power bill), but we still need at least two orders of magnitude more luminosity at 500 GeV centre of mass. In a linear collider, we achieve this by pushing very hard on the transverse beam sizes at the IP ($s^*_{x,y}$ in equation (9)):

	$s^*_x \times s^*_y$
LEP	$130 \times 6\ \mu m^2$
LC	$(200\text{-}500) \times (3\text{-}5)\ nm^2$

This represents over a 10^6 reduction in beam cross-section at the IP. The required tiny beam sizes in a linear collider have the following (immediate) consequences:

- We now require a very strong focusing (demagnification) of the beam at the IP, which in turn requires very strong focusing quadrupoles close to the IP. Chromatic and geometric aberrations must be cancelled very accurately to avoid dilution of the beam sizes.
- The extreme high charge densities of the colliding beams leads to significant beam-beam effects such as:
    - strong self-focusing (pinch) of the bunches (good);
    - instability effects which lead to tighter collision tolerances on the beams (bad);
    - a high level of *beamstrahlung* radiation which dilutes the luminosity spectrum (bad);
    - production of copious e^-e^+ pairs created by the strong field of the bunches which are a source of background for the detector (bad).
- Tight tolerances on the vibration of the accelerator components, especially the final quadrupole magnets.

Both the strong focusing and the resulting strong beam-beam effects are only achievable in a one-pass machine such as a linear collider; in storage rings, the beam-beam effects are necessarily kept small to avoid loosing the beam!

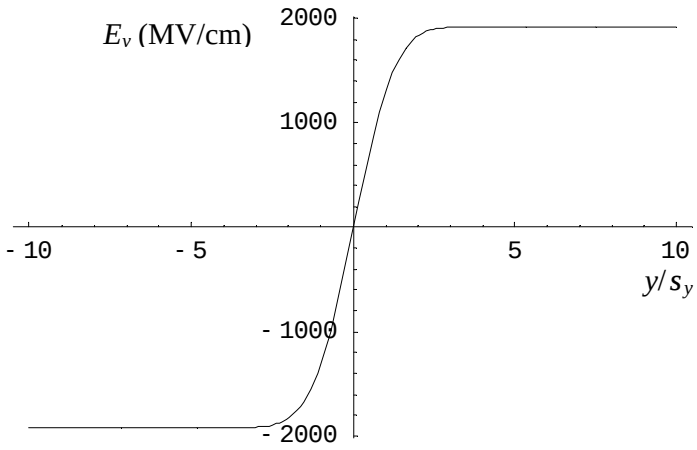


Figure 1: Typical electric field from a flat beam at the IP of a linear collider.

Figure 1 shows the typical electric field of a flat beam ($s_x^* >> s_y^*$) in a linear collider. Clearly fields up to GV/cm can be reached. Particles in the opposing bunch see this field and are deflected by it (the source of the pinch enhancement, characterised by the H_D parameter in the luminosity formula). As the particles are deflected, they radiate hard photons referred to as *beamstrahlung*; this radiation is analogous to synchrotron radiation, although in the intense beam-beam regime the classical synchrotron radiation theory cannot be applied.

The amount of beamstrahlung radiated is a critical linear collider parameter because it (a) quantifies the beam-beam backgrounds and (b) gives an indication of the dilution of the luminosity spectrum (luminosity per centre of mass energy bin). Most linear collider designs constrain the relative beamstrahlung energy loss to a few percent. The more general beam-beam effects are quantified by the so-called disruption parameter ($D_{x,y}$), defined as

$$D_{x,y} = \frac{2r_e N_b s_z}{g s_x (s_x + s_y)} \approx \frac{s_z}{f_{beam}}, \qquad (10)$$

where s_z is the bunch length and $g = E / m_o c^2$. f_{beam} is the effective focal length of the beam; hence a small disruption parameter (weak beam-beam) means that $f_{beam} >> s_z$ and the bunch acts as a 'thin lens'. Conversely a high disruption parameter ($D_{x,y} >> 1$) corresponds to a focal length which is significantly shorter than the bunch length, giving rise to a pinch enhancement and – if D is too big – an instability which significantly reduces the luminosity in the presence of small beam-beam offsets. The enhancement factor (H_D) can be estimated from the following expression:

$$H_{Dx,y} = 1 + D_{x,y}^{1/4} \left(\frac{D_{x,y}^3}{1 + D_{x,y}^3} \right) \left[\ln\left(\sqrt{D_{x,y}} + 1 \right) + 2 \ln\left(\frac{0.8 b_{x,y}}{s_z} \right) \right], \qquad (11)$$

where $b_{x,y}$ are the b-functions at the IP. For most linear collider designs, D_y is in the range 10-20, with H_D typically ~2.

The last term in the square bracket in equation (11) comes from the so-called hour-glass effect, and effectively sets a limit on the achievable beam size for a given bunch length.

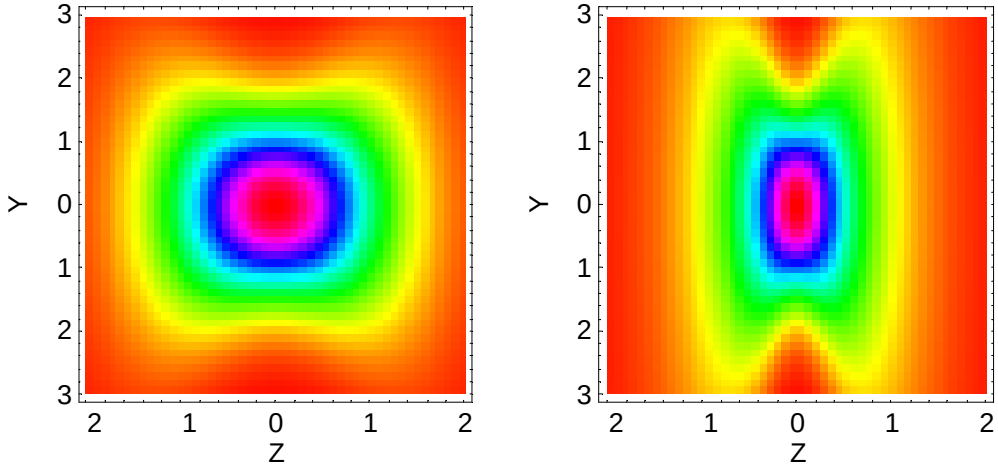


Figure 2: The 'hour glass effect'. Left shows the *z-y* distribution for a beam with $s_z = b_y$, while right shows the *z-y* distribution for $s_z = 3b_y$. Units are in nominal beam s.

Figure 2 shows density plots of the bunch at the IP in the *z-y* plane for the case of (left) $s_z = b_y$ and (right) $s_z = 3b_y$. The latter case shows a marked 'butterfly[4]' distortion compared to the former case, which would reduce the luminosity during collision. b_y can be thought of as a 'depth of focus' for the bunch, and consequently it is desirable to have $s_z \le b_y$.

The relative energy loss during collision due to beamstrahlung is approximately given by

$$d_{BS} \approx 0.86 \frac{e r_e^3}{2 m_0 c^2} \left(\frac{E_{cm}}{s_z} \right) \frac{N_b^2}{\left(s_x + s_y \right)^2} \qquad (12)$$

We would like to keep d_{BS} as small as possible while maximising the luminosity. Comparing equation (12) with our luminosity formula (for example equation (9)), we can immediately spot the standard linear collider trick: d_{BS} is a function of the sum of the two beam sizes, while luminosity is a function of the product. Hence we collide a flat ribbon-like beam with $s_x >> s_y$. As a result, the beamstrahlung is only a function of the horizontal beam size which is then constrained:

$$d_{BS} \approx 0.86 \frac{e r_e^3}{2 m_0 c^2} \left(\frac{E_{cm}}{s_z} \right) \frac{N_b^2}{s_x^2} \quad ; s_x >> s_y \qquad (13)$$

[4] to see the 'hour glass' you must rotate your head 90°; I have never understood why this is not called the 'butterfly' effect.

We then increase the luminosity (independently of the beamstrahlung) by making s_y as small as possible. Combining equations (13) and (9) we can express our luminosity scaling law in terms of the beamstrahlung energy loss d_{BS}:

$$L \propto \frac{h P_{AC}}{E_{cm}^{3/2}} \frac{\sqrt{d_{BS} s_z}}{s_y} \tag{14}$$

Summarising (so far), equation (14) indicates that for high luminosity operation we need:

- a high beam power (high P_{AC});
- a high wall-plug to beam power transfer efficiency h;
- small vertical beam size s_y at the IP;
- long bunch length[5] s_z;

In addition, a higher luminosity can be achieved at the expense of a larger energy loss due to beamstrahlung, providing we are ready to live with the consequences.

Equation (14) is almost – but not quite – our final scaling law. We have yet to introduce the important parameter of vertical emittance, and in doing so, we will see that we can use our hour-glass constraint to effectively remove the bunch length from the equation.

From basic linear optics, we can express the vertical beam size at the IP by

$$s_y = \sqrt{\frac{b_y e_y}{g}}, \tag{15}$$

where e_y is the *normalised* or *invariant* emittance, and $g = E / m_o c^2$ as before. From our discussion of the hour-glass effect, we already know that the bunch length and the vertical b-function are constrained by the relationship $s_z \le b_y$. To maximise the luminosity a sensible choice (limit) would be to set $b_y = s_z$. Using this last relationship, and replacing s_y by equation (15) in equation (14), we arrive at our final luminosity scaling law:

$$L \propto \frac{h P_{AC}}{E_{cm}} \sqrt{\frac{d_{BS}}{e_y}} H_D \quad \text{with} \quad b_y = s_z \tag{16}$$

We have re-introduced the enhancement factor for completeness. We have now successfully identified the key (fundamental) linear collider parameters which enter into the luminosity scaling, namely power conversion efficiency (h); power (P_{AC}); beamstrahlung energy loss (d_{BS}); normalised vertical emittance (e_y); and bunch

[5] we will see shortly that this is not the case.

length (s_z), which sets the scale for the vertical b-function. In addition we can now re-state our requirements for high-luminosity (for a given centre of mass energy and beamstrahlung energy loss):

- a high beam power (high P_{AC});
- a high wall-plug to beam power transfer efficiency h;
- small normalised vertical emittance e_y;
- a short bunch length s_z (and corresponding small b_y);

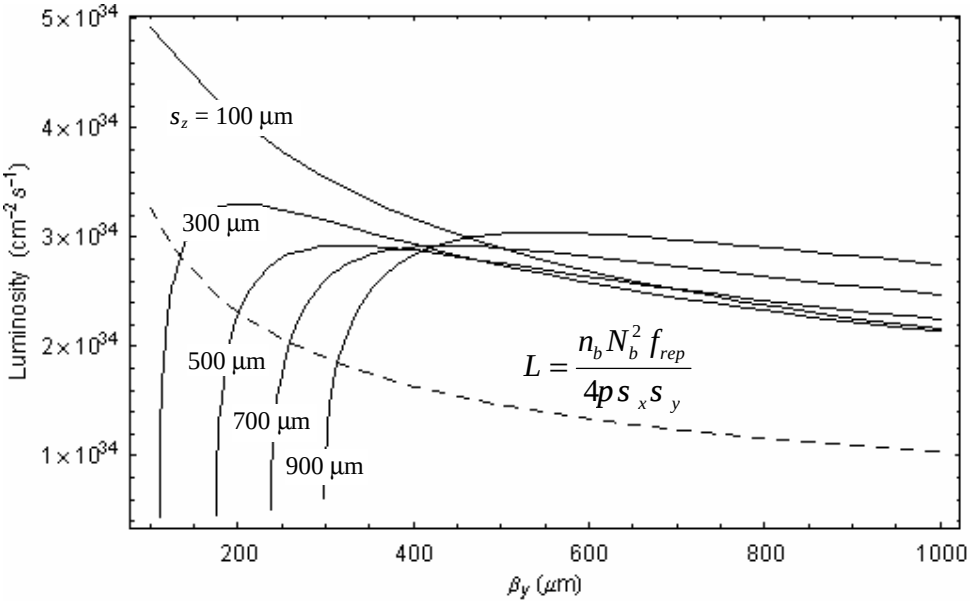


Figure 3: Luminosity as a function of b_y for various bunch lengths. Dotted line shows the *geometric* luminosity.

Figure 3 shows the luminosity as a function of b_y^* for various bunch lengths; the dotted line shows the geometric luminosity (*i.e.* no beam-beam effects, equation (5) with $H_D = 1$), which shows the expected $\sqrt{b_y^*}$ behaviour. The solid lines are calculated from equations (5) with H_D calculated from equation (11); we can make the following observations:

- for $b_y^* > s_z$, a clear pinch enhancement is visible over the geometric luminosity (dotted line);
- for $b_y^* < s_z$ the luminosity drops rapidly due to the onset of the hour-glass effect.

While it appears that a shorter bunch length is desirable, we should not forget that this is general accompanied by higher beamstrahlung (equation (12)).

Introduction to the Generic Linear Collider

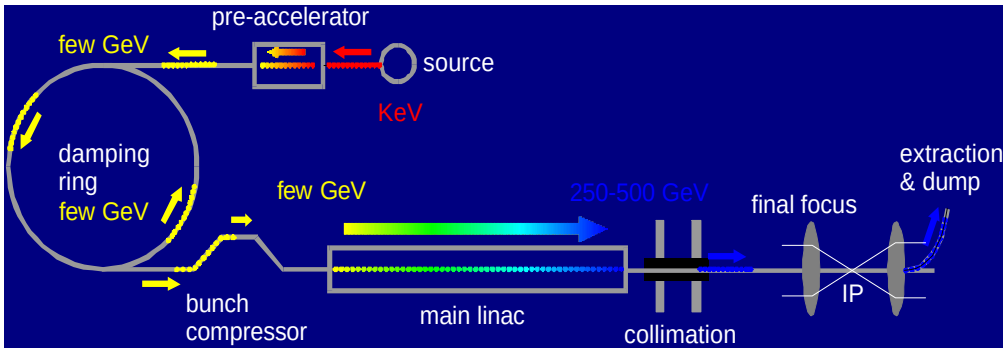


Figure 4: The Generic Linear Collider.

In the last section we introduced the important beam parameters for a linear collider via the important issue of luminosity. In this section we will briefly describe how these parameters are obtained by outlining the various sub-systems of a generic linear collider; this section will act as an introduction to the more detailed descriptions to follow in the relevant lecture units.

Figure 4 shows schematically a generic linear collider (or at least one half of it). In the sense of the beam, the various subsystems are:

- source (provides the required electrons and positron bunches with the required time structure);
- pre-accelerator which accelerates the bunches to the damping ring energy;
- a damping ring (or rings), which 'damp' (or reduce) the phase volume of the bunches;
- a bunch compressor which compresses the bunches longitudinally to the required IP bunch length;
- the main linac, which accelerates the bunches from the damping ring energy up to the desired IP energy;
- the beam delivery system (BDS) which transports the high-energy bunches to the IP where they are collided;
- an extraction line which safely transports the 'used' bunches to the dump (and optionally supports post-IP beam diagnostics).

In Figure 4, the BDS is further divided into two important sub-systems:

- a post-linac collimation system which is needed to remove the beam 'halo' which would otherwise cause unacceptable background in the detector;
- the Final Focus System (FFS), which supplies the strong focusing required to produce the nanometer-sized beams at the IP.

We will now briefly discuss each of these sub-systems. As the linac and its associated technology is so central to the linear collider (and represents the major component costs) we will deal with it first.

The Main Linac and Acceleration

Accelerating Field

The electron and positron bunches are accelerated by (*i.e.* they gain energy from) RF fields inside so-called structures. Structures can either be waveguide-like structures or resonant cavities. In both cases, the structures/cavities are so designed that the fundamental mode consists of a longitudinal electric field (E_z). There are two basic ways of using an accelerating structure: the travelling-wave (TW) or standing-wave (SW) mode.

In the TW mode, the e-m wave travels along a waveguide-like structure, with power being fed into the upstream end (see Figure 5).

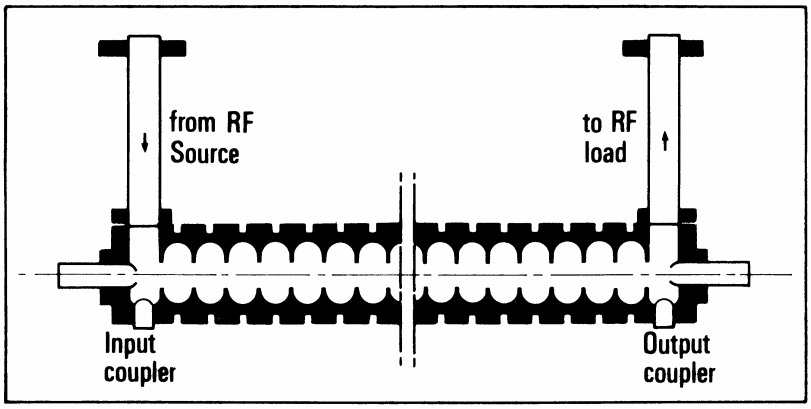


Figure 5: travelling wave structure

Providing the *phase velocity* ($v_p = w / k$) is equal to the velocity of the particles (assumed here to be the velocity of light, *c*), then the particles maintain a constant phase relation as they pass through the structure:

$$E_z(s) = E_0 \cos(f), \tag{17}$$

where *s* is the longitudinal position along the structure (= *ct*), f is RF phase, and E_0 is the peak electric field along the structure. The maximum energy gain is when f = 0, but we will see later that a non-zero (synchronous) phase is generally required for emittance preservation.

The fundamental issue in the design of a TW structure is to have an accelerating mode which has the correct phase velocity. The lowest order accelerating mode in a uniform circular waveguide is TM_{01}, but this mode has a phase velocity that is greater than *c* (remember that the product of the *group* and *phase* velocities $v_g v_p = c$ for this simple geometry, and hence $v_g < c$). To make use of this mode as an accelerating structure, we must somehow slow the wave down. In practise this is achieved by periodically inserting irises as shown in Figure 6.

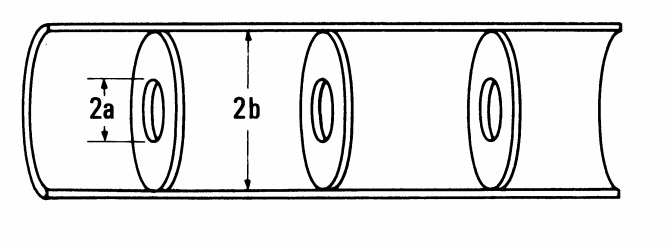


Figure 6: disk loaded waveguide

The disks act like capacitive loads in a transmission line and slow down the propagation of the wave. By tailoring the dimensions *a* and *b* the correct phase velocity for the accelerating mode can be achieved.

In *standing wave* structures (cavities), the particles see a time varying field:

$$\begin{aligned} E_z(s) &= E_0 \sin(wt + f)\sin(kz) \\ &= E_0 \sin(kz + f)\sin(kz) \end{aligned} \tag{18}$$

Acceleration is no longer constant as with the TW case, but varies along the cavity. The length of the cavities is $l/2$ so that the particle always sees an accelerating voltage as it passes from cavity to cavity. The total voltage seen per cavity is the integral of (18):

$$\begin{aligned} \Delta V_{cav} &= E_0 \int_0^{L_{cav}} \sin(kz + f)\sin(kz)dz \\ &= E_0 \int_0^{l/2} \sin(kz + f)\sin(kz)dz \\ &= \tfrac{1}{4}\, l\, E_0 \end{aligned} \tag{19}$$

Hence the effective gradient is $\Delta V_{cav}/(l/2) = E_0/2$, or half the peek field.

Important cavity parameters

An important quantity for cavity performance – and particularly efficiency – is the shunt impedance unit length (r_s), defined as

$$\frac{dP}{dz} = -\frac{E_0^2}{r_s} \tag{20}$$

where dP/dz is the power lost in the structure walls per unit length. The shunt impedance tells us how much power you need to feed a structure to maintain a specified field. Ideally we would like to make the shunt impedance as high as possible to reduce the amount of power needed to maintain the field. Note that equation (20) refers to the case of zero beam loading (no beam); this is just the power needed by the structure itself. Note that if a standing wave structure with shunt impedance r_s is used in a travelling wave mode, the shunt impedance is doubled; this is because a standing wave can be though of as the superposition of a forward and a backward travelling wave.

Another important parameter is the Q of the cavity which is defined as

$$Q \equiv 2p \, \frac{\text{stored energy}}{\text{energy lost per cycle}} \, . \tag{21}$$

If w_s is the stored energy per unit length of a structure, then

$$Q = -\frac{w \, w_s}{dP/dz} \tag{22}$$

Combining (22) with (20) gives

$$\frac{r_s}{Q} = \frac{E_0^2}{w \, w_s} \tag{23}$$

The quantity r_s / Q is important and depends only on the geometry of the cavity/structure, and not on the material or surface properties.

The shunt impedance scales with operating frequency as

$$r_s \propto \begin{cases} f^{+1/2} & \text{normal conducting} \\ f^{-1} & \text{superconducting} \end{cases} \tag{24}$$

For this more than any other reason, LC designs based on conventional (normal conducting) RF have pushed to higher frequencies (JLC/NLC 11.2 GHz, CLIC 30 GHz). For Superconducting cavities, lower frequencies are more efficient, the optimum being close to 1.3 GHz (as chosen for TESLA).

The frequency scaling of the Q-factor is

$$Q \propto \begin{cases} f^{-1/2} & \text{normal conducting} \\ f^{-2} & \text{superconducting} \end{cases} \tag{25}$$

Thus for normal conducting, the higher shunt impedance comes at the cost of a reduce Q. Finally the ratio r_s / Q scales as

$$\frac{r_s}{Q} \propto \begin{cases} f & \text{normal conducting} \\ f & \text{superconducting} \end{cases} \tag{26}$$

As expected, this scaling is the same for both normal and superconducting cavities, since r_s / Q is independent of material or surface properties.

Finally, we need to introduce the *fill time* t_f of the cavity, which is the time required for the cavity to reach the required voltage. For a TW structure, it is defined as the

length of time to fill the structure with energy (L/v_g). In terms of Q and ω it can be expressed as

$$t_f = t_0 \frac{2Q}{w} \tag{27}$$

where t_0 is the structure *attenuation* coefficient defined as

$$P_L = P_0 e^{-2t_0} \tag{28}$$

where P_0 and P_L are the input and output RF power respectively. For a SW cavity, the fill time is defined as the time for the field to charge up to 1/*e* of its final value:

$$t_f = \frac{2Q}{w} \tag{29}$$

which is the same as TW result (27) but without the attenuation factor.

Beam loading

So far we have only discussed the issue of achieving the desired RF gradient in the structure or cavity. We have seen that, with normal conducting cavities, a significant RF power is required to maintain the gradient due to energy loss in the structure walls. When we now inject a beam, the electron bunches will be accelerated and gain energy from the field; this energy must be replaced by the power source (klystrons), or a drop in the structure voltage will occur. We refer to this effect as beam loading.

Mathematically, we can write the power loss per unit meter in a structure as

$$\frac{dP}{dz} = -\frac{E_z^2}{r_s} - i_b E_z \tag{30}$$

where i_b is the peak beam current. The first term on the right hand side of (30) is just equation (20), and represents the power lost to the cavity walls. The second term is the power removed by the beam (the beam loading). For warm (conventional) RF, the wall power term dominates, while for superconducting RF, the beam loading term almost completely dominates (i.e. all the power goes into the beam).

For conventional RF, the beam loading effectively reduces the gradient (voltage) seen by the beam. We refer to the *loaded gradient* as opposed to the *unloaded gradient*: for JLC/NLC, the unloaded gradient is ~65 MV/m, while the loaded gradient is ~50 MV/m; we can interpret this as the beam *current* generating a back-phased electric field of –15 MV/m.

Figure 7 shows the gradient (electric field) along the an NLC-like structure.

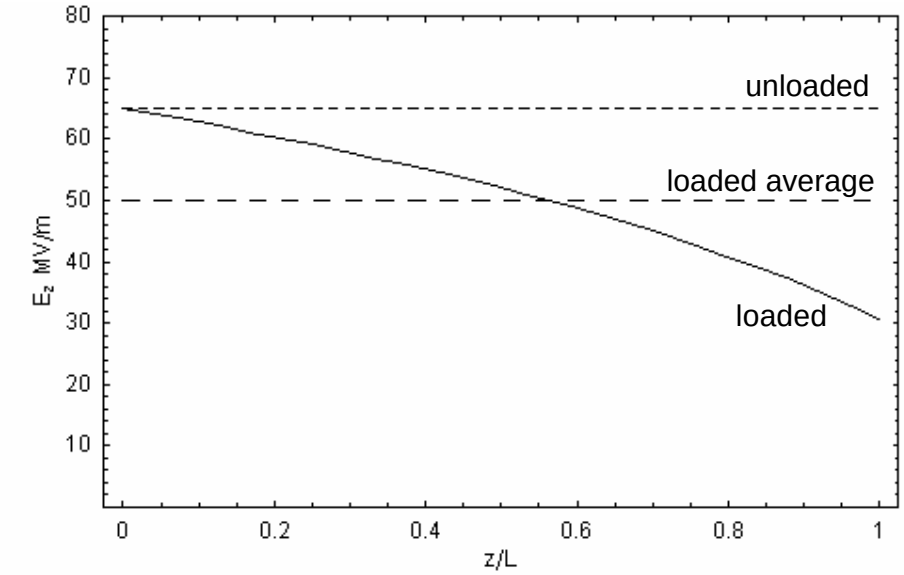


Figure 7: Electric field along an NLC-like structure.

We can clearly see how the passage of the beam causes the gradient to decrease along the length of the structure. We refer to the average as the loaded gradient. The difference between the loaded and unloaded gradient is given by

$$E_{z,u} - E_{z,l} = \frac{1}{2} i_b r_s \left(\frac{2t_0 + e^{-2t_0} + 1}{1 - e^{-2t_0}} \right) \tag{31}$$

Transient Beam Loading

Equation (31) and the curve in Figure 7 represent the steady state solution, where the beam current i_b is considered to be continuous. In a linear collider, the beam current is not continuous, but pulsed at the repetition rate. However the steady state approximation is very good for the long bunch trains typical of the current LC designs.

Unfortunately, the transient behaviour cannot be ignored, otherwise the energy difference (spread) over the bunch train would be too severe. The transient voltage manifests itself at the beginning of the bunch train, before the steady state voltage is established. We can think of the first bunches extracting a portion of the stored energy from the structure, leaving a lower voltage for the trailing bunches. The first bunch sees the full unloaded gradient, while the trailing bunches see the steady state loaded gradient. Therefore the voltage transient over the beginning of the bunch train spans the full range between loaded and unloaded gradient (15 MV/m for an NLC-like structure).

Transient beam loading is compensated by arranging for the (unloaded) gradient to increase over the bunch train, in such a way that the transient behaviour is compensated. The beam loading is dependent on the current (bunch charge), which is generally measured dynamically in the damping ring before the pulse is extracted. In

this fashion the optimum compensation can be applied on a pulse by pulse basis. Compensation to the level of a few parts in 10^3 have been achieved in conventional RF systems.

For superconducting systems, the situation is a little different. Since no power goes into the cavity walls, a long RF pulse can be used with many bunches (TESLA, has 2820 bunches in 950 μs, corresponding to bunch separation of 337 ns). On these time scales, fast feedback can be used to adjust the cavity voltage during the pulse itself. Such systems have been demonstrated to achieve a few parts in 10^4 stability over the bunch train.

Single bunch loading

Due to the finite length of a bunch in the bunch train, there is an additional single bunch beam loading effect, where the head of the bunch effectively decelerates the tail. Single bunch effects are generally calculated in the time domain by using a longitudinal wake field potential.

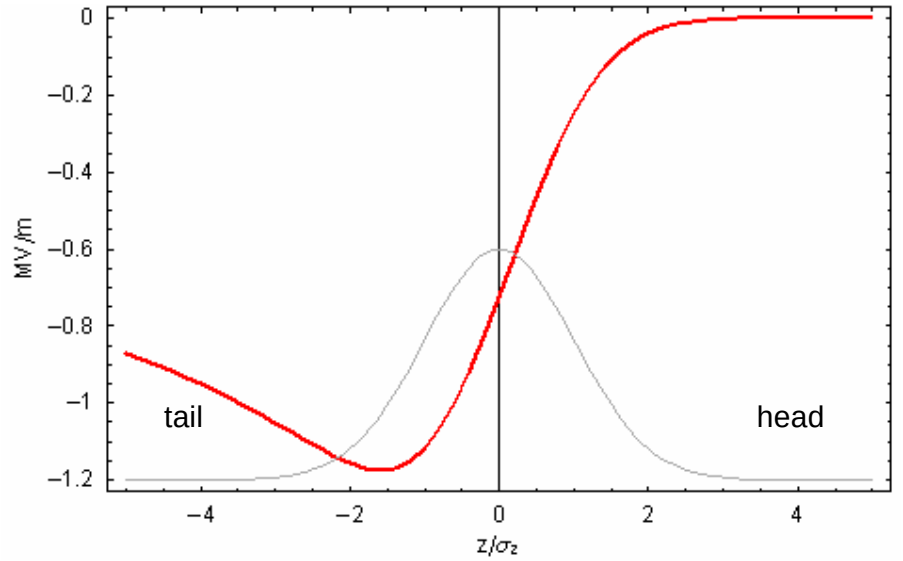


Figure 8: Longitudinal wakefield for an NLC X-band structure. The bunch length is s_z = 110 μm, and the bunch charge is $0.75 \times 10^{10} e$. The grey curve indicates the Gaussian longitudinal bunch profile.

Figure 8 shows the single bunch longitudinal wake for the NLC X-band structure. The average energy loss of the bunch per structure per unit charge is referred to as the *loss parameter K:*

$$K = \frac{1}{Q}\int_{-\infty}^{+\infty} W_{\square}(z)\, r(z)\, dz \qquad (32)$$

where $W_{\square}(z)$ is the single bunch longitudinal wake potential, and $r(z)$ is the longitudinal charge distribution (Q is the total bunch charge). For the NLC structure it is ~5.7×10^{14} $VpC^{-1}m^{-1}$ (assuming a 110 μm Gaussian bunch), or ~0.7 MV/m for the nominal bunch.

As with the multi-bunch transient beam loading, the single bunch beam loading must also be compensated to prevent an excessive energy spread within the bunch. This is achieved by riding the bunch slightly in front of the RF crest at some phase angle f. The slope of the RF at that point can be used to compensate the head-tail beam loading induced energy drop over the bunch. Figure 9 shows the effective gradient along the bunch for the optimum phase angle of –15.5° in the case of an NLC X-band structure. Figure 10 shows the mean (effective) gradient and the relative RMS energy spread of a single NLC bunch as a function of RF phase angle: the minimum energy spread is achieved at the optimum angle plotted in Figure 9.

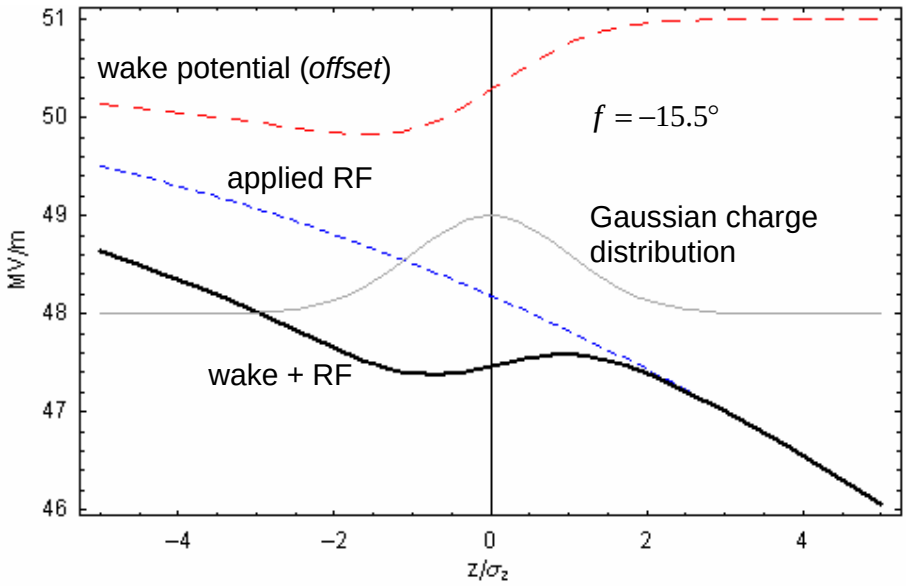


Figure 9: Example of single bunch beam loading compensation in the NLC using the RF curvature. The wake potential has been vertically offset to fit on the plot.

Transverse Wakefields

When a bunch travels through a structure with a transverse offset with respect to the structure axis, the bunch induces transverse modes which then act back on the beam. In the case of a single bunch, the modes deflect the tail of the bunch; In the multi-bunch case the modes generated by earlier bunches deflect the later ones. If not compensated or corrected for, the transverse modes will lead to a phenomenon known as beam break-up, (single-bunch or multi-bunch), which destroys the transverse beam quality (emittance).

The magnitude of the transverse wakefields are a strong function of the iris radius (a in Figure 6), or more precisely the ratio a/l. Although the exact dependence varies with the details of the structure design, the transverse wakefield has been shown to scale approximately like $a^{-3.5} l^{-0.5}$; since the dimensions of the structures (including a) scale with inversely with the RF frequency (f), the transverse wakefields ($w_\perp$) scale roughly as $w_\perp \propto f^3$. The push towards higher frequencies for greater RF efficiency

(higher shunt impedance) and high gradients (shorter linacs) thus comes at the price of much stronger wakefields, ultimately leading to tighter tolerances.

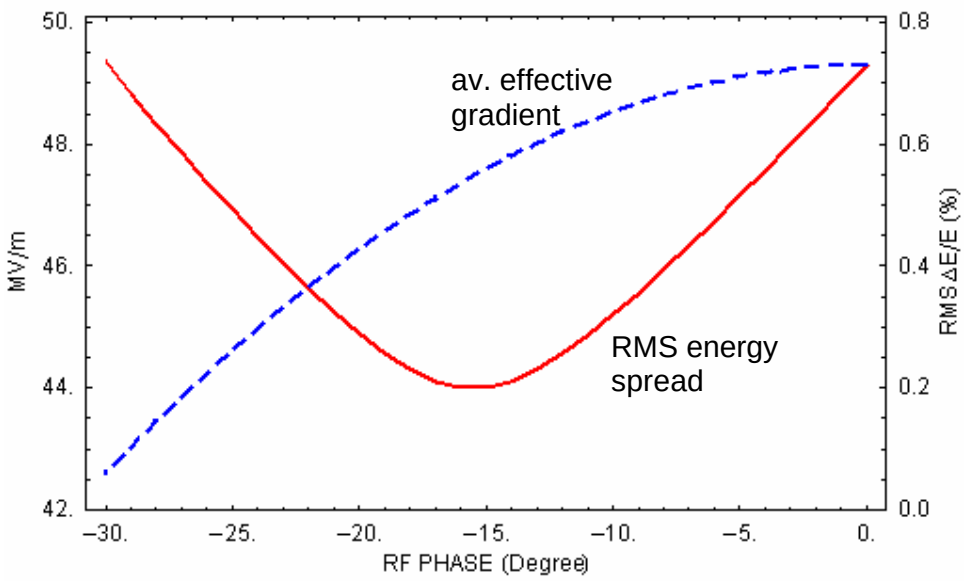


Figure 10: Mean effective gradient and RMS energy spread of a single NLC bunch as a function of the RF phase.

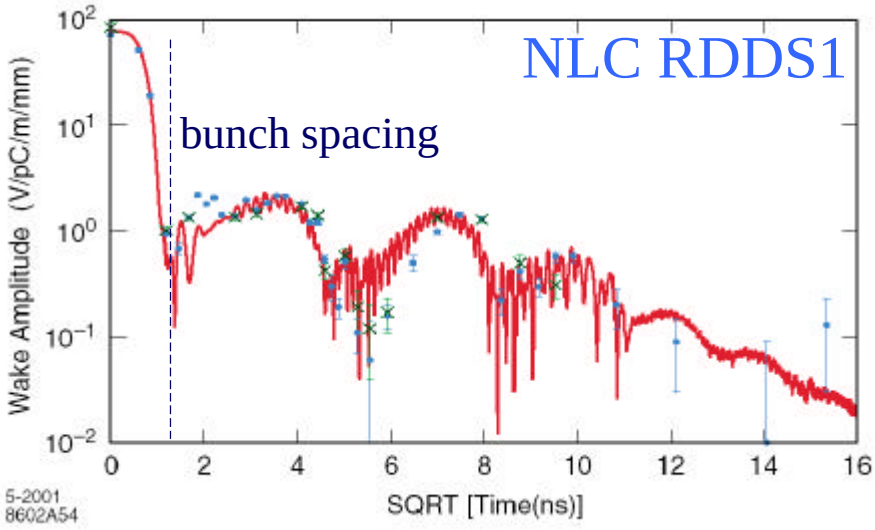


Figure 11: Effects of detuning and damping on the long-range transverse wakefields in an NLC structure.

For the multi-bunch case, the approach generally adopted is to engineer the problem away by sufficiently damping the modes generated by the passage of one bunch before the next bunch arrives. This is done in two steps: the fast ‘damping’ is achieved by randomly detuning neighbouring cells within a structure. The modes generated by these cells destructively interfere with each other and cause a rapid reduction in amplitude over the short time between bunches. However, this is not strictly damping, as the energy of the modes is still present, and due to the finite tune spread of the detuned structures, the modes will at some later time re-cohere. So on a

longer time scale, 'true' damping (so-called higher-order mode or HOM dampers) is used to extract the energy of the modes from the structures. Figure 11 shows both the measured and calculated long-range wake amplitude for an NLC damped detuned structure. The cell-to-cell detuning causes the more than two orders of magnitude decrease in amplitude with the first bunch spacing (1.4 ns).

On the time scale of a single bunch, it is not possible to damp the modes mechanically, and generally the single-bunch wakefields are dealt with by better structure to beam alignment. One important instability – that driven by a coherent betatron oscillation along the linac – can be effectively controlled by the use of Balakin, Novokhatsky and Smirnov (BNS) damping. Consider a bunch performing a coherent betatron oscillation along the linac. When the bunch is at its maximum displacement (a), the wakefield generated by the head kicks the tail of the bunch. If we simplify things by thinking of two particles separated by $2\sigma_z$, each with a charge $Q/2$, then the trailing (tail) particle will receive a kick from the head particle q, which, $\pi/2$ in phase downstream, will lead to a finite displacement $\approx bq$ (where the head particle now has zero displacement). A further $\pi/2$ in phase and both head and tail particle now have a displacement $-a$, and the head will now give a $-q$ kick to the tail. However, the original kick has now changed sign due to the π phase advance, and so the kicks add coherently. This resonant behaviour will continue down the linac, driving the tail particle to high and higher amplitudes.

The concept of BNS damping is to balance the effective de-focusing of the tail due to the wakefield, by increasing the focusing from the lattice. This is effectively achieved by decreasing the energy of the tail with respect to the head. In other words, a longitudinally correlated energy spread is introduced into the bunch, so that the wakefield kicks are balanced by the natural chromaticity of the linac FODO lattice. As a result the bunch oscillates as a rigid body down the linac, and there is no (or little) emittance growth.

For our two particle model, the energy difference between head and tail particle is given by

$$\Delta E = \frac{1}{8}\frac{W_\perp(2\sigma_z)QL_{\text{cell}}^2}{\sin^2(\pi n_b)} \tag{33}$$

where L_{cell} is the FODO cell length, and n_b is the fractional betatron tune shift per cell ($=\Delta\mu_{\text{cell}}/2\pi$). ΔE can be generated by a judicious choice of RF phase in the linac (see section above on single-bunch beam loading).

We should note that $W_\perp(z)$ is generally a non-linear function, and that in practise, it is difficult to achieve the required energy difference given by (33) for each location in the bunch (this condition, if met, is referred to as auto-phasing). However, in practise it can be achieved to a very good approximation. The cost of implementing BNS damping is in the required RF overhead needed to run off phase in the initial sections of the linac, and the higher energy spread in both the downstream parts of the linac

and the beam delivery system, ultimately leading to tighter alignment tolerances on the quadrupoles.

BNS damping addresses the problem of single-bunch beam break-up due to a coherent oscillation in the linac, but does not address the issue of random cavity or structure alignment. The additional emittance growth due to random structure transverse alignment errors with an RMS of ΔY_{RMS} is approximated by

$$\Delta e \approx dY_{\mathrm{RMS}}^2 \left[pe_0 N r_e W_\perp(2s_z) \right]^2 \frac{L_{\mathrm{acc}} \bar{b}_i}{2aG} \left[\left(\frac{E_f}{E_i} \right)^a - 1 \right] \tag{34}$$

where E_i and E_f are the initial and final beam energies, a is the scaling of the focusing lattice (typically ~0.5), G is the accelerator gradient, L_{acc} is the length of the structures, $\bar{b}_i$ is the initial average beta function, N is the number of particles per bunch, $W_\perp(2s_z)$ is the wake potential at twice the bunch length (s_z), and e_0 and r_e are the permittivity of free space and the classical electron radius respectively. From (34) we can immediately see that for a given Δe, the RMS structure alignment tolerance scales roughly as

$$\begin{aligned} dY_{\mathrm{RMS}} &\propto \frac{1}{NW_\perp} \sqrt{\frac{G}{b}} \\ &\propto \frac{f^{-3}}{N} \sqrt{\frac{G}{b}} \end{aligned} \tag{35}$$

So going to high-frequencies has a significant impact on the required alignment tolerances for the structures. These tighter tolerances are then offset by having a higher gradient (G), stronger overall focusing (and hence smaller b-functions), and smaller charge per bunch (Ne). Nevertheless, the tight tolerances cannot be met using standard mechanical survey and alignment techniques, and beam-based alignment of the structures is generally required. For both JLC-NLC and CLIC the structures are placed on girders than can be remotely translated to micron precision. Each structure will be equipped with an output coupler which allows the transverse dipole modes to be measured (the coupler acts in essence as a structure beam position monitor). The girders can be moved to effectively zero out the dipole mode. The procedure must be repeated for the ~2000 structure girders in the linacs.

Sources

Electron sources

In the first part of the lecture, we derived the scaling laws for luminosity for a linear collider. We saw that high beam powers where generally required to achieve the ambitious luminosity goals. From considerations of linac acceleration efficiency and wakefield control, we have seen that we ideally need to:

- accelerate large numbers of bunches in a single bunch train to achieve a high RF to beam power transfer efficiency;

- reduce the charge per bunch to mitigate the effects of the strong transverse wakefields.

Hence the sources must provide the long bunch trains. In addition, polarisation is mandatory for the electron source.

All LC designs propose to use a laser-driven photo-injector to provide the necessary time structure and charge per bunch. Figure 12 schematically shows the concept.

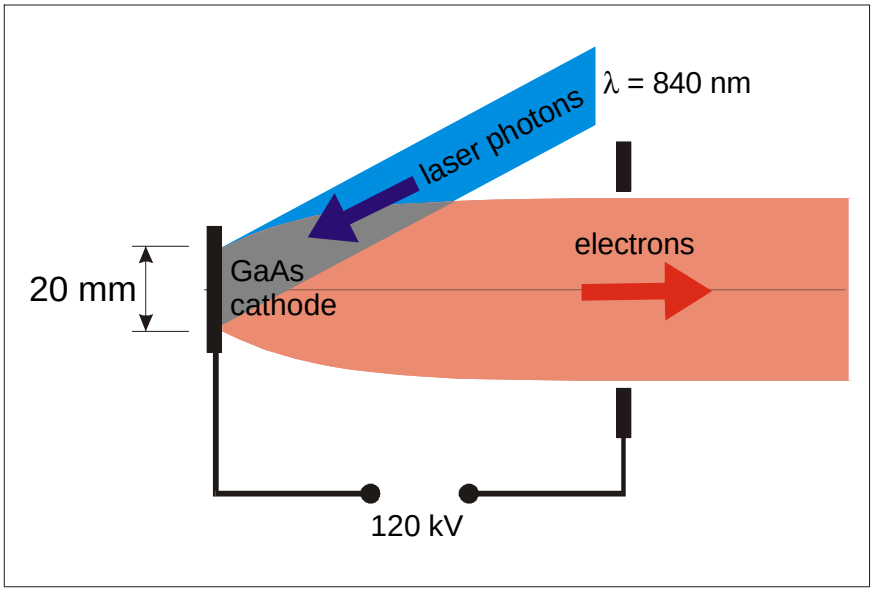


Figure 12: Concept of the laser-driven photo-injector for polarised electron production.

The emittance produced by such a gun is dominated by space-charge effects, and is typically of the order of 10^{-5} m (a factor of 10 too big in the horizontal plane, and a factor of 500 too big in the vertical!). To produce polarised electrons, a GaAs cathode is used together with a laser light of 840 nm wavelength: such sources are capable of producing over 90% polarisation. Unfortunately, the GaAs cathodes are extremely sensitive and the guns require very high vacuum (better than 10^{-11} mbar); this rules out high-brightness RF guns that produces orders of magnitude better emittance, because the typical vacuum associated with such guns is rather bad (at best 10^{-7} mbar). Development of high-brightness polarised RF guns with high vacuum would be very attractive for LC applications.

Because the gun is effectively DC, the polarised source requires a bunching section downstream of the gun as shown in figure 12. The long DC beam is first bunched using a sub-harmonic bunching section (RF), before finally being accelerated up to a ~GeV before injection into the damping rings.

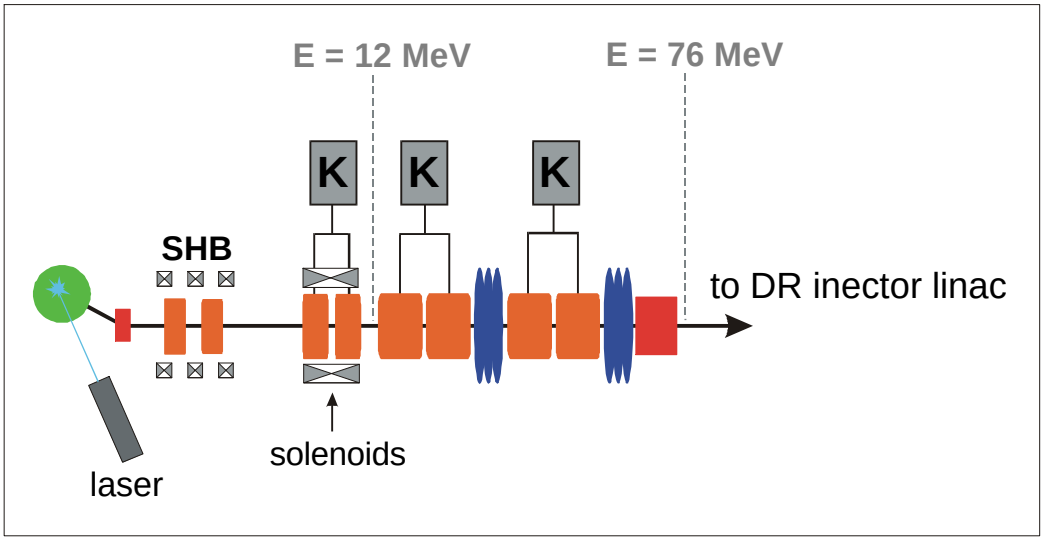


Figure 13: typical bunching and pre-acceleration section for the polarised electron source.

Positron Source

Unlike electrons, positrons must first be created in a high-energy particle reaction. The basic mechanism is pair-production: high energy photons (gammas) are converted in a target into electron-positron pairs. The positrons are collected, focused and accelerated, while the electrons are dumped. The key issue for the source is how the gamma photons are created. For the current LC designs, two approaches to the problem are foreseen:

- A *conventional* source (Figure 14), where high-energy electrons (>GeV) are allowed to strike a thick target (~4 radiation lengths). The primary electrons generate high-energy photons via bremstrahlung, and these then convert within the same target to the required electron-positron pairs.
- An undulator based source (Figure 15), where very high energy electrons (>150 GeV) are first passed through an undulator or wiggler magnet to produce the required high-energy photons. These photons are then converted in a thin target (~0.4 rad. lengths) into electron-positron pairs.

The conventional source requires a target of ~4 radiation lengths primarily to generate the photon shower (the bulk of the pair production is at the back of the target). Conversely, the undulator source requires a relatively thin target (0.4 rad. lengths) to act as converter for the photons; this has several advantages:

- much less energy (average power) deposition in the target itself (~5 kW as opposed to 22 kW for a conventional source);
- the emittance (both transverse and longitudinal) of the produces positrons are smaller due to less Coulomb scattering in the target (roughly a factor of two over a conventional source).

To produce the required positron charge, the JLC-NLC machine proposes using three targets running in parallel (driven by the same 6 GeV linac). RF deflectors will be used to send alternate bunches to the targets, and recombine the resulting positron bunches into a single train after the target and capture sections.

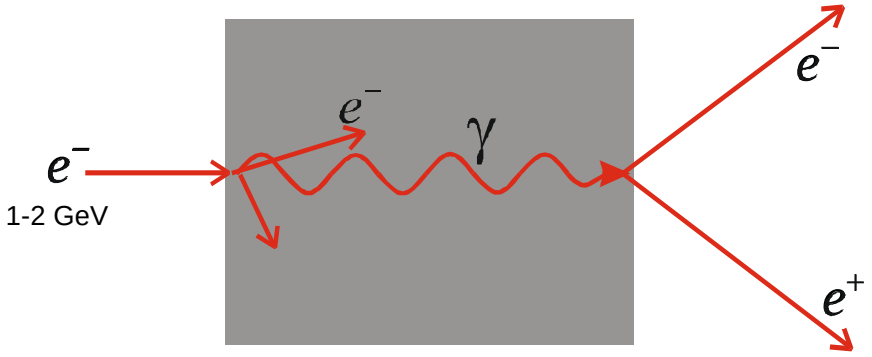


Figure 14: Conventional positron source, using a thick target. Incident electron energies are typically 2-6 GeV.

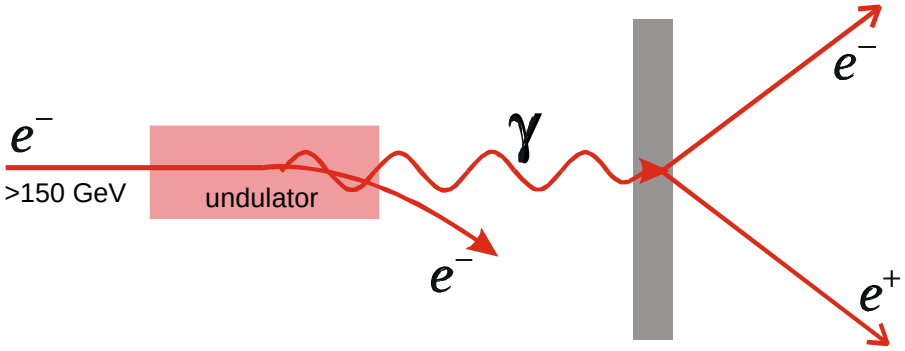


Figure 15: undulator based source.

The parallel target is necessary to reduce the average power deposition in the thick targets to below failure thresholds. The undulator source requires only a single target, due to the much reduced power deposition in the material itself.

Unfortunately the undulator source has several disadvantages:

- Generating high energy photons (gammas) using an undulator requires a high beam energy (>120 GeV). Practically the problem is solved by using the primary (luminosity) electron beam. After the undulator, the high-energy electron beam is steered around the target and either transported directly to the IP or further accelerated (depending on the location of the positron source). This introduces a fundamental coupling between the electron and positron linacs, the former being required before the later can operate fully. This coupling has potential impact for commissioning and boot-strapping the machine.
- Unlike the conventional source, which is very like that used at the SLC, the undulator source has never been realised, and due to high electron beam energy required, is unlikely to be tested before the LC is built.

One last advantage of the undulator source is the possibility of producing polarised positrons. By replacing the planar undulator with a helical undulator, whose magnetic field rotates azimuthally along the axis of the undulator with a given period. The field causes the generated photons to be longitudinally polarised, a characteristic that is

past to the electron-positron pairs. There are several R&D challenges for the polarised source, not least the construction of the required very high field helical undulator.

Damping Rings

The emittances produced by both the electron gun and the positron source are too large by several orders of magnitude. While there is at least the conceptual possibility of producing a polarised electron source based on an RF gun that just might provide the required emittance, the nature of positron production will always insure the need for at least one damping ring.

A damping ring is a storage ring in which the beams are stored for a specified time (typically 20-200 ms) before being ejected and accelerated in the main linac. During that time, the synchrotron radiation 'damps' both the longitudinal and transverse emittances. The damping behaviour can be summarised by the following equation:

$$e_f \approx e_i \exp\left(-\frac{2T}{t_D}\right) + e_{eq}\left[1 - \exp\left(-\frac{2T}{t_D}\right)\right], \tag{36}$$

where the e_i, e_{eq}, and e_f are the initial (injected), equilibrium and final (ejected) emittances, T is the storage time and t_D is the damping time, given by

$$t_D = \frac{2E}{P_g}, \tag{37}$$

where E is the damping ring energy, and P_g is the average radiated power per electron:

$$P_g \approx \frac{cC_g}{2p}\frac{E^4}{r^2}; \quad C_g \approx 8.85\times10^{-5}\,\mathrm{GeV^{-3}m}\,. \tag{38}$$

Equation (38) states that the damping time is just twice the time required for an electron to radiate all its (initial) energy.

The injected positron beam has a normalised transverse emittance of typically 0.01 m, which must be damped to ~2×10^{-8} m: a reduction by a factor of 5×10^5! From equation (36) this corresponds to about 7-8 damping times (providing the equilibrium emittance is small enough – see later).

The store time T depends on the repetition rate of the machine (f_{rep}), and the number of bunch trains stored at any one time (N_s):

$$T = \frac{N_s}{f_{rep}} \tag{39}$$

The ring must have a large enough circumference to hold the bunch trains:

$$C = cN_s n_b t_b \tag{40}$$

where n_b are the number of bunches in a bunch train, and t_b is the bunch spacing (c is the velocity of light). For JLC-NLC we have N_s = 3, $n_b t_b$ = 267 ns, corresponding to $C \approx 240\text{m}$. Assuming the ring is approximately circular, that gives us a radius of ~38 m. The ring energy is 1.98 GeV. From equations (37) and (38), we find that

$$P_g \approx 45\,\text{GeV/s}$$

and

$$t_D \approx 88\text{ms}.$$

With f_{rep} = 120 Hz, we have from (39) T = 25 ms, about a third of our damping time. We require a damping time which is roughly 25/8 ~3 ms. Clearly it is necessary to significantly decrease the damping time (by a factor of ~30!).

We could accomplish this by increasing the ring energy, since from (37) and (38) we have that $t_D \propto E^{-3}$. However, we must also consider that:

- the equilibrium emittance $e_{eq} \propto E^2 / r$; and
- the required RF power $P_{RF} \propto P_g \propto E^4$

A second (and preferred) way to increase P_g is by the addition of strong wigglers in straight sections in the ring. The average power radiated per electron is now

$$P_g = c\frac{\Delta E_{wiggler} + \Delta E_{arcs}}{L_{wiggler} + 2\pi r_{arcs}} \tag{41}$$

where $\Delta E_{wiggler}$ and ΔE_{arcs} are the energy lost per electron per turn in the wigglers and arcs respectively, and $L_{wiggler}$ is the total length of wiggler (straight sections) in the ring. Clearly $L_{wiggler} + 2\pi r_{arcs} \geq cN_s n_b t_b$ in order to contain the required bunch trains. The energy radiated by a wiggler is given by

$$\Delta E_{wiggler} \approx \frac{K_g}{2\pi} E^2 B^2 L_{wiggler}; \quad K_g \approx 8\times 10^{-6}\,\text{GeV}^{-1}\,\text{Tesla}^{-2}\text{m}^{-1} \tag{42}$$

where B^2 is the average value of the field squared over the total length of wiggler.

Example: using the previous ring parameters, insert a total of 50 m of wiggler, and estimate the required wiggler field to achieve the required factor of 30 decrease in damping time.

To decrease the damping time by a factor of 30, we must increase P_g to 45×30 = 1350 GeV/s.

The energy loss per turn $= \dfrac{240+50}{c} P_g \approx 1.3\times10^{-3}\,\text{GeV}$

The energy lost in the arcs (r = 38 m) $= 45\times\dfrac{240}{c} \approx 3.6\times10^{-5}\,\text{GeV}$

Hence the energy lost in the wigglers per turn is

$$\Delta E_{wiggler} = 1.3\times10^{-3} - 3.6\times10^{-5} \approx 1.3\times10^{-3}$$

(note that the wiggler completely dominates the damping time). From (42) we have

$$B^2 \approx \frac{2p\times(1.3\times10^{-3})}{(8\times10^{-6})\times(1.98)^2\times50} \approx 5\ \text{Tesla}^2$$
$$B \approx 2.3\text{Tesla}$$

Limits on Performance

The equilibrium emittance in the horizontal plane is fundamentally defined by the steady state, where the rate of emittance change due to the radiation damping is exactly balanced by the increase in emittance due to the quantum 'noise'. For the quantum excitations (so-called anti-damping), the growth rate is to a good approximation constant, giving rise to a linear increase in emittance with time:

$$\frac{de_x}{dt} = q_x \tag{43}$$

where q_x is a constant which is only a function of the lattice optics and the energy. We should note here that q depends explicitly on the random nature of quantum fluctuations of the emitted synchrotron radiation:

$$q_x \approx \frac{\left\langle N\left\langle u^2\right\rangle \mathsf{H}\ \right\rangle_s}{E^2} \tag{44}$$

where N is the average photon emission rate, $\left\langle u^2\right\rangle$ is the variance of the photon energy, and H is given by

$$H\quad(s) = \frac{1}{b_x(s)}\left\{D_x^2(s) + \left[b_x(s)D_{x'}(s) + a_x(s)D_x(s)\right]\right\} \tag{45}$$

which is a function of the lattice and changes as we go around the ring (with s), as do N and $\left\langle u^2\right\rangle$ (both these values depend on the local magnetic field, which clearly varies with s). The subscript s on the outer angle bracket in (44) indicates an average over the complete ring is to be taken.

As we have already mentioned, the radiation damping depends only on the average radiated power per electron and the energy (equation (37)):

$$t_D = \frac{2E}{P_g} \tag{46}$$

We should note that there are no quantum effects involved here: damping is a purely classical mechanism which depends only on average loss rates. The (horizontal) equilibrium emittance is given when $de/dt = 0$, i.e.

$$\begin{aligned} \frac{de_x}{dt} &= Q - \frac{2}{t_D} e_x = 0 \\ e_{x,eq} &= \frac{1}{2} Q t_D \end{aligned} \tag{47}$$

The horizontal equilibrium emittance depends on the presence of horizontal dispersion via equation (45). In the vertical plane, we still have the same damping mechanism, since this does not depend on quantum fluctuations but only on the average power loss. However, does the absence of vertical dispersion (at least by design) mean that $q_y = 0$, and that the vertical emittance simply damps away to zero?

Not quite. Even in the absence of dispersion there is still a fundamental quantum effect that will generate emittance. When a photon is emitted with a given momentum $\hbar w / c$, the electron must necessarily recoil to conserve this momentum. The basic theory that leads to the classical damping rates and the horizontal excitation assume that the photon is emitted along the direction of the electron, and there is no change in the electron angle. However, the photons are emitted in a cone with a typical angle of g^{-1} (~250 μr for 2 GeV). These random angle 'kicks' add emittance to the beam. For most modern storage rings (light sources) this effect is negligible, but the proposed damping rings are pushing (dangerously) close to this limit.

In practise, there are many other effects (including so-called collective effects) that limit the minimum achievable vertical emittance:

- Intrabeam scattering, which begins to become important as the charge density increases during damping.
- Instabilities, such as fast-ion or electron cloud. The former places constraints on the vacuum system, while the latter is sensitive to the surface properties of the vacuum chamber (secondary emission coefficient).
- magnet misalignments, which cause close orbit deviations leading to cross-plane coupling and spurious vertical dispersion, all of which will cause serious degradation to the vertical emittance if not corrected.

All the above points are important effects, but the last one is probably the first obstacle to be tackled. The typical alignment tolerances are on the order of 10-20 μm, well beyond what is possible to achieve with traditional mechanical survey methods. Beam based alignment is therefore mandatory for the damping rings. Much of the

'proof of principle' for a damping ring design is based on having orbit correction algorithms which – given a set of realistic installation alignment errors – allow the ring to achieve its design vertical emittance.

Bunch Compression

The choice of the damping ring lattice and energy defines the equilibrium beam phase space, i.e. the transverse emittances (*x* and *y*) and the longitudinal emittance. The longitudinal emittance is the product of the bunch length and the energy spread in the beam, both of which are fixed by the damping ring design. The bunch length produced by such rings is typically a few millimetres. We have seen in the section on luminosity scaling that the we need to make the bunch lengths short, typical a ~100 μm. Thus we need to compress the bunch longitudinally by a factor of ~40 or so. We do this in a special bunch compression section, located after the damping ring and before the main linac (note that it is important in the main linac to have $s_z \ll l_{RF}$ to keep the bunch energy spread small).

Compressing a relativistic beam is achieved by first introducing an energy 'chirp' along the bunch using an RF section phased at the zero crossing. The beam is then transported through a dispersive (non-isochronous) lattice; the relative path length differences for the low and high energy parts of the bunch cause the bunch to compress longitudinally, or more precisely, to rotate in longitudinal phase space. Figure 16 outlines the concept.

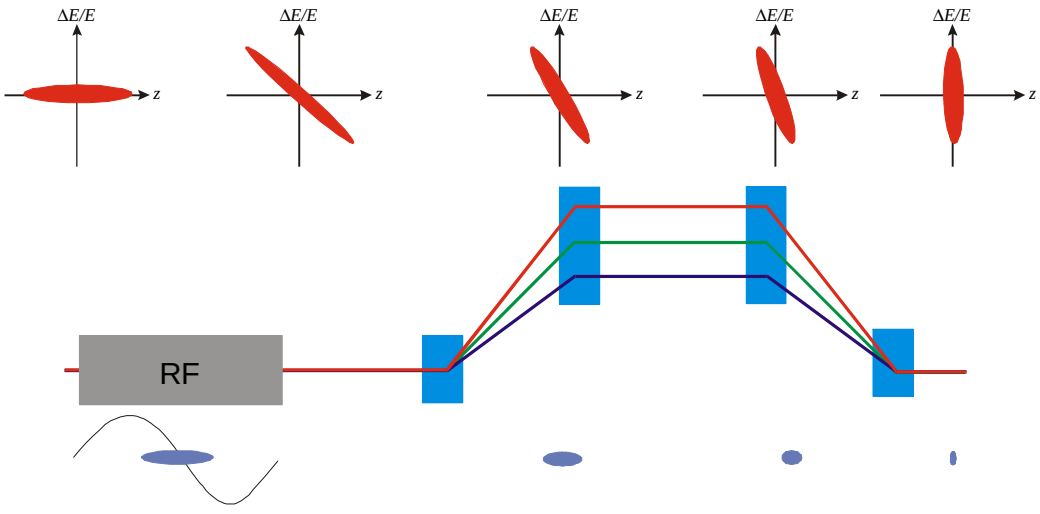


Figure 16: Bunch compression by a p/2 longitudinal phase space rotation.

By Louville's theorem, the longitudinal phase space area must be conserved. We can use this fact to ascertain some basic parameters for the (linear) compressor.

We begin with the bunch from the damping ring which has an *uncorrelated* relative energy (momentum) spread d_u and an initial bunch length $s_{z,0}$. We will define the required compression ratio $r_c \equiv s_{z,0} / s_z > 1$, where s_z is the final required bunch length. Since the product $s_z d$ is conserved, the final energy spread must also be increased r_c:

$$d = r_c d_u \tag{48}$$

The final energy spread is given by

$$d^2 = d_u^2 + d_c^2 \tag{49}$$

where d_c is the *correlated* relative energy spread introduced by the RF:

$$d_c \approx \frac{k_{RF} V_{RF} s_{z,0}}{E} \tag{50}$$

where $k_{RF} = 2p / l_{RF}$, V_{RF} is the total RF (peak) voltage, and *E* is the beam energy. We have assumed that $s_{z,0} \ll l_{RF}$, and have taken the linear slope of the RF at the zero crossing (note that implicit in this assumption is no average acceleration of the beam, and thus *E* stays constant).

Combining (48), (49) and (50) we arrive at an expression for the required RF voltage:

$$V_{RF} \approx \frac{E}{k_{RF}} \left(\frac{d_u}{s_{z,0}} \right) \sqrt{r_c^2 - 1} \approx \frac{E}{k_{RF}} \left(\frac{d_u}{s_{z,0}} \right) r_c \tag{51}$$

The quantity in parenthesis ($d_u / s_{z,0}$) is fixed by the damping ring. To reduce the required voltage for a given compression ratio, we see that we can either reduce the energy at which we perform the compression, or decrease the wavelength of the RF, although if we decrease the wavelength too much, the linear approximation implicit in (50) will become invalid. The *E* in the numerator of equation (51) is one reason while bunch compression is (in general) not performed at high energies.

For the dispersion (non-isochronous) section, we take the linear part only, which is usually written as

$$z_1 = z_0 + R_{56} d_i \tag{52}$$

where z_0, z_1 are the initial and final particle longitudinal positions, d_i is the particle relative energy (momentum) deviation, and R_{56} is the linear *longitudinal dispersion.* The minimum bunch length is achieved when the longitudinal phase space ellipse is upright (see Figure 16), i.e. when the correlation $\langle zd \rangle = 0$. From equation (52) we have

$$\langle zd \rangle_{final} = \langle zd \rangle_{initial} + R_{56} d^2 = 0 \tag{53}$$

The initial correlation is given by the RF:

$$\langle zd \rangle_{initial} = \frac{k_{RF} V_{RF}}{E} s_{z,0}^2 = d_c s_{z,0} \tag{54}$$

Remembering that $d^2 = d_u^2 + d_c^2$ (equation (49)), we arrive at an expression for R_{56}:

$$R_{56} = -\frac{d_c s_{z,0}}{d_c^2 + d_u^2} \tag{55}$$

For a high compression ratio, $d_c \gg d_u$ and we can approximate (55) as

$$R_{56} \approx \frac{s_{z,0}}{d_c} \tag{56}$$

Final Focusing

Once the bunch has been damped, compressed longitudinally and then accelerated, we must focus the beams and collide them at the IP. Remembering the luminosity scaling arguments, we maximised the luminosity by reducing the vertical beam size to the point where $\beta_y^* \approx s_z$. Hence typical IP vertical β-functions are in the range 0.1-0.4 mm. At the exit of the linac, the β-functions are much more characteristic of typical FODO lattices (~100 m): thus we must reduce (demagnify) the beam by a typical factor of $M = \sqrt{\beta_{linac} / \beta_y^*} \approx 300$. The simplest way to achieve this is by use of a telescope structure with point-to-point focusing. Lets briefly consider only the vertical plane for simplicity. From Figure 17, we can immediately write down that $M = f_1 / f_2$. If we take the focal length of the final lens $f_2 = 2\,\mathrm{m}$, then we immediately see that we typically need $f_1 \approx 600\,\mathrm{m}$, which roughly sets the length scale for the final focus system. In reality, the final focus optics is not strictly a simple telescope as above, and the same demagnification factors can generally be achieved in shorter systems. In addition, the simply 'thin lenses' in the above figure must focus in both planes, and the final 'lens' is generally formed from a quadrupole doublet (referred to as the final doublet, or FD).

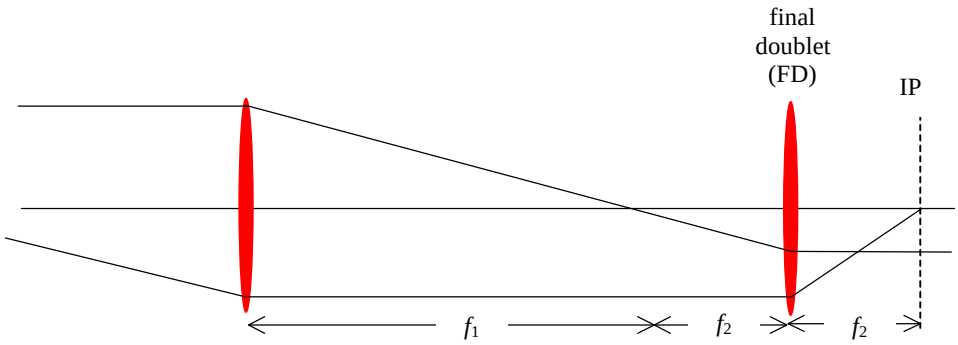


Figure 17: Telescope optics for a final focus system.

Much of the design of the complete system is constrained by the choice of focal length of the FD (f_2), which is often referred to as L^* in the literature. The FD requires magnets with very high quadrupole gradients; typically hundreds of Tesla per meter. To achieve these types of gradient generally requires superconducting or permanent magnet technology, both of which are under consideration.

Such a strong lens suffers from a high degree of chromatic aberration. Particles with different energies will be focused to different points. Consider the simplified case depicted in Figure 18. The shift in focal point is $\Delta f \approx L^* d$. Assuming that the associated change in IP angle q is negligible, the off energy particle will then have a finite displacement at the IP given by

$$\Delta y^* \approx \Delta f q = L^* q d \tag{57}$$

To estimate the impact of this aberration on the RMS vertical beam size, we must calculate the RMS aberration $\Delta y^*_{\mathrm{RMS}} = \sqrt{\left\langle \Delta y^{*2} \right\rangle}$. Assuming that there is no initial correlation between energy and angle, we have

$$\Delta y^*_{\mathrm{RMS}} \approx L^* q_{\mathrm{RMS}} d_{\mathrm{RMS}} . \tag{58}$$

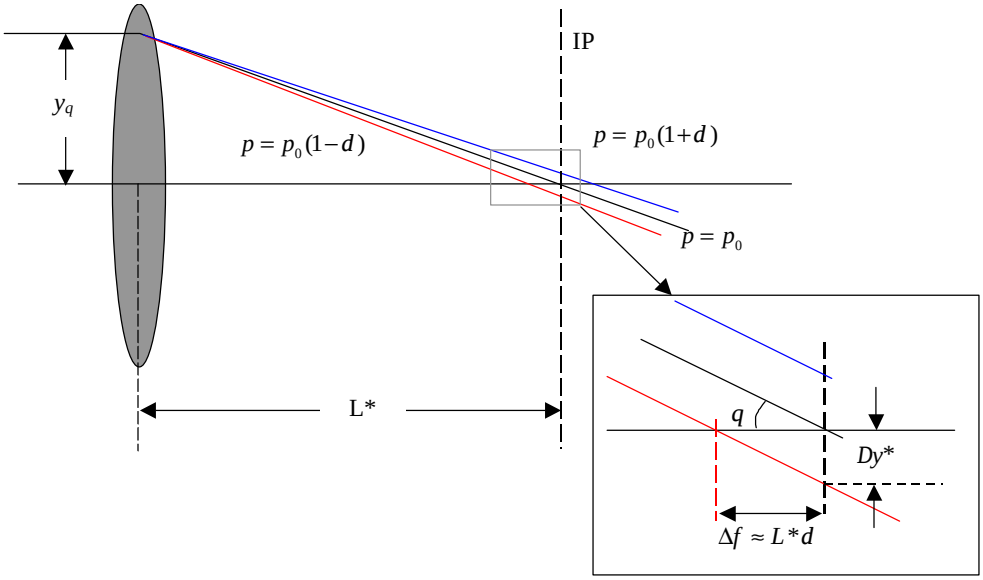


Figure 18: Chromatic aberration from the strong final doublet.

Dividing equation (58) by the nominal (linear) beam size s_y^*, we have

$$\frac{\Delta y^*_{\mathrm{RMS}}}{s_y^*} \approx L^* \left(\frac{q_{\mathrm{RMS}}}{s_y^*} \right) d_{\mathrm{RMS}} . \tag{59}$$

Using the standard relations $q_{\mathrm{RMS}} = \sqrt{e/b}$ and $s = \sqrt{eb}$, we finally arrive at

$$\frac{\Delta y^*_{\mathrm{RMS}}}{s_y^*} \approx \frac{L^*}{b_y^*} d_{\mathrm{RMS}} \tag{60}$$

With L^*= 2 m, $b_y^* = 0.1\mathrm{mm}$, and $d_{\mathrm{RMS}} = 0.3\%$, we arrive at

$$\frac{\Delta y^*_{\mathrm{RMS}}}{s_y^*} \approx 60$$

Thus, if left uncorrected, the chromatic 'aberration' of the FD would completely dominate the IP vertical beam size, increasing it by a factor of ~60! Note that the effect is only dependent on L^* and b_y^* (for a fixed relative energy spread). We have already said that we'd like to make $b_y^* \approx s_z$ to maximise the luminosity, which leaves only reducing L^*. Unfortunately this is not practical since (i) the magnet technology becomes intractable, and (ii) some space must be left for the physics detector (considerations of which general try to force the FD further away from the IP, thus increasing L^*). There are potential exotic solutions for effectively making L^* very short : plasma lenses at the IP, or the so-called dynamic focusing schemes, which utilise a second (low-energy) bunch as a lens just in front of the IP. These schemes are

considered beyond the next generation 0.5-1 TeV machines, but may prove attractive for future multi-TeV machines, somewhere in the distant future.

The current solution is to optically correct the FD chromaticity using strong sextupole magnets in a dispersive region of lattice.

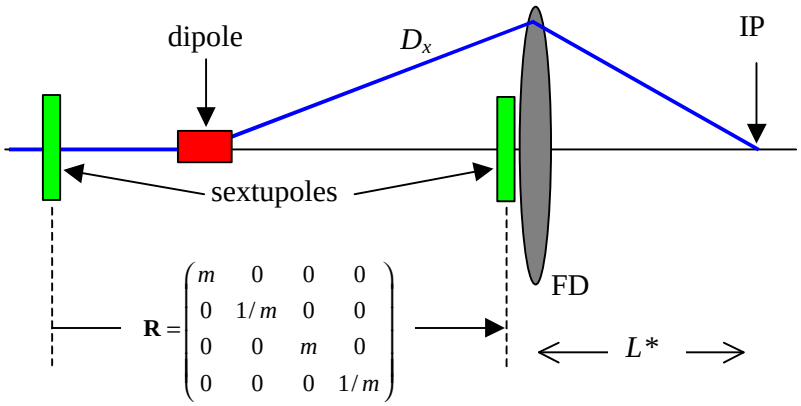


Figure 19: Concept for a local chromatic correction scheme for the FD.

Figure 19 shows the concept for the FD chromatic correction currently used in all LC designs. Horizontal dispersion is generated at the FD location by weak dipoles judiciously placed to cause the dispersion to be zero at the IP. A sextupole (or sextupoles) is then placed adjacent to the FD. In the presence of horizontal dispersion, the non-linear kicks from a thin-lens sextupole of integrated strength K_s are given by

$$\begin{aligned} \Delta x' &= -\frac{1}{2}K_s\left(x^2 - y^2\right) - D_x K_s x d - \frac{1}{2}D_x^2 d^2 \\ \Delta y' &= +K_s xy + D_x K_s y d \end{aligned} \tag{61}$$

The second term in both expressions (xd and yd) are the first-order chromatic kicks we will use to cancel the similar kicks from the FD. Again for a thin-lens FD:

$$\begin{aligned} \Delta x' &= \frac{xd}{L^*} + \frac{D_x d^2}{L^*} \\ \Delta y' &= -\frac{yd}{L^*} \end{aligned} \tag{62}$$

Adding the kick expressions given in (61) and (62) together, and choosing the values of K_s and D_x such that

$$\frac{1}{L^*} - D_x K_s = 0\,, \tag{63}$$

the first-order chromatic kicks vanish as required. The sextupole-FD combination is now chromatically corrected to first-order in d. Unfortunately, there are still residual non-linear terms which will cause significant aberrations if left uncorrected:

$$\Delta x' = -\frac{1}{2D_x L^*}\left(x^2 - y^2\right) + \frac{1}{2}\frac{D_x}{L^*}d^2$$
$$\Delta y' = +\frac{1}{D_x L^*}xy \qquad (64)$$

The pure geometric (*d*-independent) terms can be cancelled by placing one or more sextupoles upstream at the same phase as the FD (as shown conceptually in Figure 19). The linear map should have the form shown in Figure 19, i.e. the magnification in both *x* and *y* planes should be the same (in this simplified example, *m*). Then the strength of the upstream compensating sextupole is $K_{s,comp} = K_s / m = 1/(mL^* D_x)$. The non-linear dispersion in the horizontal plane (d^2 term) can be cancelled either by arranging for a finite dispersion function at some upstream quadrupoles at the same phase as the doublet, or allowing a small dispersion at a sextupole (or both).

In a real system, four to five sextupoles are generally used to balance (zero) all the second-order terms in both planes; the system is therefore a second-order achromat. The trick is to achieve this and simultaneously minimise the third- and higher-order aberrations which arise from interactions between the various non-linear terms; this normally requires considerable experience on the part of the designer.

Fundamental limits: The Oide Effect.

We can reasonably ask the question if there are any fundamental limits to the minimum achievable vertical beam size at the IP from a given system (notwithstanding the constraint that $b_y^* \geq s_z$; we can in principle always compress the bunch further). K. Oide derived one such limit based on synchrotron radiation emitted in the FD. As high amplitude particles travel through FD, they emit synchrotron radiation photons, loosing some energy in the process. The particle now travel on a slightly different trajectory due to the high chromaticity of the FD, and are focused to a different point on the axis (this is completely analogous to the above discussion on the FD chromatic aberration, except that here the momentum error is 'born' in the FD itself). The random quantum fluctuations in the doublet cause an increase in the RMS vertical beam height at the IP. As b_y^* is reduced, the corresponding b at the doublet increases; the particles then (on average) see higher magnetic fields, and the effect of synchrotron radiation also increases. Thus there should be a minimum value of b_y^* corresponding to a minimum s_y^*, below which, the effect of the radiation begins to dominate and increase the vertical IP beam size.

Oide showed that the minimum beam size and corresponding b-function is given by

$$s_{y,\min}^* \approx 1.83\left(r_e \lambda\!\!\!^{-}_e F\right)^{1/7} e_{y,n}^{5/7}$$
$$b_{y,\min}^* \approx 2.39\left(r_e \lambda\!\!\!^{-}_e F\right)^{2/7} e_{y,n}^{3/7} \qquad (65)$$

where *F* is a number dependent on the details of the FD focusing scheme; $e_{y,n}$ is the normalised vertical emittance, and $r_e, \lambda\!\!\!^{-}_e$ have their usual meanings. *F* has a typical

value of ~7, with a minimum of ~0.1. The rather weak 1/7-power dependence on F however makes the minimum beam size fairly insensitive to the actual details of the FD design.

One remarkable observation from (65) is that both expressions are independent on beam energy. This is counter intuitive (and surprising), as most synchrotron radiation effects scale as very high powers of E (the emittance growth from a dipole magnet scales as E^6!).

Stability and Feedback

The tiny vertical emittances and nanometer IP beam sizes place extremely tight tolerances on magnet alignment. Generally speaking, the tolerances required to achieve the design luminosities cannot be met using state-of-the-art survey and installation techniques, and heavy use of beam-based alignment is required.

Having successfully beam-base aligned our linear collider and tuned up our luminosity, the environment will then attempt to ruin our good work. Particularly, ground motion (vibration) will attempt to move the accelerator components away from where we put them, degrading the luminosity as it does so.

We generally divide the effects of component motion into two regimes: a shift in the orbit causing the nanometer beams to move out of collision at the IP; and a degradation of the beam quality (beam size) due to spurious dispersion, cross-plane coupling or wakefields. Normally the former is on times scales much faster than the latter.

To get a feel for the magnitude of the effects, lets consider the effect of moving a quadrupole vertically. The offset quadrupole gives a coherent kick to the beam, causing a downstream betatron oscillation. The resulting offset at the IP is given by

$$\Delta y^* = K_Q y \sqrt{b_Q b_y^*} \sin(\Delta f), \tag{66}$$

where K_Q is the integrated quadrupole strength, y is the quadrupole offset, b_Q and b_y^* are the vertical b-functions at the quadrupole and IP respectively, and Δf is phase of the quadrupole with respect to the IP. By dividing (66) by the nominal beam size at the IP $s_y^* = \sqrt{b_y^* e_{y,n}/g}$, we find:

$$\frac{\Delta y^*}{s_y^*} = K_Q y \sqrt{\frac{b_Q g}{e_{y,n}}} \sin(\Delta f) \tag{67}$$

We require that the LHS of (67) be less than one (typically 0.3). In the beam delivery system, the b-functions range from hundreds of meters to thousands of kilometres at the special case of the FD. Typical tolerance numbers for y range from 10 to 100 nm.

Lets now take a more qualitative look at the random vibration of the linac quadrupoles. The final motion at the IP is now the sum of the effects from all the quads:

$$\frac{\Delta y^*}{s_y^*} = \frac{1}{\sqrt{e_{y,n}}} \sum_{i=1}^{N_Q} K_{Q,i} y_i \sqrt{b_{Q,i} g_i} \sin(\Delta f_i) \tag{68}$$

We are interested in the RMS motion at the IP. Squaring (68) and taking the average, we find

$$\frac{\langle \Delta y^{*2} \rangle}{s_y^{*2}} = \frac{K_Q^2 s_Y^2}{e_{y,n}} \sum_i^{N_Q} b_{Q,i} g_i \sin^2(\Delta f_i) \tag{69}$$

where $s_Y^2 = \langle y_i^2 \rangle$ is the variance of the quadrupole vibration (assumed uncorrelated and equal for all quads). We have also assumed that quad strength K_Q is constant and taken it outside the summation. To make an estimate of the tolerance on the vibration amplitude, we approximate the summation of the quads by

$$\sum_i^{N_Q} b_{Q,i} g_i \sin^2(\Delta f_i) \approx \frac{N_Q \overline{bg}}{2} \tag{70}$$

We justify this approximation by assuming that change in energy (g) along the lattice is constant and linear, and that phase advance is smooth and $\sin^2(\Delta f)$ averages to ½. Hence (69) (after taken the square root) becomes

$$\frac{\Delta y^*_{\mathrm{RMS}}}{s_y^*} \approx \sqrt{\frac{N_Q \overline{bg}}{2e_{y,n}}} K_Q s_Y \le 0.3 \tag{71}$$

Now lets take some example numbers:

N_Q	400
$\overline{b}$	100 m
$\overline{E}$	125 GeV (corresponding to $\overline{g} = 2.4\times10^5$)
$e_{y,n}$	30 nm
K_Q	0.03 m

Substituting the above numbers into (71) leads to

$$s_Y \le 25\mathrm{nm} \tag{72}$$

An extreme case is the FD itself. Due to the parallel-to-point focusing arrangement of this ‘lens’, there is a one-to-one correspondence between the offset of FD and the

resulting offset at the IP. Hence the vibration tolerance on the FD is of the order of ~1 nm!

Although these nm tolerances look formidable, we can fortunately significantly relax them by the use of beam-based orbit correction feedback. There will almost certainly be many such feedback systems in any future LC. The most important one is the beam-beam feedback system at the IP which keeps the nanometre beams in collision.

The feedback makes use of the strong mutual beam-beam kick that a relative offset of the two beams at the IP produces. Figure 20 shows the basic set-up. Such a feedback system was successfully demonstrated at the SLC.

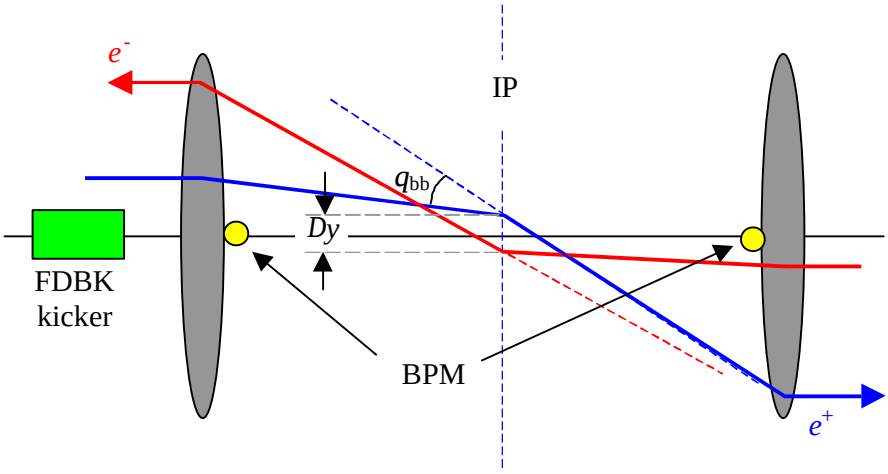


Figure 20: Concept of the beam-beam feedback system at the IP.

The two BPMs placed at the exit of the FDs are used to measure the strong beam-beam kick angle at the IP (q_{bb}) due to a relative offset of the beams (*Dy*). A kicker (corrector) magnet then adjust the offset of one beam to null (zero) the beam-beam kick, at which point the beams should be colliding head on.

When we discuss the use of feedback systems, we must immediately consider the frequency of the 'noise' that we are trying to correct. All feedback systems have a well defined frequency response (bandwidth). For an LC, we are in reality designing high-pass filters which attenuate beam motion below a certain cut-off frequency; above this frequency, the beam motion will in general be amplified. Figure 21 shows the frequency response for an idealised feedback system. The logarithmic frequency scale is indicated as a fraction of the sample rate (machine repetition rate) which ranges from 5 Hz for TESLA, to 200 Hz for CLIC (the high frequency colliders can operate at a much higher repetition rate than a superconducting machine). Several values of gain are plotted: gain can be thought of as the fraction of correction calculate for pulse *n* that is applied to pulse *n*+1. Hence gain=1 means the full 100% correction is applied on the next pulse. One immediate observation is that a higher gain will attenuate up to higher frequencies, but will at the same time significantly amplify those frequencies above the 'cut-off' (which I define here as the point where the response is unity). A lower gain has a lower cut-off, but has significantly less amplification at higher frequencies.

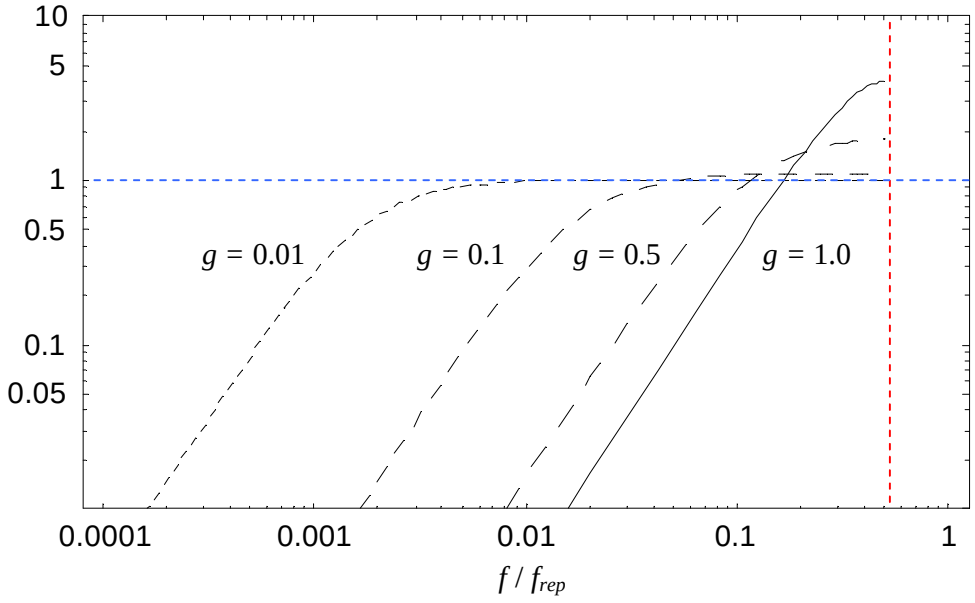


Figure 21: Frequency response curves for a simple one pulse delay feedback system. The curves represent different feedback gains. The frequency scale is relative to the machine repetition rate; the red dotted line represents the Nyquist frequency ($f_{rep}/2$).

The choice of feedback parameters (primarily gain) depends on the expected noise spectrum. Fortunately the amplitude of the ground motion vibration spectrum rapidly decreases with frequency ($\propto f^2$), and so we generally only need to worry about low frequencies below a few Hz. Figure 22 shows the ground motion power spectra measured at several accelerator sites around the world. We can see that above ~1 Hz the amplitudes are typically less that 10 nm (100 nm in the worst case). At lower frequencies where the amplitudes are microns, the ground motion is highly correlated over relatively long distances (long wavelength surface waves). The effect of the correlation is to significantly reduce the impact on the collider performance, since the magnets move together.

From Figure 21 we see that a good choice of gain is 0.1 which will attenuate frequencies below $1/20^{th}$ of the repetition rate (a good rule of thumb). Hence the high RF frequency machines can typically attenuate beam motion less that a few Hz, which with a good choice of site (Figure 22) would seem sufficient. For TESLA with a repetition rate of 5 Hz, the cut-off is 0.25 Hz, which would appear insufficient for the purpose of stabilising the beam. However, TESLA can use the extremely long bunch train (2820 bunches in 950 μs) to very effectively perform feedback within the train itself; this system has a bandwidth of ~300 kHz, well above any vibration spectrum.

Although feedback alone would be sufficient to deal with the beam jitter generated from most quadrupoles in the machine, there still remains the problem of the FD with its ~1 nm tolerance. The FD is a very special case, and accordingly requires special attention. Mechanical stabilisation is foreseen for the warm RF machines (laser interferometer stabilisation systems using piezoelectric movers, and other passive damping devices are under investigation). For TESLA the intra-train fast feedback in principle will take care of 100 nm of FD motion (it is unlikely the FD will vibrate at

frequencies higher that 300 kHz!), but even here some stabilisation would probably prove beneficial.

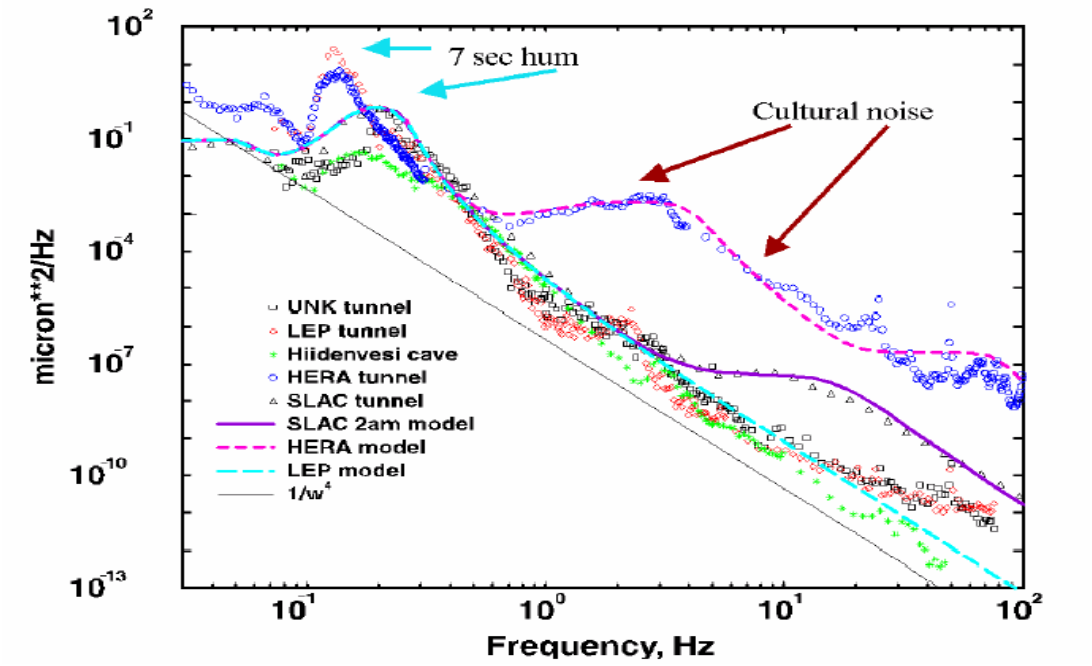


Figure 22: Ground motion spectra measured at various accelerator sites around the world.

Finally, we should mention long-term stability. In the previous discussion we have been mostly concerned with relatively fast vibration. However at very low frequencies, slow diffusive ground motion can move magnets by microns over the course of days. A model that is widely used for this slow ground motion in the so-called ATL law: consider two points on the ground separated by a distance L. after a time T, the variance of relative change in height of the two points (Δy) is given by

$$\left\langle \Delta y^2 \right\rangle = ATL \tag{73}$$

Where A is a constant depending on the site characteristics. Typical measured values of A range from 10^{-5} to 10^{-7} $\mu m^2/m/s$. Since ground motion is random, we can only talk about statistical quantities such as the variance.

Figure 23 shows the simulated effect of slow diffusive ground motion (based on (73)). Each curve represents the average over 20 different seeds of ground motion. The no feedback plot shows what will happen if no correction is made: the luminosity rapidly drops to 50% of its initial (design) value after 20 seconds. This is completely driven by the beams moving out of collision. Next we turn on the IP beam-beam feedback system to keep the beams colliding. The luminosity still drops but the time scale is now on the order of tens of minutes to an hour. The cause is the (relatively) slow decay of the upstream orbit which then generates spurious dispersion and cross-plane coupling, increasing the vertical beam size at the IP. To counter act this degradation, we must control (maintain) the orbit in all the magnets upstream. We achieve this by a slow orbit correction. Once we turn this correction on, we find that

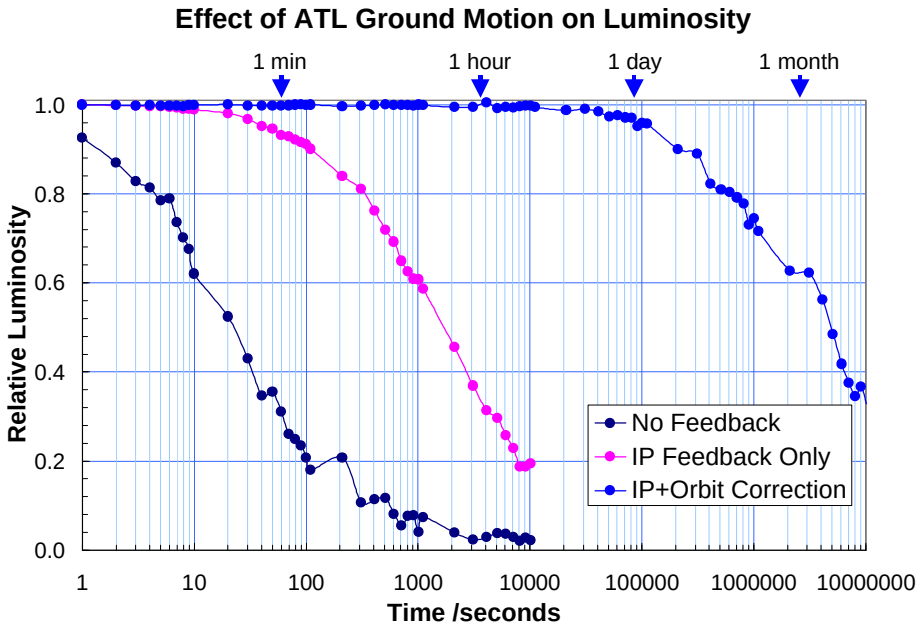


Figure 23: Simulated effect of slow ground motion on collider performance (in this case TESLA). Only the effect of the beam delivery system magnets are shown.

the luminosity is stable up to days. The slow reduction still observed is due (in this case) to the orbit correction algorithm not being completely dispersion free, and this residual dispersion eventually becomes large enough to reduce the luminosity. However, on this time scale use of semi-invasive tuning can be used to remove the dispersion without any significant hit in luminosity.

Here endeth the first lecture ☺