Linear Colliders as Factories for Z^0 and Heavier Particles

G. Guignard

Cern, Geneva, Switzerland

Abstract. After a brief review of the physics potential of linear colliders as Z^0 and heavier particles factories, the conditions required for the luminosity and the background to achieve good detection conditions are discussed. Considering an interval of RF frequency ranging from about 10 GHz to 30 GHz, the reasons for operating at such a high frequency when high accelerating gradient is required are recalled. Effects of strong wake fields on the linac energy-spread, transverse beam-stability and multibunch dynamics are summarized. The state of the art in building accelerating structures and keeping the trajectory within tight limits is given. Prospects for microwave klystrons as high-power RF sources for normal-conducting e^+e^- linacs are described briefly. Developments concerning a two-stage accelerator scheme proposed at CERN for generating the necessary power are presented. Finally, consistent sets of guessed parameters for linear-collider factories are proposed to estimate the performance that can be expected.

1 Perspectives and Main Issues

1.1 Introduction on Physics Potential

According to the physicists interested in the field of particle interactions [1], the open problems in the Standard Model at high energies can at present be divided schematically into three sectors: the fermion sector with the search for the top quark and its properties, the gauge boson sector including the $W^{\pm}$ and Z couplings as well as their static properties, and the Higgs sector with a quest either for their characteristic properties if they exist or for WW scattering if they do not. Beyond the Standard Model, possible extensions concern the gauge symmetry with the emergence of new gauge bosons $\left(W_R^{\pm}, Z_R\right)$, heavy neutrinos together with exotic quarks and leptons on the one hand, and the supersymmetry implying Higgs spectrum, scalar partners of quarks and leptons and spin-1/2 partners of gauge bosons and Higgs particles on the other hand.

For this kind of research, both hadron colliders and the lepton colliders can be used and are complementary given their specific characteristics and the domains of energy they are likely to cover. Considering the first type, there are actually two proton–proton colliders in the design or project stage respectively, the Large Hadron Collider at CERN (LHC) and the Superconducting Super Collider (SSC) at the SSC Laboratory in Dallas. The LHC is designed for a centre-of-mass energy $\sqrt{s}$ of 15.4

TeV and the SSC is aiming at 40 TeV for the same quantity. These figures correspond to an effective part available for the reaction smaller than or equal to about 2 TeV in the first case and perhaps 5 TeV in the second. By comparison, the design or studies of lepton colliders are far less advanced; it is nevertheless clear that they will be linear colliders owing to the fact that cost and size of rings for energies beyond LEP energies become prohibitive because of the unavoidable strong synchrotron radiation emitted by electrons and positrons travelling along a curve. Nowadays, the energies that look reasonable and are most often considered by both the engineers and the physicists are lumped around two levels which could be associated with two phases of linear colliders; a phase I with $\sqrt{s}$ around 300 to 500 GeV and a phase II with $\sqrt{s}$ between 1 and 2 TeV say. Phase I, adequate for top and Higgs, would be complementary to proton–proton colliders and phase II equivalent and complementary to pp colliders, though very promising in the search for novel phenomena and particles.

Since we are talking about linear colliders as factories, the target luminosity should be at least an order of magnitude above that of LEP (Large Electron–Positron ring at CERN), say. The LEP luminosity is at present about 10^{31} $cm^{-2}s^{-1}$ and will likely not exceed 10^{32} for a long time in the future, even with the multibunch scheme known as the "pretzel scheme".

Hence, the linear-collider factories should aim at a luminosity $\mathcal{L}$ of 10^{33} $cm^{-2}s^{-1}$ or higher. Starting with this tentative figure, one will try to convince the reader all the way through this review of the conceptual studies currently being carried out on linear colliders, that the desired goal at the energies considered is not out of reach. Before doing this, however, let us briefly indicate what such a luminosity means for the physics research.

Let us come back to the three sectors of the Standard Model. The mass of the top quark, that is a scale of the electroweak symmetry breaking, plays a key role in the fermion sector of the model and is a gate to the physics at high energies beyond the model. This mass is large, above 91 GeV and in the interval 130 to 150 GeV (±30 GeV) according to recent experiments carried out at CERN and Fermilab for instance. With a luminosity of 10^{33} $cm^{-2}s^{-1}$, the expected production rate would be of the order of 10 000 top/year. Top production, production threshold and decay modes could be studied and, for the mass range quoted above, predictions say that the mass of the top could possibly be determined with an uncertainty smaller than 500 MeV, to be compared with the uncertainty of about 5 GeV expected in a proton–proton collider. Considering next the interactions of electroweak bosons, there are many independent coupling parameters and possible constraints on these parameters to be determined. Linear colliders are interesting in this case, for they would provide bounds for these parameters that are more stringent than at hadron colliders. The third sector concerns the Higgs in the Standard Model, i.e. with masses below 700 GeV approximately. Let us note that in hadron colliders the search for Higgs is easy for masses between 140 and 700 GeV, but difficult below 140 GeV (and very difficult above 700 GeV). As complementary instruments, linear colliders look ideal for the Higgs search in an intermediate mass range from below 140 GeV continuing up to 350 GeV if $\sqrt{s}$ = 500 GeV. Considering again $\mathcal{L} = 10^{33}$ $cm^{-2}s^{-1}$ (corresponding to

an integrated luminosity of about 10 fb^{-1}/year), the production rate of Higgs with masses near 140 GeV would be ~ 1 000 events per year.

The next field to mention concerns the extensions of the Standard Model. Within extended gauge theories, linear colliders seem interesting in the search for novel leptons in all cases, precision studies (masses, couplings) of all novel fermions, and as factories of the neutral Z´ (issued from the theory) if the centre-of-mass energy is sufficient. By comparison, pp colliders are ideal in this case for direct production of vector bosons up to about 5 TeV. Next comes the supersymmetry that aims at the unification of bosons with fermions and is a plausible step towards quantum theory of gravity. The associated particle spectrum includes five physical Higgs particles (three neutral and two charged, with a non-zero probability that all masses be below 250 GeV), sleptons and squarks (spin-0 partners to fermions), and gluinos and Higgsinos (spin-1/2 partners to gauge and Higgs bosons). Again, linear colliders look ideal for the search for these Higgs in the intermediate mass range (while pp colliders are appropriate for higher masses up to 800 GeV) and for measurements of spin or couplings.

In summary, linear colliders that could serve as Z^0 factories (though then in competition with existing ones) could also be very interesting as top and Higgs factories provided the appropriate performance can be reached, and the questions of the energy spread and of the number of underlying $\gamma\gamma$ events are solved. It indeed appears that precision measurements of the top's characteristics and coupling parameters are then possible as well as the search for and exploration of new phenomena in the lepton /Higgs sector.

1.2 Background and Interaction-Point Issues

During the collisions, beam–beam forces enhance the luminosity and the angular opening of the spent beam. But, at the same time, electrons and positrons radiate "beamstrahlung" photons in the bending field generated by the opposite bunch. These photons have basically two detrimental effects: they carry energy from the electrons and positrons, create parasitic e^+e^- pairs and generate a background of $\gamma\gamma$ processes. The beam parameters at the interaction point (I.P.) must therefore be adjusted to keep these effects at an acceptable level.

The first detrimental effect corresponds to a degradation of the resolution on the centre-of-mass (c.m.) energy of the collision, described by the luminosity distribution versus s (c.m. energy square) termed "differential luminosity". This degradation corresponds to a dilution of the luminosity towards low energies [2] that can be characterized by three most significant parameters:

i) the pinch enhancement factor H_D of the luminosity due to the bunch field-forces,
ii) the average fractional energy loss δ_E due to beam–beam radiation ("beamstrahlung"),
iii) two thirds of the fractional average critical photon energy in the classical regime (noted Y).

Flat beams with a high aspect ratio σ_x/σ_y make it possible to decrease all three parameters with respect to the values they take with round beams. The reduction of

the enhancement factor H_D might seem unfavourable at first view, but it is associated with a beneficial reduction of the luminosity dilution and of the average energy loss δ_E (which remains large with respect to the acceptance of the final-focus system that precedes the I.P., but smaller than the one determined by the initial-state radiative corrections). This means that with flat beams there are more annihilation events with a total energy close to the c.m. energy available. For instance, with the CLIC parameters and an aspect ratio of ~ 11, the fraction of luminosity above 95% of the c.m. energy is about 70%.

The second important effect is the presence of background events in addition to the wanted e^+e^- annihilation processes [3]. It turns out that background is due to the two-photon process, since the incident electron and positron beams in a linear collider carry with them virtual photons, usually called "Weiszäcker–Williams" photons, as well as real "beamstrahlung" photons that can interact together in one of the following ways:

$$\gamma\gamma \rightarrow \text{hadrons} \quad \text{(real–real)}$$
$$\gamma e^{\pm} \rightarrow e^{\pm}\,\text{hadrons} \quad \text{(real–virtual)}$$
$$e^+e^- \rightarrow e^+e^-\,\text{hadrons} \quad \text{(virtual–virtual)}$$

The convolution of the virtual and real photon energy spectrum gives the effective differential luminosity due to two-photon events, which can then be compared with the annihilation differential luminosity. If the former is too high, the annihilation events' study can be difficult, either because the readout trigger is swamped by two-photon events that cannot easily be distinguished from the wanted events, or because too many valuable annihilation triggers would contain secondaries from simultaneous two-photon interaction. It is moreover not excluded that particles be produced at high transverse momenta by two-photon collisions in which photons transform virtually to vector mesons, and constituent quarks or gluons from the meson collide to produce high-transverse-momentum (p_T) "minijets". Nevertheless, the dominant process by which photons interact to produce hadrons is through a state in which both photons become virtual hadrons. Since the interaction volume is typical of hadron dimensions (~1 fm^3), the final products tend to have limited p_T. By contrast, e^+e^- annihilations produce quark–antiquark pairs in a pointlike process and the quarks generate hadrons in two jets with high transverse momenta. This difference and the fact that the $\gamma\gamma$ frame relative to the laboratory is boosted with final particles coming out at rather small angles suggest that event selection criteria can be found to suppress or reduce two-photon background relative to annihilations.

Possible detection criteria have recently been studied with the support of numerical simulations [3]. Any general-purpose detector is sensitive only to tracks emitted with angles greater than some limit θ_{min} optimized to obtain acceptable parasitic e^+e^- pairs (θ_{min} of the order of 150 mrad). Taking this into account, trigger requirements to keep $\gamma\gamma$ background low can be based on the total energy and transverse momentum in the final state, depending on what precedes. Total energy, close to $\sqrt{s}$ only in the absence of beamstrahlung degradation and the escape of secondaries, can for instance be set to be above a threshold of the order of 5% of nominal $\sqrt{s}$ to prevent $\gamma\gamma$ events from triggering without affecting too much the efficiency for annihilation events. To help in this challenge, triggers can also be based on p_T

requirements for the secondaries; one possibility is to define a minimum requirement on the "transverse energy" sum, while another way consists of requiring that at least one track in the event have a momentum p_T larger than some specified threshold of the order of 1% of $\sqrt{s}$. Monte Carlo results indicate that it should not be difficult to get clean annihilation measurements in linear colliders in this way, in particular for events at energies above 0.9 s, assuming one is prepared to use higher thresholds than those quoted and to lose the corresponding small fraction of good events taking place at lower energies.

Although $\gamma\gamma$ events are sufficiently different from e^+e^- events to make their contributions to the trigger rate insignificant, the probability of such an event occurring within the resolving time of the detector could be so high that every legitimate annihilation would be superposed on a two-photon event (in particular for colliders with high luminosity per bunch-crossing). This effect can be characterized by the mean number of two-photon events per bunch-crossing, also called occupancy

$$\Omega \sim \mathcal{L}_b\left(1+n_\gamma\right)^2 \sim \frac{RN^2}{f_{rep}(1+R)^2}$$

where $\mathcal{L}_b$ is the luminosity per bunch crossing and n_γ the average number of photons per incident particle. To reduce detector occupancy to the desired level, it is hence judicious to lower the number of particles N per bunch, increase the repetition rate f_{rep}, and, last but not least, use flatter beams (higher aspect ratio R that also limits the luminosity dilution, as mentioned above). Reducing N while keeping the same total luminosity might imply the presence of several bunches per RF pulse. This would, however, reduce the occupancy only if the detector can resolve events coming from the successive bunch-crossings. Once these few parameters have been selected, the experimenter can still use different strategies to minimize the effects of the underlying $\gamma\gamma$ events, like increasing the cut-off angle θ_{min} (background cuts are faster than signal cuts) and profiting from the jet structure of the annihilation events with balanced p_T.

If it is obviously essential to achieve workable experimetal conditions, the beam focusing at the I.P. also contains critical issues [2] briefly recalled hereafter. The required high-demagnification telescope induces transverse aberrations that necessitate the presence of a chromaticity correction section with sextupoles and subsidiary bending magnets to create dispersion. It also implies tight tolerances on field errors and misalignments to limit luminosity losses due to beam offset and coupling. In some cases, wake fields in small-aperture elements may generate a too large emittance growth. But designs of final-focus systems taking these effects into account have been worked out.

1.3 Accelerating Gradient and RF Frequency

The total average RF power that is required in one linac is given by

$$P_{RF} = \frac{P_b}{g^2\eta} \qquad (1)$$

where g^2 is the filling efficiency of the accelerating sections (i.e. the fraction of input energy left at the end of the fill time τ_0) and η the fractional energy extraction by the beam [4]. The beam power P_b is proportional to the final particle energy eU, the number N of particles in the beam and the repetition rate f_{rep},

$$P_b = eUNf_{rep} \tag{2}$$

while the fraction of stored energy extracted by a charge Ne is given by

$$\eta = \frac{NeZ_0\omega_0^2}{2\pi c\, E_0}. \tag{3}$$

The quantity Z_0 is frequency-independent and only depends on the shape of the accelerating structure; it is the shunt impedance over Q factor per RF wavelength and is related to the shunt impedance over Q factor per unit length r'_0 via

$$\omega_0 Z_0 = 2\pi c\, r'_0. \tag{4}$$

The circular frequency ω_0 is equal to 2π times the RF frequency f_0, and E_0 stands for the accelerating gradient.

Since a high-energy extraction is desirable, the expression (3) advocates for high RF frequencies at a given beam charge. In particular, when aiming at very high gradients in order to limit the total length of the linac, very high frequencies f_0 are required. The highest possible value is eventually limited by the manufacturing and alignment tolerances as well as the wake fields, and this limit constrains the gradient E_0 actually attainable.

Other constraints on E_0 come from the difficulty in generating the required power. The peak power per unit length $\hat{P}/L_0$, in a structure of shunt impedance per unit length R'_0 is indeed given by

$$\frac{\hat{P}}{L_0} = \frac{E_0^2}{g^2 \alpha R'_0 \eta_2} \tag{5}$$

where α is the power flow attenuation constant and η_2 is the efficiency of the energy transfer from the source to the linac structure. The power requirements (5) related to the desired E_0 may be large and it is a challenge to develop power sources delivering the necessary energy and working at high frequency.

The scaling of E_0 with ω_0^2 suggested above keeps constant the fraction η and the average power P_{RF}, by virtue of Eqs. (1)–(3). In these conditions the peak power per unit length $\hat{P}/L_0$ is proportional to

$$\frac{\hat{P}}{L_0} \sim \omega_0^{-1/2} E_0^2 \sim \omega_0^{7/2} \tag{6}$$

which includes the fact that R'_0 varies with $\sqrt{\omega_0}$.

Increasing E_0 in this way can be considered until a limit either on the manufacturing tolerances or on the development of power sources is reached, as mentioned already.

A third limitation may arise from the wake fields whose peak values depend on ω_0, the iris aperture a and the loss factor k_0 as follows [4]:

$$\hat{W}_T^{\delta} \sim \frac{k_0}{\omega_0 a^2} \sim \frac{R_0}{Q_0 a^2} \sim \frac{\omega_0}{a^2} \sim \omega_0^3$$
$$\hat{W}_L^{\delta} \sim k_0 \sim \frac{\omega_0 R_0}{Q_0} \sim \omega_0^2 \qquad (7)$$

The rapid increase of the point-charge wakes (mainly transverse) with ω_0, and the concomitant beam blow-up are the sources of this limitation.

Since both the achievable gradient and the detrimental effects increase with ω_0, the choice of the RF frequency results from a compromise. The frequency range considered in this study report goes from about 10 GHz to 30 GHz, while the accelerating gradient is supposed to be between 50 and 100 MV/m. This corresponds to linacs suitable for either CLIC (CERN Linear Collider using a Drive Linac) or NLC (Next Linear Collider using pulsed RF generators) as envisaged at SLAC, KEK or IHEP. Table I gives considered values of the parameters discussed above, for the different proposals.

Table I. Examples of parameters

		E_0 (MV/m)	ω_0 (GHz)	$\hat{P} / L_0$ (MW/m)
CLIC		80–100	30.0	150
NLC:	SLAC (NLC)	50	11.4	60–240
	KEK (JLC)	100	11.4	
	IHEP (VLEPP)	100	14.0	

2 Beam Dynamics in the Linacs

2.1 Wake Fields and Single-Bunch Dynamics

As mentioned above, high-current beams induce in a high-gradient accelerator strong electromagnetic fields that increase rapidly with the RF frequency. These wake fields are responsible for energy loss, energy spread and transverse blow-up. Longitudinal wakes directly influence the distribution of energy inside the bunch, its contribution diminishing the accelerating field seen by the particles. Since these wakes are not uniform within a bunch, the energy loss is accompanied by an energy spread that induces variation of the focusing strength and dispersion of the trajectories inside the bunch in the presence of external magnetic quadrupoles. Transverse wakes deflect parts of the bunch and these deflections depend on the particle momentum and position with respect to the accelerator axis. Because of the energy spread and trajectory dispersion, dipole wakes produce kicks changing along the bunch and eventually producing an apparent beam blow-up (the "head" of the bunch deflecting the "tail"). Quadrupole wakes may also be present but they have been either considered as negligible or not yet studied in the different proposals.

Some compensation of the energy spread σ_E using wake potential versus RF sine wave is possible to high orders [4,5]. This is strongly desirable if the momentum acceptance at the exit of the linac is limited, as for instance in a final-focus system

that typically accepts a $\Delta p/p$ of $\pm$ 2‰ to $\pm$ 5‰. The total accelerating gradient seen by a particle at position z in the bunch can be written:

$$G(z) = G_{RF} \cos(\omega_0 \frac{z}{c} - \phi_0) - W_L(z) \tag{8}$$

and the balance between the RF wave and the longitudinal wake will obviously depend on the phase ϕ_0, but also on the bunch charge N and bunch length σ_z that enter in W_L. Knowing G(z) it is possible to find ϕ_0 and σ_z values which minimize the spread σ_E for a given N. Furthermore, the energy distribution $\nu(E)$ can be calculated [5] in order to study its properties and dependence on G_{RF}, ϕ_0, σ_z and N, by using

$$\nu(E) = \frac{1}{N}\frac{dN}{dE} = \frac{1}{N}\frac{dN}{dz}\frac{dz}{dE} = \frac{\rho(z)}{dE/dz} = \frac{\rho(z)}{e\,dG(z)/dz} \tag{9}$$

where $\rho(z)$ is the charge distribution (Gaussian). Results concerning CLIC are given as examples [5]. Figure 1 shows G(z) for different charges N with ϕ_0 = 7–8.5° and σ_z = 0.14–0.17 mm. Figure 2 gives the corresponding energy distributions for $N = 5 \times 10^9$ and 6×10^9 per bunch. The flatter G(z), the smaller σ_E, and, in these two cases, the values obtained are about 4.6‰ and 7.7‰ respectively (in the absence of truncation of $\nu(E)$). Note also that this adjustment ends up with a minimal tail population and a somewhat narrow core size (Fig. 2).

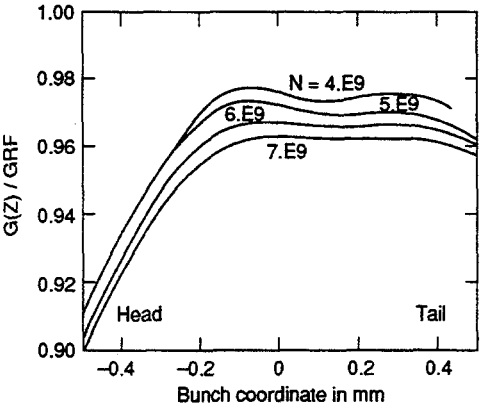

Fig. 1. Total gradient minimizing the energy spread σ_E

Transverse instabilities due to dipole wakes are all the more critical the smaller the emittance. In the linac of colliders, the vertical emittance may be very small since the beam is usually flat. Having a flat beam or a large aspect ratio $R = \sigma_x/\sigma_y$ at the collision point reduces the average energy loss due to synchroton radiation. This energy degradation due to beamstrahlung induces a large energy spread that should not, however, exceed the energy spreading related to background processes (imposing a limit of ~5% for tolerable σ_E from beamstrahlung). Furthermore, a large ratio R allows the avoidance of an excessive repetition rate that varies for constant luminosity and beam–beam radiation σ_E as follows:

$$f_{rep}(R) = f_{rep}(R=1)\frac{4}{RH_y} \tag{10}$$

where H_y is the vertical pinch enhancement factor (typically between 2 and 2.5). For these reasons, values of R and normalized vertical emittance $\gamma\varepsilon_y$ are respectively large and small in the proposals quoted (Table II).

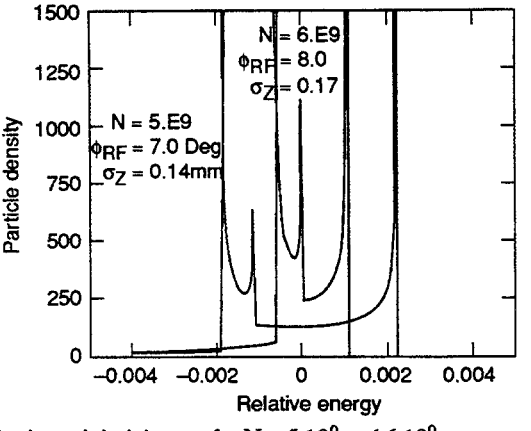

Fig. 2. Energy distribution minimizing σ_E for $N = 5{\cdot}10^9$ and $6{\cdot}10^9$

Table II. Beam aspect ratio at final focus and V-emittance

	R	$\gamma\varepsilon_y$ (rad m)
CLIC	10–20	$2{\cdot}10^{-7}$
NLC	100	$3{\cdot}10^{-8}$

The presence of strong dipole wakes (almost 20 times larger in CLIC than in NLC) implies large kicks originating from misalignments of the structure and off-centred trajectories. The corresponding equation of the transverse motion of single particles is given by

$$y'' + k_0^2 y = \left[k_0^2 - k^2(z,s)\right]y + \frac{r_0}{\gamma}\int_{-\infty}^{z} \rho W_T^{\delta}\left(z^* - z\right)y\left(z^*,s\right)\,dz^* \qquad (11)$$

where k^2 characterizes the linac focusing (k_0^2 being its "average" value, independent of z) and the wake-field effect is integrated over the heading part of the bunch of charge distribution ρ. To counteract this effect, the idea consists of obtaining a coherent motion by imposing the same oscillation period to all particles. This condition, called autophasing by its author [6], comes directly from inspecting Eq. (11):

$$k^2(z,s) = k_0^2 + \frac{r_0}{\gamma}\int_{-\infty}^{z} \rho W_T^{\delta}\left(z^* - z\right)\,dz^* \qquad (12)$$

All the proposals for high-gradient, high-frequency linear accelerators strive to satisfy condition (12) or its linearized version [7]

$$\left.\frac{\partial k^2}{\partial z}\right)_{z=0} = \frac{r_0}{\gamma} N \left.\frac{\partial W_T}{\partial z}\right)_{z=0} \qquad (13)$$

to limit beam blow-up. There are basically two ways for achieving this variation of the focusing strength with the position z inside the bunch:

1) Using the external focusing of a magnetic FODO lattice, the change of k^2 with z can be obtained via an imposed energy spread σ_{BNS}, since $k^2 = k_0^2 / p$ if p is the particle momentum. The required energy spread may come in turn from the dependence on z [Eq. (8)] of the accelerating gradient G combined with an adjustment of ϕ_0. A negative phase ϕ_0 is needed to ensure that the bunch tail is more focused than the head, according to Eq. (12). The subsequent σ_{BNS} ensuring stability is as large as 5% in CLIC, but only 0.2–0.6% in NLC proposals (in proportion to their wake fields). Note that this requirement conflicts with the minimization of σ_E discussed above and based on a positive phase ϕ_0, in particular for high-RF frequency linacs.

2) An elegant way to avoid the conflict with σ_E minimization and simultaneously create the spread in k^2 without the detour of a large energy spread does exist. It consists of generating part of the transverse focusing directly from RF fields oscillating at the frequency of the accelerating fields, in so-called microwave quadrupoles [8]. Since the radial electric field in a narrow slit vanishes in the mid-plane, the effective magnetic gradient due to the axial electric field and deduced from Maxwell's equations is given by [8]

$$G_m(T/m) = \frac{\pi}{c\lambda_{RF}} E_0(MV/m)\sin\phi_1 \qquad (14)$$

where E_0 is the peak accelerating gradient, λ_{RF} the RF wavelength and ϕ_1 the RF phase angle measured from the top. In the direction perpendicular to the slit, the electric field is doubled (compared with circular aperture) and overcompensates the magnetic gradient by exactly a factor 2, thus forming a quadrupole. In practice an oval cavity with circular aperture (Fig. 3) is preferred to a circular cavity with slotted iris [8], in order to have the required radius and surface finish at the aperture.

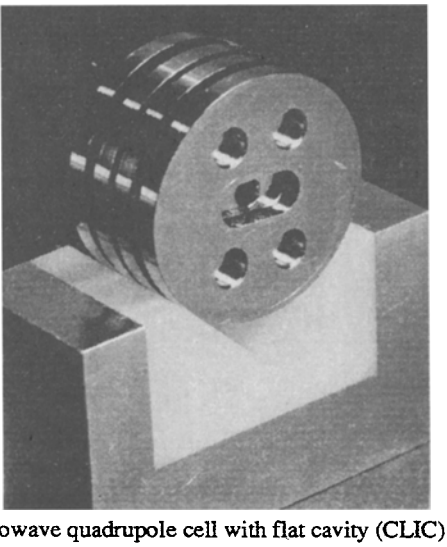

Fig. 3. Microwave quadrupole cell with flat cavity (CLIC)

In CLIC, where transverse wakes are large, it is proposed to generate the spread of k^2 with microwave quadrupoles using a phase ϕ_1 close to the phase ϕ_0 that minimizes σ_E, i.e. running near the maximum accelerating voltage. In this solution [9], the main basic focusing k_0^2 is created by external magnetic quadrupoles, while the microwave quadrupoles are only responsible for the variation $k^2 - k_0^2$. Hence, transverse instabilities can be damped (Fig. 4) keeping the energy spread low.

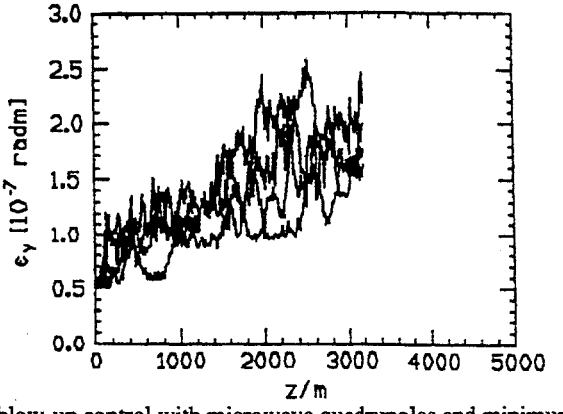

Fig. 4. Example of blow-up control with microwave quadrupoles and minimum σ_E

2.2 Wake Fields and Multibunch Dynamics

Accelerating long bunch trains that extend over a period comparable to the filling time τ_0 of a cavity section may provoke instabilities of the whole train along the linac, due to interactions with RF fields. The shape, timing and modulation of the RF pulse, as well as long-range longitudinal wake field, are responsible for bunch-to-bunch energy variations. The subsequent energy spread can cause filamentation and emittance growth beyond the acceptance at the exit of the linac. Interactions with resonant transverse electromagnetic fields in disk-loaded waveguides, in particular with the so-called HEM_{11} dipole modes, produce increasing deflections along the bunch train that drive a transverse instability, called beam break-up.

The bunch-to-bunch energy variation produces effects that are more severe with long trains. The fundamental mode, as well as high-order modes of longitudinal wakes, is at the origin of inter-bunch beam loading and the actual RF pulse influences the energy spectrum. There is consequently a need to control bunch-to-bunch energy spreads and some compensation schemes have been studied [10,11]:

1) Matched filling, i.e. adjustment of the injection timing of the bunch train with respect to the RF pulse and appropriate choice of the bunch spacing. The idea is to have sufficient extra energy in the RF section fill between bunches to cope for the energy lost in accelerating the preceding bunches.

2) Staggered timing, i.e. delay of a subset of klystrons so that some accelerating sections are only partially filled during build-up of the beam-loading voltage to its steady-state value. The number of delayed klystrons is selected to produce a voltage equal to about twice the steady-state beam-loading voltage in the linac.

3) Modulation of RF input, i.e. phase adjustments or small klystron variations during the time when the bunch train is passing through a cavity section. This makes use of the propagation out of the section of the leading edge of the pulse while the train is passing over.

The results of such compensation schemes depend on the bunch length with respect to the section filling-time and on the bunch separation. In the NLC for instance [10], the first method applies preferably to short trains (10 bunches of 10^{10} particles/bunch, lasting about 10% of τ_0). Figure 5 shows the energy deviation obtained with bunch separation of 16 RF wavelength, 5ns RF pulse rise-time and dispersion of RF frequency components. The fractional energy deviation remains below about 3‰ but ~75% of the maximum accelerating gradient is used. With long trains (70 bunches or more), the second method seems more appropriate and Fig. 6 shows results obtained in the same conditions of rise-time and dispersion. If the total energy deviation is about the same, only ~65% of the maximum gradient (assumed to be 50 MV/m) is available. Multibunch dynamics in CLIC has not yet been studied. Performance might probably require only 2–4 bunches per beam (perhaps up to 10) separated by 10–20 λ_{RF} ($\lambda_{RF}/c \cong 33$ ps) while the RF section filling-time is 11.1 ns. Hence, the first method seems applicable but this must still be checked.

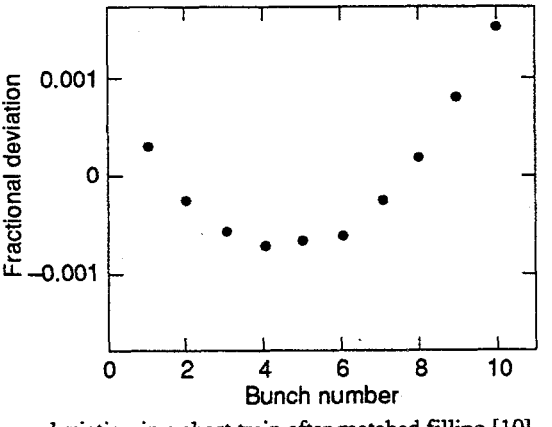

Fig. 5. Fractional energy deviation in a short train after matched filling [10]

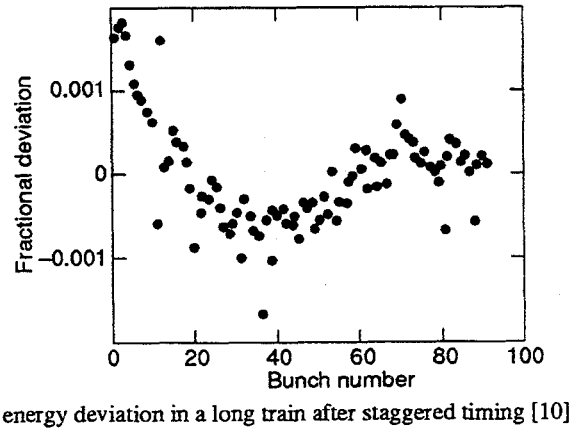

Fig. 6. Fractional energy deviation in a long train after staggered timing [10]

Cumulative beam break-up due to long-range dipole modes of transverse wake fields can be severe. The transverse beam modulation is carried along the linac from accelerating section to accelerating section, through the beam. Blow-up then occurs and manifests itself along the linac as an amplitude growth from the head to the tail of the bunch train. Two possible cures have been investigated [12,13,14]:

1) damped structures; modified disk-loaded waveguides in which the power of the wake field modes is coupled out to lossy regions through radial slots in the disks and/or azimuthal rectangular waveguides, whereby the external Q-factor of the undesirable HEM mode is lowered to values typically below 20. This was studied in SLAC and KEK, the latter requiring Q-values between 15 and 70 for the predominant TM_{11} modes [15] (in order to limit the emittance growth within a factor $\sqrt{2}$ and alignment tolerances within 80 μm). The limitation of this method might come from the low Q-values required and the large number of cells involved, the practical difficulties increasing with the RF frequency.

2) staggered tuning [12]; variation in the cell dimensions in each accelerating section resulting in a cell-to-cell spread (by 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 frequency distribution seems to be a truncated Gaussian, since the initial roll-off of the wake is strong, with low partial recoherence within the length of a (short) bunch train.

Whilst the fabrication of damped structures has been tested [14], the possibility of staggered tuning has been investigated by numerical simulations [16] and experimental measurements [17]. They concern a detuned 50-cavity disk-loaded structure with iris diameter ranging from 0.83 cm to 1.22 cm and Gaussian HEM frequency population centred at 14.45 GHz for a standard deviation of 1.07 GHz. The Advanced Accelerator Test Facility at Argonne makes it possible to measure the energy variation and the horizontal position of a witness bunch following the driving bunch in a time interval between 0 and 1 ns. The former yields the longitudinal wake potential and the latter gives the transverse potential as the structure is swept horizontally. Figure 7 compares calculation with experiment [17] and confirms roll-off expectation in this particular case, though it is not clear if recoherence takes place at larger distances z from the driving bunch (in NLC bunch separation is about 4 cm and a "short" train would cover ~ 3.5 m). No emittance blow-up simulations are known to the author for RF frequencies between 10 and 30 GHz, but the gain expected from staggered tuning can be illustrated by simulation results [18] obtained with 180 bunches, at lower frequency (3 GHz) and with $N_f = 25$ (Fig. 8). The effectiveness of staggered tuning is visible from the fact that without it about half of the bunches have too large amplitudes at the end of the linac, and from the difference in the scale used for the top picture (mm) and for the bottom one (μm).

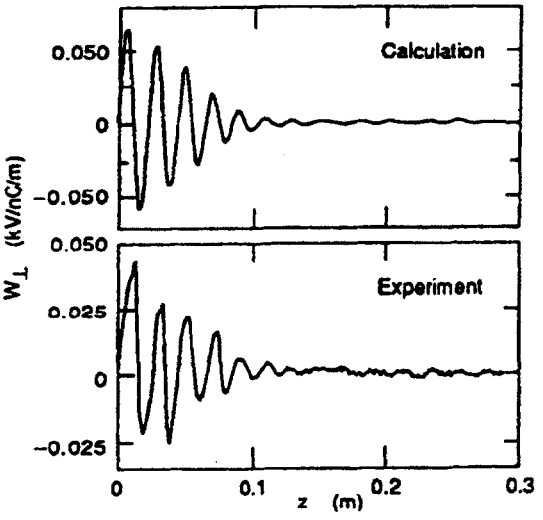

Fig. 7. Calculated and measured transverse wake potential for a detuned 50-cavity structure [17]

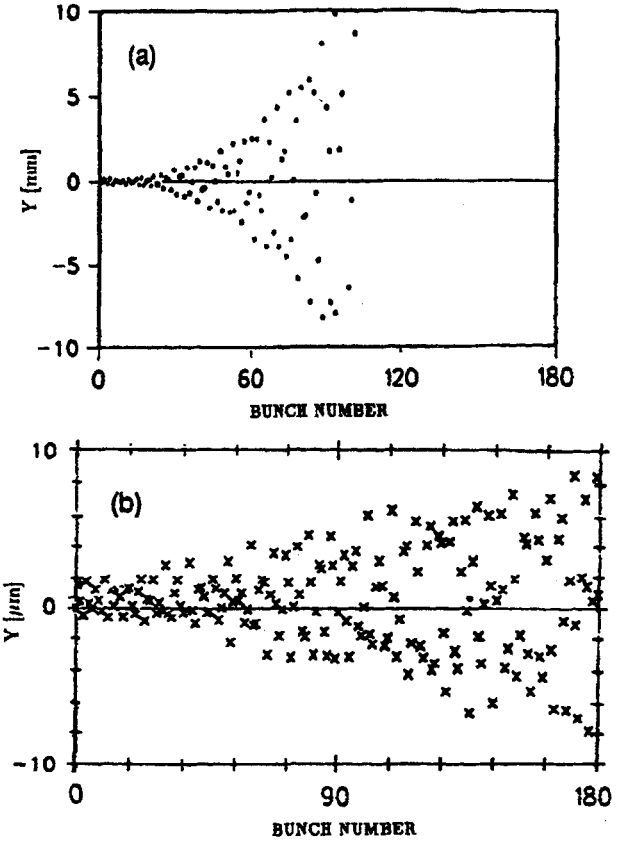

Fig. 8. Transverse bunch offsets at linac end, without (a) and with (b) staggered tuning [18]

3 Linac Hardware Challenges

3.1 Constraints on the Fabrication of Accelerating Sections

The requirement for high gradient at high frequency calls for tight tolerances in the fabrication of the accelerating cavities. Approximately the same principles have been adopted in the design of the accelerator sections proposed at CLIC and at the JLC or VLEPP versions of NLC. The most promising manufacturing method, tested at CLIC, is the brazing of machined copper cups, and the construction of prototypes of CLIC structures [19] proved that this technique could be successfully acquired by industry.

For the CLIC 30-GHz structure, the tolerances on the cell dimensions (iris diameter of 4 mm and cell diameter of 8.7 mm) are of the order of 2–5 μm (in order to limit the total phase error to 5° over a section) and on the surface finish of about 0.05 μm (to obtain 95% of the nominal Q-value). The high-precision copper cups were made on Pneumo diamond-tool lathes, with laser interferometric feedback with 25 nm resolution. The machining accuracy consistently achieved by manufacturers would eliminate the need for dimple tuning, since the phase shift error was approximately 0.1°/cell. High-quality brazed joins were produced, with two diffusion-bonded annular surfaces at the inside and outer edges of the disks, to prevent braze leakage. Unacceptable frequency changes could be avoided during brazing operations. Complete structures (Fig. 9) with radial holes and vacuum manifolds for pumping and channels for cooling have been manufactured and measurements in the laboratory confirmed the expected parameter values [19]. In the JLC design, for example, damping slots could be incorporated in addition.

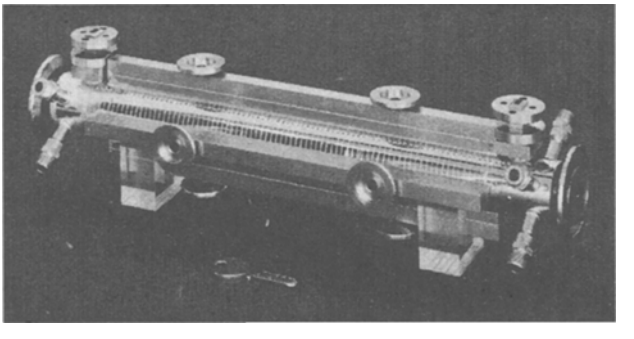

Fig. 9. Finished prototype of a 30 GHz accelerating section (CLIC)

Field-emitted electrons can create multipactoring resonance discharge at well defined field levels, and also dark currents due to the capture, bunching and acceleration of these electrons, eventually producing a parasitic beam. Recently, work was carried out on possible scaling laws [20] for these phenomena with the operating frequency. Basically, both phenomena scale with the frequency ω_0 for exact geometry scaling, but are influenced by the cavity shape near the axis for the same iris opening a. Working at high frequency with a large iris opening (or large a/λ_{RF} ratio) should be favourable in this respect. Numerical estimates [20] for CLIC cavities at 30 GHz indicate absence of electron capture and dark currents up to a

gradient as high as 1400 MV/m. A possible limitation around 100 MV/m seems rather to come from second- or third-order multipactoring, in which emitted electrons lose energy by successive impacts, drift to the outer diameter, and eventually give rise to a breakdown of the structure. Further investigations are needed to better understand these phenomena that already look more critical at the lower end of the frequency interval considered.

3.2 Tolerance on the Alignment of Linac Elements

The dominant effect associated with misalignments is the blow-up of the projected emittance resulting from the incoherent motions due to transverse dipole wake fields (as already mentioned above). The acceptable growth of the normalized emittance $\gamma\varepsilon_y$ may vary from only 14% to a factor 3 or 4 depending on the proposal requirements. The subsequent alignment tolerances can be deduced from these numbers assuming first a simple one-to-one trajectory correction aiming at centring the beam in each position monitor (BPM) by moving the preceding lattice quadrupole. To give examples, the corresponding alignment tolerances for CLIC and NLC (SLAC version) are the following (r.m.s. values):

CLIC 3 μm on quadrupoles, 5 μm on structures,
NLC 7 μm on quadrupoles, 4 μm on structures.

In order to relax these tolerances while keeping the same acceptable emittance growth, a trajectory correction more efficient than the one-to-one method has to be applied. SLAC [21] proposed compensating for the dispersion while correcting the orbit and minimizing the wake-field dilutions caused by the corrected trajectory. The minimization procedure developed for achieving this is a weighted least-squares that minimizes the following sum:

$$\Phi = \sum_{j\in\{BPM\}} \frac{\left(\Delta y_{QF}\right)_j^2}{2\sigma_{prec}^2} + \frac{\left(\Delta y_{QD}\right)_j^2}{2\sigma_{prec}^2} + \frac{y_j^2}{\sigma_{al}^2 + \sigma_{prec}^2} = \Phi_{min} \qquad (15)$$

where y_j is the corrected trajectory amplitude, Δy_{QF} the difference trajectory resulting from both QF-variations and corrector adjustments (idem for Δy_{QD}, related to QD-variations however), σ_{prec} is the r.m.s. precision of the BPM readings and σ_{al} the r.m.s. BPM misalignments. Since the QF- and QD-fields are opposite and the QF- and QD-strengths are both supposed to be decreased when measuring Δy_j, the sum of these two variation terms mimics the effects of the dispersion (trajectory shift with momentum or quadrupole-strength deviation), while their difference mimics the effects of the wake field (sign depending only on the side where the trajectory is off-centred and not on the quadrupole field polarity). With this algorithm, the NLC (SLAC) alignment tolerances could be relaxed to 70 μm for both the quadrupoles and the accelerating structures. If all the quadrupoles are simultaneously detuned, only one difference Δy corresponding to the dispersion appears in the function Φ. Minimizing it was called dispersion-free correction while minimizing (15) was termed wake-free correction by its author [21]. Figure 10 shows beam distributions after one-to-one, dispersion-free and wake-free corrections in NLC. Plots on the left

are projections onto the y-y′ phase plane, while right-hand plots are projections onto the y-z plane (z: longitudinal coordinate). One sees the strong dilution in (a), reduced to wake-field tail bend in (b) and minimized in (c).

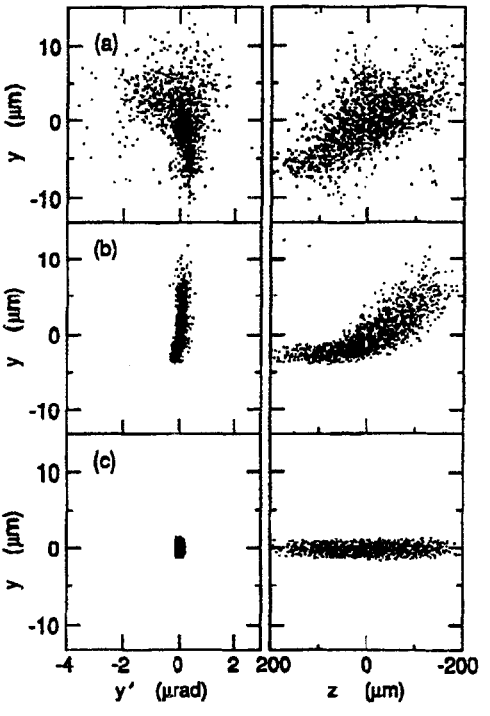

Fig. 10. Beam distributions after one-to-one (a), dispersion-free (b), and wake-free (c) correction in NLC [21]

CLIC started from this idea and focused on an achromatic trajectory correction [22] to higher order, developing the trajectory differences Δy_j (measured) and ΔY_j (due to corrections) in $\delta = \Delta p/p$,

$$\Delta y_j = \sum_n a_n^j \delta^n \ , \quad \Delta Y_j = \sum_n A_n^j \delta^n \tag{16}$$

Calling y_j the measured trajectory and Y_j the one due to corrections, the function to be minimized becomes,

$$\Phi = \sum_{j \in \{BPM\}_y} \left\{ w_0 \frac{y_j + Y_j}{\sigma_{al}^2 + \sigma_{prec}^2} + \sum_{n \geq 1} w_n \delta^{2n} \frac{a_n^j + A_n^j}{2\sigma_{prec}^2} \right\} = \Phi_{min} \tag{17}$$

where the sum applies to the quadrupoles focusing in the plane considered, in order to avoid too large a wake dilution. Each variation term of the second sum in Eq. (17) represents the dispersion order by order. In the applications, second-order corrections are actually implemented. The efficiency of this global correction is illustrated in Fig. 11 showing the residual trajectory and the emittance dilution in a sector of CLIC, for 5 μm alignment tolerances on the quadrupoles [23]. Combining this with a one-to-

one algorithm realigning the quadrupoles with the beam and a peculiar lattice scaling, the CLIC alignment tolerances could probably be relaxed to about 50 μm for the quadrupoles and 10 μm for the accelerating structures.

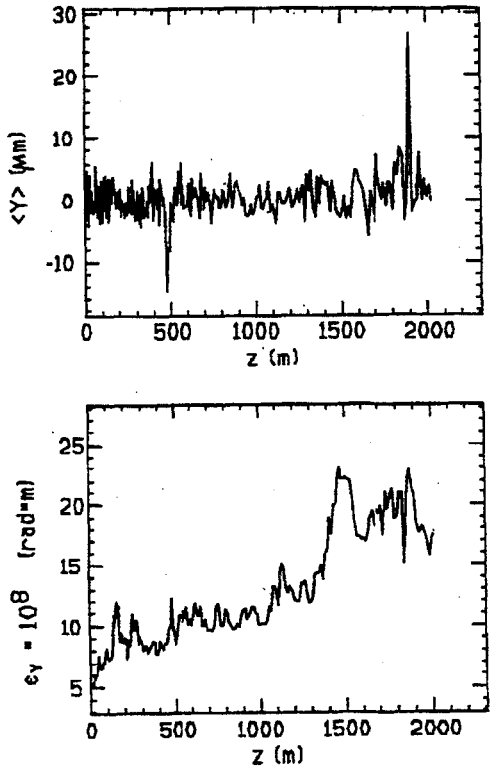

Fig. 11. Trajectory (top) and blow-up (bottom) after global correction in a 2 km-long CLIC sector

4 Power Generation

4.1 Microwave Klystrons as RF Power Sources

Conceptual designs of linacs for future colliders operating around 11 or 14 GHz call for microwave klystrons able to deliver as much as 100–200 MW in pulse lengths of the order of 1 μs. These requirements cannot be satisfied with existing microwave tubes and new klystron designs meet a certain number of challenges briefly recalled hereafter. The maximum power capability is limited by the area available to dissipate beam or RF losses and shrinking with the inverse of ω_0^2. Good power transfer efficiency from the beam to the output circuit and possible release of the intrabeam space charge forces favour high RF voltage V as well as low perveance defined by $I/V^{3/2}$; this implies a large RF gradient across the output gap. Permitting greater beam current I and power makes it necessary to achieve a higher beam convergence that involves better confinement and more precise beam optics. Finally, these high-

current, high-voltage conditions increase the risk of failure mechanisms limiting the power; this concerns possible RF breakdowns mainly in the output circuit as well as intrapulse heating due to beam interception, the two mechanisms being interrelated. Typical figures considered are 40–50% for the power transfer efficiency, about 1 or 2 $\mu A/V^{3/2}$ for the perveance, beam convergence as high as 200, a gradient lower than 6 MV/cm in the output gap and a scraped beam fraction below 1%.

There are different means envisaged for trying to find solutions. Working at low perveance by increasing V is an important element for beam control, as empirically demonstrated for operational RF sources. Using excellent beam optics near the output circuit and reducing gap voltages by replacing the resonant cavity by multiple gaps or travelling-wave output are possible improvements. The klystron gun voltage can be divided by many intermediate anodes in order to provide the correct potential profile for beam formation and focusing, as in the VLEPP klystron design [24]. In this intricate design, the beam is composed of many separate "beamlets" produced by different regions of the cathode and is switched by a non-intercepting control electrode (offering the possibility of pulsing the klystron using a quasi-d.c., high-voltage supply and a low-voltage modulator in series with the grid). In order to achieve very low perveance, a "cluster klystron" is proposed by a BNL–SLAC collaboration [25]; it is a collection of 42 separate beams, each comprising a 40-MW klystron and all sharing a common superconducting solenoid. Finally, the use of lumped shavers in order to control beam interception might be necessary.

Some characteristics of high-power klystron development projects, incorporating (separately) the features mentioned, are given in Table III. Achievements so far, as quoted in a recent review paper [26], are briefly summarized below, together with the reasons for the limitation observed. The SLAC XC klystron achieved 40 MW, 0.8 μs or 72 MW, 0.1 μs pulses, performance being limited by RF breakdown in the double-gap output circuit stimulated by beam interception. In the KEK klystron, the output power was restricted to 22 MW by failure of the gun ceramic insulator. In the VLEPP klystron, 50 MW has been achieved, limited by a substantial beam loss that amounts to 70 out of 200 A. The cluster klystron is still at the design stage.

Table III. High-power klystron projects

Klystron	Wave-length (cm)	Power out (MW)	Pulse length (μs)	Voltage (kV)	Microper-veance ($\mu A/V^{3/2}$)	Frequency (GHz)
SLAC 5045	10.5	67	3.5	350	2.0	2.86
SLAC	10.5	150	1.0	450	2.0	2.86
SLAC XC	2.6	100	1.0	550	1.2	11.4
KEK	2.6	120	1.0	550	1.2	11.4
VLEPP	2.1	150	1.0	1000	0.3	14.2
Cluster	2.6	1680	1.0	400	0.4	11.4

4.2 Drive Linac as High-Frequency Power Source

Owing to difficulties met in the development of microwave klystrons operating in the frequency range of 11–14 GHz, it is unthinkable to use similar tubes to deliver the 30-GHz, ~150 MW/m peak power per unit length required in the CLIC linac structure. The CLIC scheme for generating the necessary power is based on a two-stage accelerator [27]; there is a drive beam that runs parallel to the main beam and ensures power flow to the main linac. The drive linac contains strings of travelling-wave transfer structures, in which short and intense bunches induce the required power that is then fed to the main linac, and sectors of superconducting (SC) cavities supplying energy to the beam when necessary. The use of SC cavities to reboost the drive beam as well as to accelerate it up to its initial energy is dictated by the concern of good extraction efficiency; LEP-type cavities operating at ~ 350 MHz with a gradient of ~ 6 MV/m are considered. The drive-beam energy should be of the order of a few GeV, which implies no longitudinal mixing inside the bunches and no phase slip with respect to the main beam. Owing to the unavoidable intermittent reacceleration, the drive beam has to be arranged in discrete trains of dense bunchlets that all contribute to the build-up of a decelerating field in the transfer structure and are separated by the 30-GHz wavelength. To generate a pulse of length equal to the main structure filling time τ_0 (11.1 ns), four such trains are needed, separated by the 350-MHz wavelength. The total charge needed per train is about 1.65 μC and the number of bunchlets per train depends on the matching of the decelerating field build-up to the SC cavity accelerating gradient. Simply minimizing the deviation of the linear build-up from the sinusoidal acceleration wave limits this number to 11, but the use of voltage harmonics schemes [28] (their sum giving almost a linear function as shown in Fig. 12) might allow for 43 bunchlets (with correspondingly lower charge per bunchlet).

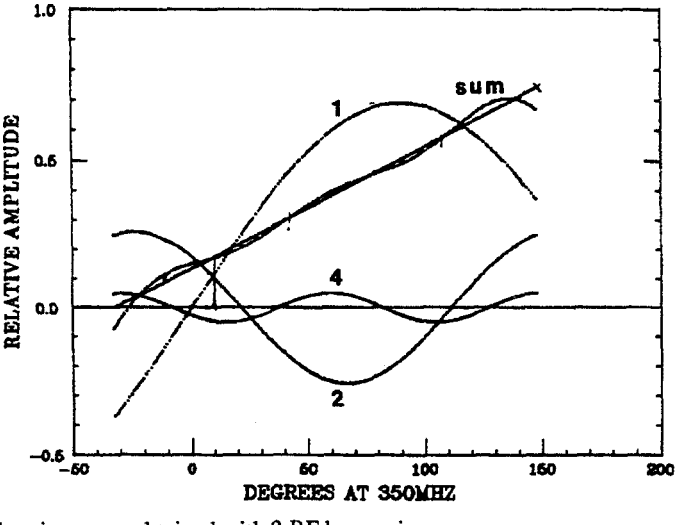

Fig. 12. Accelerating ramp obtained with 3 RF harmonics

This drive-linac conceptual design implies certain challenges which have been addressed, mainly the bunchlet generation, the transfer structure design and the beam dynamics control. Owing to the difficulty of generating short (1 mm r.m.s.) and dense (up to 10^{12} particles) bunchlets, a test facility (CTF) has been built [29]. It includes an RF gun, a beam line acting as magnetic spectrometer, acceleration to 60 MeV at 3 GHz and RF power generation at 30 GHz (using prototype structures). A charge up to 30 nC per beam should be obtained using a laser-driven photocathode, synchronized with the RF. During first tests using a prototype of the main linac structure instead of an actual transfer structure, a 2.7 MW–12 ns power pulse was extracted. The most recent transfer structure design proposed [30], is based on power-collecting rectangular waveguides that run along the outside of the beam pipe (either on each side or above and below) and are coupled to the inside via slits about 0.5 m long (Fig. 13). The phase velocity in the waveguide is adjusted by periodic indentations. Numerical calculations and model work are being carried out in parallel to check the possibility of generating the required flat power pulse and the amplitude of the wake fields. This design must indeed provide the low impedance needed (~ 4.5 Ω/m for R′/Q) and the required decelerating field per bunchlet (~ 65 kV/m for a population of 10^{12}), while minimizing the undesirable wake-field modes that could compromise beam stability. The dynamics of such a beam includes special features: the energy differences between bunchlets are unusually large owing to the increasing decelerating field, the energy spread within each bunchlet is wide since the bunch length is not short with respect to λ_{RF}, and the amplitudes of the synchronous wake fields may be disturbing. The impact of these features on the beam transport and dynamics, in the presence of alignment imperfections, trajectory correction, variable wakes and magnetic focusing, is being investigated by numerical simulations [31]. Figure 14 gives an example of initial- and final-energy distributions in a train of 11 bunches travelling over ~ 3.5 km. Phase plots of the emittances (Fig. 15) show that all the bunchlets remain within the beam-pipe acceptance (circle tangent to the frame), with the assumptions retained for this calculation. Further investigations are needed to check the feasibility of the scheme.

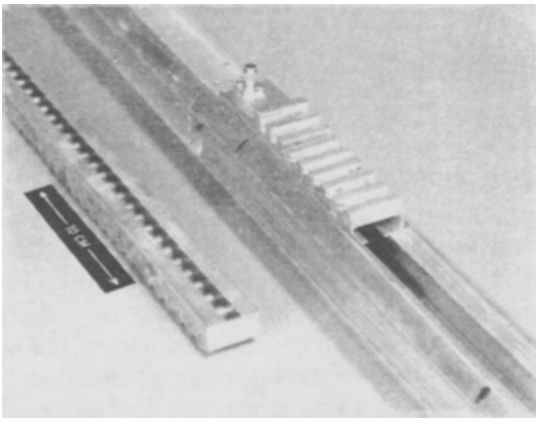

Fig. 13. Model of the most recent transfer structure design

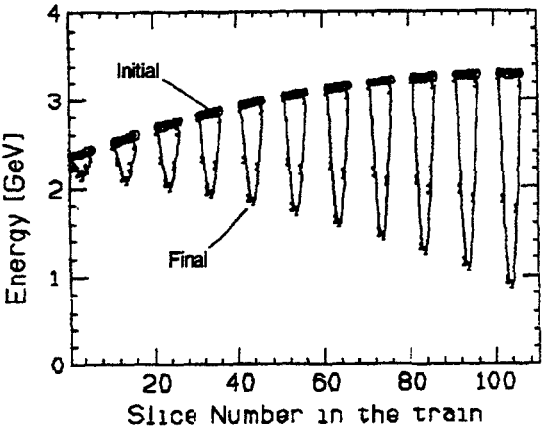

Fig. 14. Initial and final energy distributions in a drive-beam train

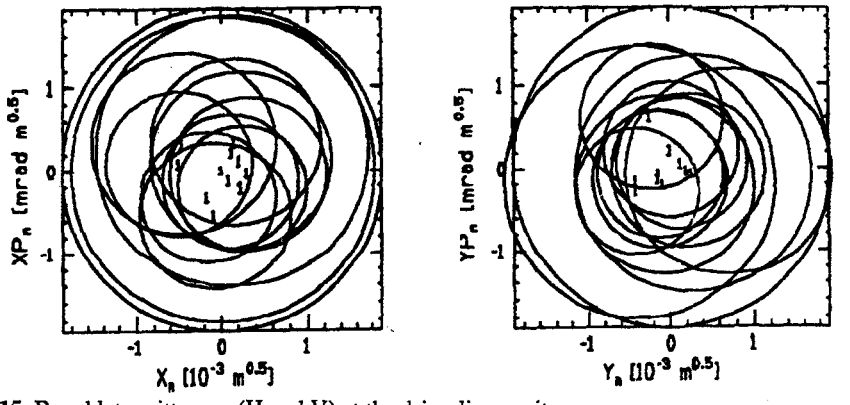

Fig. 15. Bunchlet emittances (H and V) at the drive-linac exit

5 Attempt at Defining Factory Characteristics

5.1 Linear Collider Schematic Layout

A possible layout often envisaged in the Linear Collider studies is sketched in Fig. 16. The injection systems, which we will not describe in this paper in any detail, could for instance be grouped together in a central position. They include two RF guns producing electrons, primary linacs accelerating the particles to an energy of the order of 1.5 to 3 GeV and damping rings running at these energies also. To produce the required positrons, a converter target is required that generates photons then transformed into e^+e^- pairs. A second linac then accelerates the collected positrons to the same energy as the main electrons. For reasons of simplicity, the converter target and second linac are not represented on the side of the e^+ injection that must begin with a primary acceleration of e^- up to the target. The very high rate production of the particles is often a challenge, mainly for the positrons (~ 10^{14} $e^{\pm}$/sec in CLIC, for

instance). Damping rings are imposed, at least for the positrons, in order to reach the tiny transverse phase-space dimensions needed. Longitudinal dimensions required are also small and one or two bunch compressors have to be added at the exit of the damping rings to further reduce the bunch length. The first compressors acting at the ring energies are followed by pre-accelerations to something like 9 to 16 GeV in preliminary linacs (at 3 GHz), in order to limit the energy spread in the second compressors. It has been proposed to put the second compressors at the extremities of a kind of very long trombone (implying transfer lines behind the preliminary linacs) which could be extended by stages the day the final energy of the collider is raised. At the exit of these compressors and the injection into the main linac, the main beams must have the required characteristics that lay in the following interval approximately

$$\sigma_z \cong 0.1\text{–}0.2 \text{ mm}$$
$$\gamma\varepsilon_x \cong 2\text{–}5 \times 10^{-6} \text{ rad} \cdot \text{m}$$
$$\gamma\varepsilon_y \cong 5\text{–}20 \times 10^{-8} \text{ rad} \cdot \text{m}$$

considering X-band colliders based on small spot sizes at the I.P. and small beam power. In this case, the main beams are then accelerated to the final energy through X-band linacs (10–30 GHz) with high gradient (50–100 MV/m). At the end of the linacs, the beams traverse the final-focus system and are focused at the I.P. to dimensions of the order of σ_x = 70–300 nm and σ_y = 3–16 nm. Multibunch operation (above 1 and up to 90 bunches say) might be necessary to reach the desired luminosity and the corresponding beam power would be around 2–4 MW/beam. The experimental areas and the beam dumps could possibly be located in the neighbourhood of the injection systems, as sketched in Fig. 16.

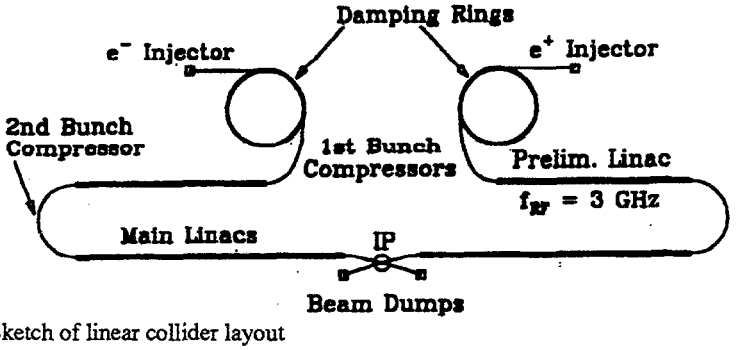

Fig. 16. Sketch of linear collider layout

5.2 Guessed Parameters and Performance for Linear Colliders as Factories

For the purpose of this paper, and to give indications to the reader about what looks possible today, guesses were specifically made on the basis of the CLIC feasibility studies, mainly for the final-focus system [2]. For the estimates of the relevant parameters and the projected performance, three levels of centre-of-mass energy have been retained, in agreement with the interests of the experimental physics:

1) $\sqrt{s} = 100$ GeV Z^0 factory
2) $\sqrt{s} = 250$ GeV Top factory
3) $\sqrt{s} = 500$ GeV Higgs factory

The main linacs would have increasing lengths, using the idea of the trombone, with the same injection and interaction point complex, for instance. Assuming an accelerating gradient of 80 MV/m, the active length of each linac would be between 625 m for case 1 and 3125 m for case 3. Adding approximately 20% for the unavoidable focusing quadrupoles and for the drift spaces would raise this length to 750 m and 3750 m, respectively. Ultimately, adding the space required by the final-focus system (400 to 450 m) would bring the length per "trombone" to something like 1200 m in case 1 and 4200 m in case 3.

Assuming that the properties of the beam at the exit of the linacs are those required, the linear collider performance depends on the adjustment capability of the final-focus system, for the energies considered. We have already mentioned the two main parts of this system, the telescope with a large demagnification ($\tilde{>}$ 50) and the chromatic correction sections with cells containing sextupoles and a phase shift of π. Aberrations and momentum bandwidths limit the maximum β-value acceptable in the quadrupoles, hence the minimum β-value (β^*) at the crossing point. Next, the synchrotron radiation in the last quadrupole before the interaction means that the vertical beam size σ_y^* at the I.P. (smaller than σ_x^* by choice) goes through a minimum even if β_y^* is further reduced (this phenomenon known as the Oide effect limits the benefits of strong focusing). Finally, limiting beamstrahlung energy-spread implies, as we have seen, flat beams and conditions on σ_y^* to reach the desired luminosity. In this framework, investigations were carried out by O. Napoly [32] to provide educated estimates of possible sets of parameters. The starting point was the two normalized emittances we can reasonably expect at the crossing point after considering the control of the emittance growth in the linac and of the luminosity dilution due to beam–beam radiation,

$$\gamma\varepsilon_x \cong 1.5 - 2.0 \times 10^{-6} \text{ rad} \cdot \text{m}.$$
$$\gamma\varepsilon_y \cong 1.5 \times 10^{-7} \text{ rad} \cdot \text{m}$$

Analysing the Oide effect tells us for which values of the β-functions at the I.P. the beam sizes reach a minimum. However, the minimum horizontal value β_x^* that can be considered is fixed in fact by the tolerable chromatic aberrations. Moreover, the control of the average energy spread during collision (associated with beamstrahlung) implies the retention of β_x^*-values depending on the centre-of-mass energy. Therefore, the β_x^* are all different for the energy levels we are interested in and for the energy resolutions δ_E we are aiming at. Then, in the vertical plane, the condition that the β_y^*-value must not be smaller than the bunch length σ_z (0.17 mm in CLIC) imposes its value to ~ 170 μm. Hence, both β^* are above the Oide limit in all cases.

When all these final-focus parameters are selected, it is possible to calculate the nominal luminosity per bunch using the simple formula

$$\mathcal{L}_b^n \sim \frac{f_{rep}\, N^2}{4\pi\, \sigma_x^*\, \sigma_y^*} \tag{18}$$

On the one hand, correct tracking that includes aberrations, synchrotron radiation in quadrupoles, real particle distributions and losses tends to give luminosity values lower than $\mathcal{L}_b^n$ (18), on the other hand, pinch effect associated with the focusing due to fields from the other interacting bunch enhances the luminosity with respect to $\mathcal{L}_b^n$ by a factor of the order of 2 for flat beams. Altogether, the nominal value $\mathcal{L}_b^n$ gives a slightly pessimistic estimate, too low by only 20% say, a difference that can be neglected in our approximate projections.

Using the final-focus criteria quoted above and the equation (18) for a first estimate of the luminosity per bunch, it is possible to make lists of guessed parameters and performance for the three "factories" imagined above (Table IV). For the lower-energy factories (Z^0 and top) physics may require good energy resolution and small energy spread δ_E; therefore the price to pay in luminosity when reducing δ_E is also indicated in Table IV. For the Higgs factory however, there is no point in reducing δ_E below the limit of about 5% coming from the always-present 2γ events.

Table IV. List of guessed parameters at different energies

	Z^0 factory	Top factory	Higgs factory
$\sqrt{s}$ (GeV)	100	250	500
σ_y^* (nm)	16	10	8
σ_x^* (nm)	70	70	90
β_x^* (μm)	327	600	2200
δ_E (%)	4.2	7.3	5.9
$\mathcal{L}_b^n$ ($cm^{-2}s^{-1}$)	4×10^{32}	7×10^{32}	6.8×10^{32}
σ_x^* (nm)	280	300	–
β_x^* (mm)	5.2	11	–
δ_E (%)	0.28	0.55	–
$\mathcal{L}_b^n$ ($cm^{-2}s^{-1}$)	1.1×10^{32}	1.6×10^{32}	–

Speculating about the possibility to reduce by staggered tuning, in multibunch mode, the effects of the transverse wake fields on the follower-bunches of a train in a CLIC-type collider, one could imagine to have up to 11 bunches per RF pulse. These bunches would be separated by 10 RF periods perhaps, last for a maximum of 3.3 ns (compared with 11 ns filling time), and increase proportionally the total luminosity. In the most optimistic case, the total luminosities would range from 1.2 to 4.4×10^{33} $cm^{-2}s^{-1}$ for Z^0-factories, from 1.8 to 7.7×10^{33} for top factories and reach 7.5×10^{33} for so-called Higgs factories. These approximate projections made as an educational exercise show that the performance one can hope to expect is in promising agreement with the physics requirements.

6 Conclusions

The choice of the RF-frequency in a high-gradient normal-conducting linac will probably fall in the 10–30 GHz interval and result from a compromise between the opposing requirements of saving power and minimizing harmful effects. A lot of studies and simulations improved the knowledge of the mechanisms involved in wake field effects and beam stability in the presence of one bunch or several bunches per beam. They revealed a number of promising correction possibilities aiming at low energy-spread and small transverse emittances at the exit of the linac. The question of the tight alignment tolerances required remains challenging, but looks solvable if good alignment strategies are defined. The idea of using microwave quadrupoles for stabilizing the beam in the presence of strong wake fields has been reinforced by numerical simulations. Model work on high-frequency accelerating structures proved that they could be manufactured by industry and recent studies indicate that the risk of dark currents decreases at higher frequency. Microwave klystrons for 250-GeV and 50-MV/m linacs seem feasible in the near future, as do the companion RF-pulse compression systems. Prospects for the peak power requirements at 500 GeV and 100 MV/m are however more distant, since technical limitations have to be overcome. For the CLIC scheme based on a two-stage accelerator, significant progress has been made on the study of the drive-beam dynamics and the design of the transfer structures. In this case, challenging issues are the generation of the required short and dense bunchlets, and the development of efficient transfer structures that are least harmful to the drive beam. Workable experimental conditions seem achievable and designs of final-focus systems have been worked out. Projected performance of linear colliders as factories was estimated for centre-of-mass energies corresponding to the production of Z^0, top and Higgs particles. It gives the encouraging result that the physicists' expectations can possibly be satisfied.

Acknowledgements

The author is particularly grateful to K. Berkelman, C. Fischer, N. Holtkamp, T.L. Lavine, O. Napoly, K.A. Thompson, L. Thorndahl and I. Wilson who have kindly made their most recent results available to him.

References

[1] P. Zerwas, Proc. ECFA Workshop on Linear Colliders, Garmisch-Partenkirchen, 1992.

[2] O. Napoly, CLIC Note 144, CERN, 1991; CLIC Notes 155 and 166, CERN, 1992;
O. Napoly et al., Proc. Part. Acc. Conf., San Fransico, 1991.

[3] K. Berkelman, CLIC Note 154, CERN, 1991; CLIC Notes 164 and 168, CERN, 1992.

[4] R.B. Palmer, SLAC-PUB-4295, 1987;
W. Schnell, SLAC/AP-61, 1987.

[5] K.L.F. Bane, SLAC-AP-76, 1989;
G. Guignard and C. Fischer, Proc. Part. Acc. Conf., San Francisco, 1991, Vol. 5, p. 3231.

[6] V.E. Balakin, Inst. of Nucl. Phys., Novosibirsk, Preprint 88-100, 1988.

[7] H. Henke and W. Schnell, CERN-LEP-RF/86-18, 1986.

[8] W. Schnell and I. Wilson, Proc. Part. Acc. Conf., San Francisco, 1991, Vol. 5, p. 3237.

[9] G. Guignard, CERN SL/91-19 (AP), 1991.

[10] K.A. Thompson, SLAC AAS Note 71, 1992.

[11] R. Ruth, SLAC-PUB-4541, 1988.

[12] R.B. Neal (editor), The Stanford Two-Mile Accelerator (W.A. Benjamin Inc., New York, 1968).

[13] R.B. Palmer, Proc. DPF Summer Study, Snowmass, SLAC-PUB-4542, 1988.

[14] H. Deruyter et al., Proc. Linear Acc. Conf., Albuquerque, 1990, p. 132.

[15] T. Higo et al., Proc. Linear Acc. Conf., Albuquerque, 1990, p. 147;
T. Taniuchi et al., KEK Preprint 91–152, 1991.

[16] K.A. Thompson and J.W. Wang, Proc. Part. Acc. Conf., San Francisco, 1991, Vol. 1, p. 431.

[17] J.W. Wang et al., Proc. Part. Acc. Conf., San Francisco, 1991, Vol. 5, p. 3219.

[18] N. Holtkamp, Private Communication;
T. Weiland (spokesman) et al., DESY 91-153, 1991.

[19] I. Wilson, W. Wuensh, C. Achard, Proc. EPAC 90, Nice, 1990, p. 943;
I. Wilson and W. Wuensh, CERN-SL/90-103 (RFL), 1990.

[20] R. Parody, Private communication, CERN, 1991.

[21] T. Raubenheimer and R. Ruth, SLAC-PUB-5355, 1990;
T. Raubenheimer, SLAC-387, UC-414, 1991.

[22] C. Fischer and G. Guignard, Proc. EPAC 92, Berlin, 1992.

[23] C. Fischer, Private Communication.

[24] L.N. Arapov, V.E. Balakin, Y. Kazakov, Proc. EPAC 92, Berlin, 1992.

[25] R.B. Palmer, W.B. Herrmannsfeldt, K. R. Eppley, Part. Acc. 30, 197–209, 1990.

[26] T.L. Lavine, Proc. EPAC 92, Berlin, 1992.

[27] W. Schnell, CERN-LEP-RF/88-59, 1988.

[28] L. Thorndahl, CLIC Note 152, CERN, 1991.

[29] Y. Baconnier et al., Proc. Linear Acc. Conf., Albuquerque, 1990, p. 733.

[30] G. Carron and L. Thorndahl, Proc. EPAC 92, Berlin, 1992.

[31] G. Guignard, CERN SL/92-22 (AP), 1992.

[32] O. Napoly, Private Communication.