THE NEXT LINEAR COLLIDER*

Ronald D. Ruth
Stanford Linear Accelerator Center

1. Introduction

There is now broad agreement in the high energy physics community that to continue exploring the energy frontier in e^+e^- interactions, we will have to abandon circular colliders and adopt linear colliders. This realization has led to active research throughout the world towards the next generation of linear colliders. The past few years have seen great strides in our understanding of both the accelerator physics and the technology of linear colliders. We are now at the point where we can discuss in fair detail the design of such a "Next Linear Collider," or NLC.[1-5]

The two key design parameters of the NLC are its energy and luminosity. A broad consensus has emerged over the past couple of years that the energy should be 0.5 TeV (total electron plus positron energy), upgradable to at least 1.0 TeV. One reason for this choice of energy range is the great potential of such a collider for significant high-energy physics research in the era of the SSC. Another is that this energy range is a natural next step; it is a factor of 5 to 10 beyond that of the present Stanford Linear Collider (SLC). In order to obtain a sufficient event rate to perform detailed measurements, the luminosity of the collider should increase with the square of its energy. For an NLC in the TeV energy range, a luminosity of $10^{33} - 10^{34}$ $cm^{-2}s^{-1}$ is required.

A factor of 5–10 energy increase can be obtained in two ways: by increasing the collider length to 10–20 times that of the SLC (3 km), or by raising its accelerating field to 5–10 times the SLC gradient (20 MV/m). The present consensus is that we should first increase the accelerating field by about a factor of three to five—up to about 50 to100 MV/m. To limit the RF power required, this field should be provided by structures similar to those used in the SLC but at a higher RF frequency of 10–30 GHz. At SLAC, the frequency choice for the NLC is 11.4 GHz, or four times the present SLC frequency. Of course, the ultimate

* Work supported by the Department of Energy, contract DE-AC03-76SF00515.

tradeoff between length and accelerating field is governed by the overall cost and the upgradeability. A broad optimum occurs at the point where the linear costs (accelerating structure, magnets, tunnel, etc.) equal the cost of the RF power source.

The choice of luminosity range also greatly influences the design of the linear collider. In principle, one could increase the luminosity simply by raising the repetition rate of the accelerator, but the wall-plug power increases in direct proportion. In a reasonable design, the wall-plug power should fall in the range 100–200 MW. Given this constraint, the best way to increase the luminosity is to shrink the beam size at the interaction point (IP). In addition, the beam cross section must be kept flat at the IP in order to minimize the amount of "beamstrahlung" radiation emitted as energetic electrons or positrons interact with the electromagnetic field of the opposing bunch.

The luminosity can be further enhanced by accelerating several bunches on each machine cycle. A single bunch of particles can, in practice, extract only a few percent of the energy available in the accelerating structure. With additional bunches, we get both greater luminosity and higher efficiency of energy transfer to the beam. The number of particles in each bunch, another factor that directly affects the luminosity, is limited by the RF energy that can be stored in the accelerating structure and by the amount of beamstrahlung radiation that can be tolerated. The obvious solution is to generate trains of successive bunches, each with a fairly moderate number of particles.

Given these goals and constraints, we can now sketch a rough design of a linear collider able to achieve both the desired energy and luminosity. A possible layout is shown in Fig. 1. There are two complete linear accelerators, one for electrons and the other for positrons. Each linac is supplied with particle beams by a damping ring followed by a preacceleration section consisting of two bunch compressors and a 16 GeV linac. After passing through the main linacs and final focus system, the beams collide at a small crossing angle inside a large particle detector like the SLD.

To illustrate the basic features of the NLC operation, let's follow some electron bunches through the collider. A sequence of 10 bunches or so is created at the source and accelerated up to about 1.8 GeV in a preaccelerator. This "batch" of bunches is then injected into a damping ring that serves to reduce the transverse

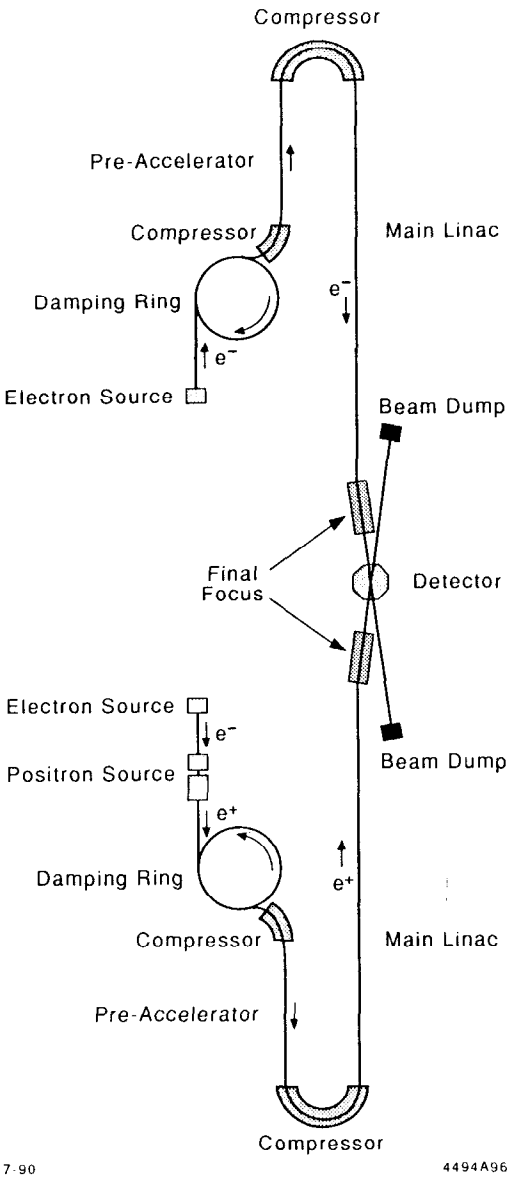

Fig. 1. Schematic diagram of the NLC.

and longitudinal phase space occupied by the electrons in each bunch. At the proper moment, these bunches are extracted from the ring and then compressed along their direction of motion by a bunch compressor, after which they are accelerated up to about 16 GeV and compressed a second time just prior to injection into the main, high-gradient, linac. The entire batch is carefully steered and focused as the electrons are accelerated up to full energy in the linac. Precision magnets in the final focus system squeeze the bunches down by about a factor of 300 just before they collide at the IP with similar bunches of positrons. Except for the fact that they were created differently, from the shower of particles that occurs when a bunch of electrons hits a metal target, these high-energy positron bunches have followed a similar evolution. After the beams collide, their debris is channeled out of the detector area and into shielded dumps.

1.1 NLC Parameter Options

The parameters for an NLC are not definite yet; however, over the past few years we have narrowed down the range of possibilities considerably. At SLAC, we have recently reviewed several options for an NLC which has an initial energy of 0.5 TeV in the CM and an upgraded energy of 1.0–1.5 TeV. Table 1 lists three parameter options: the first two columns are for 0.5 TeV in the CM, while the final column is for 1.0 TeV. In Option 1, a short linear collider is constructed with the full acceleration gradient of 100 MV/m. This can be upgraded to Option 3 by doubling the length of the linac while keeping the injection system fixed. In Option 2, a long linear collider is constructed with a reduced acceleration gradient of 50 MV/m. This can be upgraded to Option 3 by the addition of power sources to the linac. In both upgrade paths, the final focus must be modified somewhat.

Option 1 is quite short and may be less expensive than Option 2, but we are required to face all the problems of the high acceleration gradient and the required high peak power RF sources. In Option 2, we relax the requirements for RF power by a factor of four and begin with a reduced acceleration gradient. The price is an initially longer accelerator with the increased conventional construction.

We have found in the design process that it is very important to realize that the intensity and emittance at the final focus are quite different than those in the damping ring. To model this, the intensity has been allowed to decrease as shown in Table 1. In addition, the emittance at the final focus is assumed to be diluted by about 65%.

Table 1. NLC Parameter Options.

Option	1	2	3
Energy	$\frac{1}{4}+\frac{1}{4}$ TeV	$\frac{1}{4}+\frac{1}{4}$ TeV	$\frac{1}{2}+\frac{1}{2}$ TeV
Luminosity	2×10^{33}	2×10^{33}	1×10^{34}
Linac Length	7 km	14 km	14 km
Accel. Gradient	100 MV/m	50 MV/m	100 MV/m
RF Frequency	11.4 GHz	11.4 GHz	11.4 GHz
# Particles/bunch:DR	2×10^{10}	1×10^{10}	2×10^{10}
Linac	1.8×10^{10}	9×10^{9}	1.8×10^{10}
FF	1.5×10^{10}	7×10^{9}	1.5×10^{10}
# Bunches, n_b	10	10	10
Repetition Freq.	120 Hz	180 Hz	180 Hz
Wall Plug Power	66 MW	50 MW	200 MW
IP Beam Size: σ_y	4 nm	4 nm	2.5 nm
σ_x	320 nm	200 nm	220 nm
σ_z	100 μm	100 μm	100 μm

To discuss the NLC in more detail, we divide the problem into the two primary parameters: the energy and the luminosity. In the next section we discuss how to obtain the energy in an NLC.

2. The Energy

As discussed in the Introduction, the energy of the linear collider is obtained through a combination of length and acceleration gradient,

$$E = \mathcal{E}_z L \ , \tag{1}$$

where $\mathcal{E}_z$ is the acceleration gradient and L is the length of the linac. This relation is over-simplified, for reasons which we discuss later, in that the average

acceleration gradient may differ from the peak and the length may include space for focusing magnets, etc.

The acceleration is obtained with the use of radio frequency (RF) structures as shown in Fig. 2. The structure shown is a travelling wave structure. It is basically a long copper cylinder periodically interrupted by disks with holes along the center line. Every so often (every 1.5 m or so), the structure is interrupted by a feed for fresh RF power and a load to remove the depleted upstream power.

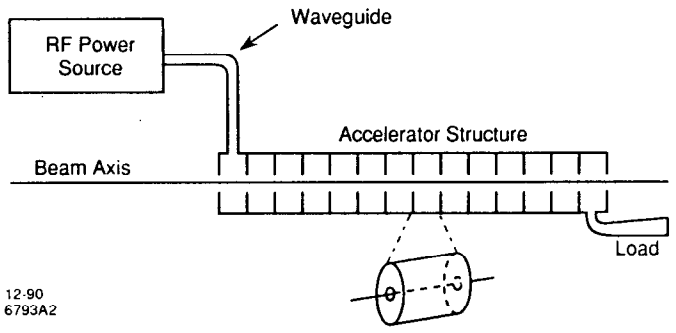

Fig. 2. Diagram of a travelling wave RF structure.

The RF power is provided by the RF source and is transported to the accelerating structure in a copper waveguide. The structure is designed as a travelling wave structure and, as such, has a characteristic phase velocity and group velocity. The phase velocity is designed to be the speed of light. In this way, if a relativistic electron enters the structure at the correct phase for acceleration, then it will be continually accelerated as it traverses the entire structure. The group velocity v_g is the rate at which the structure is filled with energy; it is the velocity of the envelope of the RF pulse as it traverses the travelling wave structure. If L_s is the length of the structure, then the time to fill the structure T_f is given by

$$T_f = \frac{L_s}{v_g} . \tag{2}$$

For cases of interest, the group velocity is somewhat less than one tenth of the velocity of light.

2.1 The Extraction of Energy

The energy gain of a test particle at the head of a bunch traversing a structure of length L_s is

$$\Delta E = \mathcal{E}_z L_s \cos\varphi , \tag{3}$$

where φ is the phase on the RF wave. The trailing particles see not only the field supplied by external sources, but also the field induced by the bunch itself, the longitudinal wakefield. The bunch charge induces fields in all the synchronous modes of oscillation of the accelerating structure. The field induced in the fundamental accelerating mode is

$$\mathcal{E}_{wake} = -2k\,q\cos(\omega z/c)e^{-\lambda z} , \tag{4}$$

where k is called the loss parameter, q is the charge in a bunch which is short compared to the wavelength of the RF, ω is the RF frequency, z is the distance behind the point-like bunch, and λ is the decay constant due to losses in the structure walls. The wakefield ahead of a speed-of-light bunch vanishes due to causality. The total wakefield induced consists of a sum over all the synchronous modes of the structure; the dominant term is given in Eq. (4).

This field induced by the particle bunch causes problems which must be dealt with. The particles at the head of the bunch feel the full acceleration, while those at the tail feel an accelerating field reduced by $2kq$. If the particles are being accelerated on the crest of the RF, this causes a reduction of the average energy gained by the bunch, and it also causes an energy spread:

$$\begin{aligned} \Delta E_{ave} &= (\mathcal{E}_z - kq)L_s \\ (\Delta E)_{spread} &= \pm kqL_s . \end{aligned} \tag{5}$$

These effects are due to the extraction of energy from the RF wave. The field induced by the bunch reduces the electric field in the structure an amount which

corresponds to the energy extracted by the bunch of particles. The fraction of energy extracted from a full structure by a bunch at the crest of the RF is

$$\eta_0 = 1 - \frac{(\mathcal{E}_z - 2kq)^2}{\mathcal{E}_z^2} \simeq \frac{4kq}{\mathcal{E}_z} \quad \text{for small } \eta_0 \ . \tag{6}$$

The reduction of the average value of the energy gain can be compensated either by increasing the accelerating field $\mathcal{E}_z$ or by adding more accelerator sections to make up the lost energy. The spread of energy in a bunch can be compensated for by shifting the phase of the bunch on the RF wave. In this way, with very little loss of acceleration, it is possible to obtain a slope sufficient to cancel the variation induced by the bunch wakefield. For a uniform particle distribution with a full width $\Delta\varphi$, the phase offset is

$$\sin\varphi_0 = \frac{2kq}{\Delta\varphi\mathcal{E}_z} \ . \tag{7}$$

Provided that the phase offset is small, this compensation technique works well. In order to achieve a small phase offset, the bunch can be lengthened or the intensity reduced. With very long bunches, the curvature of the RF must be taken into account and can, in fact, be used to provide compensation of the nonlinear variation of energy along the bunch. With short bunches, only the linear variation can be cancelled, which leaves a residual nonlinear energy variation. This residual must be kept smaller than the energy acceptance of the final focus system.

2.2 Multi-Bunch Energy Compensation[6]

In the Introduction, we discussed the acceleration of a short train of bunches in order to improve the luminosity by extracting more energy from the RF structure. From the analysis of the previous section, the second bunch must have an energy which is lower by

$$\Delta E_2 = -2kqL_s \ . \tag{8}$$

Once again, this is simply due to the extraction of energy from the RF wave.

This problem can be solved by changing the effective structure length for the two bunches. If the structure is partially filled when the first bunch passes through, and if the additional energy entering the structure prior to the passage of the second bunch matches the energy extracted by the first bunch, then the second bunch will have the same energy as the first. This is illustrated in Fig. 3; the cross-hatched areas of the external field and wakefield must match in order to compensate the energy difference. This technique is used at the SLC to adjust the relative energy of the positrons and electrons.

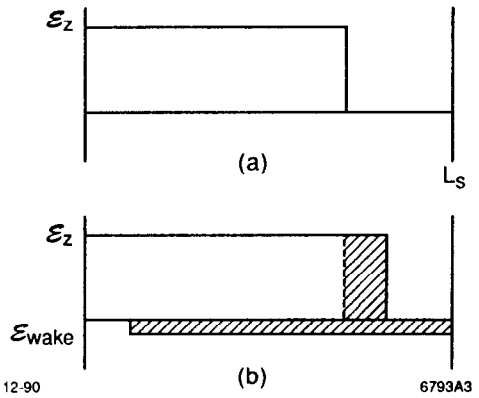

Fig. 3. The electric field profile in the structure (a) just before passage of bunch one and (b) just before passage of bunch two.

2.3 Multibunch Beam Breakup

Let us assume that we can match the energy of a short train of bunches as described in the previous section. There are still other problems caused by the wakefield induced by a bunch of particles. If the bunch is offset in the structure, then it induces a deflecting force behind it which is proportional to the offset of the bunch. This transverse wakefield is similar in form to the longitudinal and consists of the sum of many modes which cause deflection,

$$W_\perp(z) = \sum_n W_n \sin(\omega_n z/c) e^{-\lambda_n z} \ , \tag{9}$$

where W_n is the strength of a particular mode, z is the distance behind a short

bunch, ω_n is the is the mode frequency and λ_n is the decay constant for mode n. The transverse wakefield differs from the longitudinal in that the transverse is sine-like while the longitudinal is cosine-like.

The transverse wakefield can cause an instability known as beam breakup.[7,8] It is caused by the combination of the coupling from bunch to bunch and also by a resonance effect. To see this, consider just two bunches. If these two are offset coming into the structure, the magnet focusing causes them to oscillate with a wavelength $2\pi\beta$. The first bunch oscillates freely down the linac according to

$$\frac{d^2x_1}{ds^2} + \frac{x_1}{\beta^2} = 0 \,. \tag{10}$$

The second bunch is also focused but, in addition, is deflected by the leading bunch's wakefield

$$\frac{d^2x_2}{ds^2} + \frac{x_2}{\beta^2} = \frac{Ne^2W_\perp(l)}{E}\, x_1 \,, \tag{11}$$

where N is the number of particles in bunch one and E is the energy of the bunches. The deflecting force is proportional to the position of bunch one. Because bunch one oscillates in the focusing system, the force on the right hand side of Eq. (11) oscillates and bunch two is driven at resonance. Therefore, the amplitude of the second bunch grows linearly down the accelerator.

The effect is similar for many bunches: bunch three is driven on resonance by bunch one and bunch two and so on. The result is that the bunches at the end of the train can reach large amplitudes unless something is done to ameliorate the problem. The solution is to eliminate, to the extent possible, the force coupling the bunches together. This can be done by a special design of the RF structure and is the subject of the next section.

2.4 Accelerating Structures

As discussed earlier, the job of the RF structure is to accelerate the beam. As such, it is usually optimized to provide the greatest acceleration for the lowest RF power. In addition, the design can have a large impact on the stability of a single bunch (to be discussed in Section 3.6.2) and on the stability of a train of bunches. To assure the stability of a train of bunches, we would like to reduce the deflecting field induced by a bunch as much as possible before the passage of the next bunch.

This can be accomplished in two ways (see Fig. 4). In the first method, shown in Fig. 4(a), the cavity design is altered so that the deflecting fields are strongly coupled to external waveguides. After a bunch passage, the fields in the cavity die out quickly as they are propagated out the waveguide into a matched load. The design shown in Fig. 4(a) shows radial waveguides coupled via slots cut in the irises of the accelerator.[9]

The second technique, shown in Fig. 4(b) relies on the cancellation of the deflections from cell to cell. If the cells in a single short structure are designed so that the deflecting modes oscillate at different frequencies, then the average deflection over the structure effectively damps due to the decoherence of the various cell wakefields. The initial decoherence time is just the inverse of the spread in frequency. This technique is illustrated in Fig. 4(b), where the change in frequency is accomplished with three radial slots of varying depth cut into the irises of the structure.

Damped structures similar to that shown in Fig. 4(a) have been constructed at SLAC and have achieved a quality factor Q $\sim$ 8 for the dominant higher-order mode.[10,11] This damping is completely sufficient to eliminate the beam breakup discussed in the previous section.[7,8]

The second technique of detuning is an alternative, and possibly simpler, technique which is presently under investigation at SLAC.

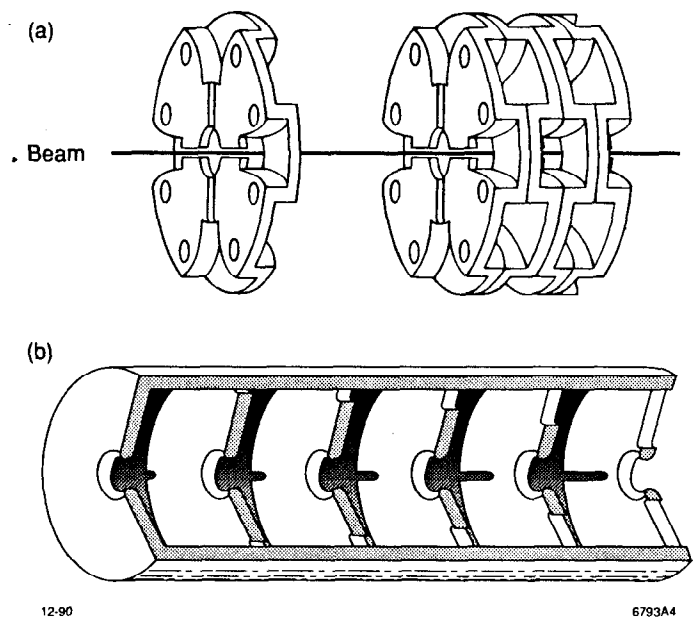

Fig. 4. Two methods of wakefield damping: (a) radial waveguides transmit the energy out of the structure; (b) variation of cell construction causes decoherence and effective damping.

2.5 RF Power Sources

To achieve the desired acceleration gradient for the Next Linear Collider, RF power sources must be provided which give the required peak power and pulse length at the desired frequency. In the designs in Table 1 we find acceleration gradients of 50–100 MV/m at an RF frequency of 11.4 GHz. To achieve the larger gradient, it is necessary to provide about 350 MW of peak RF power in a pulse of about 100 ns in length to be fed into a structure about 1.5 m in length.

In the SLC, the acceleration is accomplished with 2.8 GHz accelerating structures 3 m in length. Each of these is fed by a 40 MW pulse about 1 μs long which yields an acceleration gradient of about 20 MV/m. In order to increase the acceleration gradient in such a structure to 100 MV/m, it would be necessary to increase the RF peak power and the stored energy by a factor of 25.

At the higher frequency of 11.4 GHZ, the energy density must once again increase by a factor of 25; however, the cross-sectional area drops by a factor of 1/16. Thus, the energy per unit length in the RF structure only changes by a factor of two, provided the higher frequency is utilized. Although the energy is changed very little, the structure of the RF pulse is very different from that at the SLC. The necessary RF pulses are higher in peak power, but shorter in duration and feed a shorter accelerator structure. The primary challenge for the RF power system is to provide a source with the characteristics described above.

There are basically two approaches to this problem as outlined in Fig. 5. The first approach, shown in Fig. 5(a), uses RF pulse compression. With this technique, a modulated power pulse of $\sim$ 1 μs is provided by a conventional pulsed power transformer, a modulator. This pulse is converted to an RF pulse of the same duration by some device such as a klystron. After the RF is created, it can be compressed by RF pulse compression to the desired pulse length with a correspondingly higher peak power.

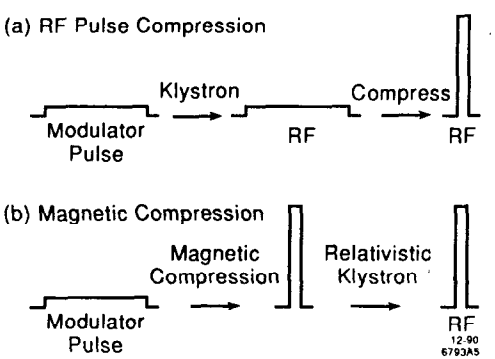

Fig. 5. Two methods of producing short high peak power RF pulses.

The second technique, magnetic compression, begins with the same modulated power pulse and then compresses this pulse using a technique called magnetic pulse compression; the time structure is achieved before the creation of RF. After this stage, RF can be created by a device such as a relativistic klystron or by an array of multiple power sources driven in parallel. This second technique has been

under experimental investigation by a collaboration of SLAC, LLNL, and LBL. The relativistic klystron achieved a power output of 330 MW with a 20 ns pulse.[12] This technique, however, is presently not considered a candidate for the power source due to inefficiency and cost. The remainder of this section is devoted to the first alternative.

2.5.1 *The Klystron*

A block diagram of the RF power system with RF pulse compression is shown in Fig. 6. The modulator power supply is conventional and is similar in most respects to those used at SLAC for the SLC. Therefore, we will begin the discussion with the klystron.

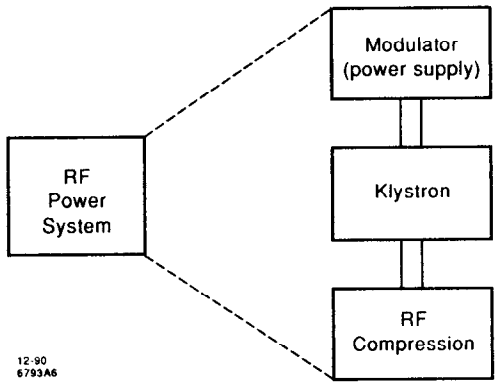

Fig. 6. Block diagram of RF power system with RF pulse compression.

A schematic diagram of a klystron is shown in Fig. 7. Put very simply, a klystron is a narrow-band, high-gain radio frequency amplifier. To achieve this amplification, an electron beam is created by the voltage induced by the modulator across the cathode and anode. The electrons are accelerated to an energy of about 400 KV with a current of about 500 A; they are transported down a narrow tube with a solenoid magnet providing the focusing. A small amount of RF power applied to the input cavity ($\lesssim$ 1 kW) modulates the beam energy at the RF frequency. Due to the induced velocity difference, the faster electrons catch up to

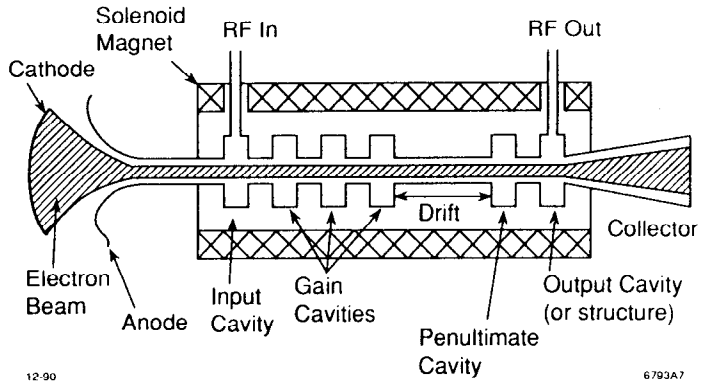

Fig. 7. Schematic diagram of a klystron.

those that were decelerated. This creates a small density modulation of the beam as it enters the first gain cavity.

This cavity is resonantly excited by the RF electric field of the modulated beam to a field of approximately 10 times that in the input cavity. This field acting back on the beam provides much deeper bunching by the time the beam reaches the second gain cavity. This process continues until the final gain cavity, where the energy of the beam is modulated by as much as 30% of its value. This modulated beam bunches strongly in the final drift region, is further compressed by the penultimate cavity and then enters the output structure. This may be one or more standing wave cavities, or it may be a travelling wave structure. The beam induces a field in the output structure; however, this structure is designed so that the phase of the RF field is such as to decelerate the sequence of bunches entering it. This deceleration extracts the RF energy in the bunches and transfers it to the cavity fields which are coupled to an external waveguide. The RF power flows out the waveguide and can be transported for further use. The beam is deposited in a water-cooled collector with approximately one half of its energy removed by this process. With the parameters given above, the klystron produces about 100 MW of RF power.

Klystrons similar in all respects to the one just described have been built at SLAC and have achieved 75 MW in short pulses and 50 MW in long pulses as of this writing.[13] The design goal is to achieve a 100 MW klystron at 11.4 GHz

with a pulse length of about 1 μs. This pulse length is much too long to be used directly in the acceleration process; we need RF pulse compression.

2.5.2 *RF Pulse Compression*

The object of RF pulse compression is to convert a long RF pulse of moderate power into a short RF pulse with high power. Ideally, a factor of five decrease in pulse length could yield a factor of five increase in peak power. Due to inefficiencies, the factor is always somewhat less. The RF pulse compression system SLED (SLAC Energy Doubler) is presently used at SLAC to boost the klystron power by about a factor of three before powering the SLAC linac. This system uses storage cavities to allow the RF to build up. A phase switch from the klystron effectively releases the energy. Unfortunately, this system gives a pulse shape which is sharply spiked due to the exponential decay of the fields in the storage cavities. For an NLC it is useful to have a flat-top pulse to control multibunch energy spread.

This flat-top pulse can be obtained by two different methods. The first method, called binary pulse compression (BPC), uses delay lines to delay the leading portion of an RF pulse so that it is coincident in time with the trailing portion. This yields an RF pulse which is one half as long, but with nearly twice the power. This process can be repeated in a sequence to achieve more and more multiplication. Due to losses in components and waveguides, the method is limited to about three compressions.

Figure 8 shows a schematic diagram of a two-stage BPC system which was constructed at SLAC.[14–16] The 3db hybrid shown in Fig. 8 is a four-port device which combines two power inputs into one or another output port depending upon relative phase. In this way phase shifts can be used as high power RF switches. A three-stage system of analogous design has been constructed at SLAC and has achieved a multiplication factor of 5.5 while reducing the pulse length by a factor of eight.[17] This system, together with two 100-MW klystrons, would produce RF power sufficient for a 5-m long accelerator with an acceleration gradient of about 100 MV/m. High-power tests of this three-stage system are continuing.

One disadvantage of the BPC method of pulse compression is that it uses rather long delay lines. The waveguides which are used have a group velocity very close to the speed of light, and they are only used once as transmissive delay lines.

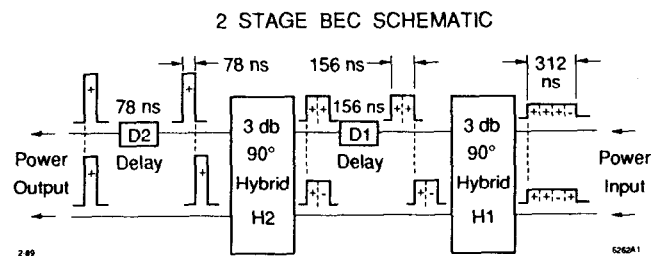

Fig. 8. Schematic diagram of a BPC system.

This problem has led to the development of a new pulse compression scheme called SLED II.[18] The system as shown in Fig. 9 is similar to the SLED system at SLAC except that the cavities for storing the RF are replaced by resonant delay lines. Each of these delay lines has a round-trip delay time equal to the output pulse length. A resonant buildup of energy stored in the lines takes place during an input pulse length which is an integral number of delay periods, typically in the range of four to eight. A phase reversal of the input pulse effectively releases the stored energy to produce a flat-top output pulse during the final delay period. An example of a SLED II pulse compression by a factor of four is shown in Fig. 10. Measurements from a low-power SLED II system with a power gain of four have shown excellent agreement with theory.

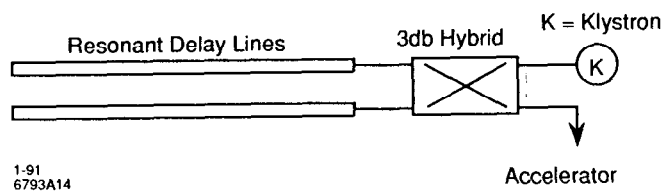

Fig. 9. A block diagram of SLED II.

Comparing SLED II with BPC, the amount of waveguide delay line to achieve a similar compression is reduced by more than a factor of five. This is due to the reflective nature of the scheme; the delay lines are used repeatedly as the RF wave builds up. In addition, this method can be staged by placing the SLED II systems

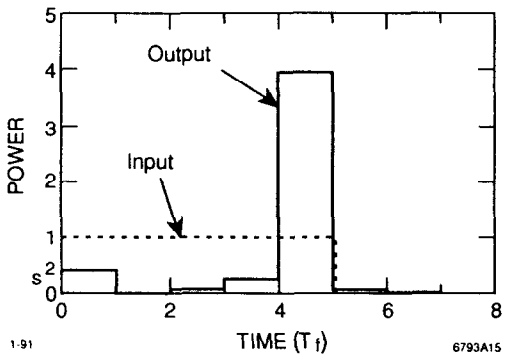

Fig. 10. SLED II pulse compression.

in series to provide even larger compressions if necessary. A high-power SLED II pulse compression system will be constructed at SLAC in 1991 to investigate this promising technique further.

3. The Luminosity

In the first two sections, we discussed the basic layout and how to obtain the energy in a linear collider; in this section, we discuss how to obtain the luminosity. Although the luminosity depends upon beam properties at the interaction point, those properties depend upon beam dynamics throughout the entire linear collider; therefore, we must trace this influence throughout the collider. Before doing that, however, let's examine the luminosity formula. Luminosity for a linear collider is the same as for a circular collider except that there is an additional term, an enhancement factor due to the mutual pinching of the beams. The luminosity is given by

$$\mathcal{L} = \frac{N_+ N_- f_{rep} n_b}{4\pi\sigma_x\sigma_y} H_D \ , \qquad (12)$$

where $N_\pm$ is the number of positrons/electrons per bunch, f_{rep} is the repetition frequency, n_b is the number of bunches accelerated on each cycle of the accelerator, H_D is the pinch enhancement factor, and finally, σ_x and σ_y are the rms beam size of the gaussian spot at the interaction point. Each bunch is assumed to collide with only one other bunch in the opposing bunch train.

The object is to increase the luminosity to 10^{33}–10^{34} cm^{-2}s^{-1}, for the energy range $\frac{1}{2}$ to 1 TeV. To do that we must increase the numerator of Eq. (12) and decrease the denominator as much as possible. For the numerator, we have at our disposal the number of particles per bunch, the repetition rate, and the number of bunches on each cycle, but we must satisfy the constraint that the wall-plug power is in the range of 100–200 MW. For the denominator, we can decrease the cross-sectional area by decreasing σ_x and σ_y, but to do this we must keep the beam flat to control beamstrahlung.

In the next few sections we discuss each term in the luminosity formula. The discussion of beam size is subdivided into several sections. In the next section we begin with the numerator of Eq. (12).

3.1 Intensity and Repetition Rate

First let's discuss the single bunch intensity $N_\pm$ and the repetition rate f_{rep}. From conservation of energy, we must have

$$eNE_{cm}f_{rep} = \eta_{rf}\eta_b P_{wall} \ , \qquad (13)$$

where η_{rf} is the efficiency for converting wall-plug power to RF power, η_b is the fraction of the energy extracted by a single bunch, and P_{wall} is the total wall-plug power supplied to the linacs. The wall-plug-to-RF efficiency, η_{rf}, is about 20% for the projected RF system. This is a fairly realistic estimate including all of the factors in the power system which were discussed in the first section. There are new ideas which could raise this to perhaps 30–40%, however, with the system shown in Section 2.5, η_{rf} is about 20%.

For somewhat different reasons, the single-bunch extraction efficiency is limited to about 2%. In Section 2.1, we discussed the single-bunch energy spread which is induced by longitudinal wakefields. Although the linear part can be compensated by shifting the RF phase to obtain a linear slope, the higher order effects are difficult to compensate. This limits single bunch energy extraction to a few percent.

For the purpose of this discussion, let's select a wall-plug power of 150 MW for an $E_{cm} = 1$ TeV.

Because the required bunches have a very small transverse dimension, it is necessary to control their offset pulse-to-pulse with a feedback system. In order for this feedback system to work efficiently, the sample rate must be at least six times the rate at which the beam centroid is changing. Because ground motion is an important source of bunch motion, and because the spectrum drops off rapidly above 10 Hz, the repetition rate of the accelerator must be greater than 60 Hz. In order to have it sufficiently greater, we set the repetition frequency to 180 Hz. It could be dropped as low as 120 Hz; however, 60 Hz is probably too low. Substituting the previous parameter values in Eq. (13), we find that the maximum number of particles per bunch is $N_{\pm} \simeq 2 \times 10^{10}$.

3.2 The Number of Bunches

As discussed in the Introduction, the designs for the NLC include the acceleration of many bunches on each cycle of the collider. The purpose of this is, of course, to increase the luminosity linearly with increasing number of bunches. If there were no constraints, the largest luminosity would be obtained by placing all the charge in the bunch train into one bunch because in this case there is quadratic gain with increasing intensity. As discussed in the previous section, the single bunch intensity is limited by the amount of energy it can extract while retaining a small relative energy spread. It turns out that this intensity is also consistent with transverse stability (Section 3.6.2) and with beam-beam effects (Section 3.7.1). Thus, the quadratic gain is stopped by these bounds; however, since there is about 98% of the energy left in the structure, it is possible to continue to gain linearly by increasing the number of bunches.

A large number of bunches brings along a host of other complications. Some of these were discussed earlier. The bunches must be stable transversely which means that the structure must be designed in a special way (Sections 2.3 and 2.4). The energy spread bunch-to-bunch must be controlled. Although the solution presented in Section 2.2 does keep the energy spread small, only about 20% of the energy can be extracted without introducing more complicated compensation techniques. This limits the number of bunches to about 10; although the single bunch intensity can be traded off somewhat with the number of bunches.

The RF pulse must be of rather high quality. Systematic phase and amplitude variations over the bunch train must be less than about 2% (such tolerances are not unrealistic with the power sources discussed). Because a significant fraction of the fields felt by the trailing bunches are due to the leading bunches, the intensity of the bunch train must be controlled with a precision less than 2%. The damping rings which produce these trains of bunches must be able to accelerate them without instability. If small position or energy changes occur, a compensation system must be developed to assure that the bunches enter the final focus system on the same trajectory and with the same energy. The final focus system must be designed so that the distant crossings of bunches do not disrupt the primary collisions at the interaction point.

Although the addition of many bunches appears to be "free" in that we simply use energy that would normally be wasted, it introduces complexity into every subsystem of the entire collider. The benefit is an order of magnitude increase in the luminosity.

3.3 The Beam Size

The transverse size of a beam in an accelerator is determined by two basic parameters: the emittance ϵ and the beta function β,

$$\sigma = \sqrt{\epsilon\beta} \, . \tag{14}$$

The emittance is a parameter that is proportional to the area occupied by the beam distribution in transverse phase space (x, p_x). It is defined by

$$\epsilon_x = \frac{1}{p_0}[< x^2 >< p_x^2 > - < xp_x >^2]^{\frac{1}{2}} \, , \tag{15}$$

where x is the transverse position, p_x is the corresponding transverse momentum, and p_0 is the central momentum of the bunch of particles. The angle brackets in Eq. (15) indicate an average over the distribution of particles in a bunch. Because the quantity in the square brackets is an adiabatic invariant (in the absence of synchrotron radiation), the emittance decreases inversely with the momentum of the bunch in a linear accelerator.

The longitudinal emittance is defined in a similar way,

$$\epsilon_z = \frac{1}{p_0}[< z^2 >< \Delta p^2 > - < z\Delta p >^2]^{\frac{1}{2}} , \qquad (16)$$

where z is the longitudinal deviation from a central position within the bunch, and Δp is the deviation of the particle momentum from a central momentum. Once again, the quantity in the square brackets is an adiabatic invariant, which causes ϵ_z to decrease inversely with the beam momentum in a linear accelerator. In the special case of a high-energy electron linac, the longitudinal distribution and the bunch length are fixed because the particles all travel at essentially the speed of light. In this case, the fractional momentum spread varies inversely with the beam momentum.

The beta function β was first introduced by Courant and Snyder in their description of the alternating gradient focusing of particle beams.[19] The parameter not only determines the particle beam size through Eq. (14), it also determines the instantaneous wavelength of the oscillations of particles within the beam envelope as they traverse the focusing magnets (wavelength = $2\pi\beta$).

The beta function also plays an important role at the interaction point. In a magnet-free region, it has the particularly simple form

$$\beta(s) = \beta^* + \frac{(s - s_0)^2}{\beta^*} , \qquad (17)$$

where β^* is the minimum value of $\beta(s)$ and s_0 is the location of that minimum, the IP in this case. According to Eq. (14), the beam size near the interaction point is therefore

$$\sigma^2(s) = \epsilon\beta^* + \frac{\epsilon}{\beta^*}(s - s_0)^2 . \qquad (18)$$

From this form, it is obvious that β^* is the depth of focus because the beam size increases by $\sqrt{2}$ when $s - s_0 = \beta^*$. Thus, the beta function plays two important roles at the IP—it determines both the spot size and depth of focus.

3.4 The Damping Ring[20,21]

The damping ring serves to reduce the emittance of the bunches of particles in all three degrees of freedom. It is an electron storage ring similar in all essential features to the storage rings used for colliding beams or synchrotron light production. The particles in an electron storage ring radiate a substantial fraction of their energy on each turn—energy that is restored by RF accelerating cavities. In the process of radiation, the particles lose energy from all three degrees of freedom, but it is restored only along one, the direction of motion; the proper amount is supplied at a single RF phase for a particle with the design energy, which leads to damping in all three dimensions. The fact that radiation is emitted as discrete quanta, however, introduces stochastic noise that causes diffusion of particle trajectories.

The competition between these damping and diffusion effects leads to an equilibrium value for the emittance of an electron storage ring. Damping rings are designed to enhance the damping effects using strong magnetic fields (such as those in wiggler magnets), while limiting the diffusion by the special design of the transverse focusing in the ring. In addition, there is a unique feature of electron storage rings that can be used to advantage. Due to the lack of vertical bending, the vertical emittance of the beam is much smaller than the horizontal—typically two orders of magnitude smaller. Such naturally flat beams are a key feature of many NLC designs.

One possible design for a future damping ring is about a factor of five larger and operates at an energy 50 percent higher than that of the SLC damping rings (see Fig. 11). The final emittance of the beam is more than an order of magnitude smaller than that of the SLC beams, which leads to much smaller sizes. In fact, the vertical extent of a beam emerging from this damping ring would be a few microns, or about equal to the final spot size at the SLC interaction point.

Another key difference is the simultaneous damping of many batches of bunches. In the SLC, at most two bunches are damped simultaneously, whereas this NLC ring will damp 10 batches of 10 bunches all at once. This feature allows a longer damping time for any given bunch, because we can extract the "oldest" batch and inject a new "young" batch while leaving those in their "adolescence" to continue damping undisturbed.

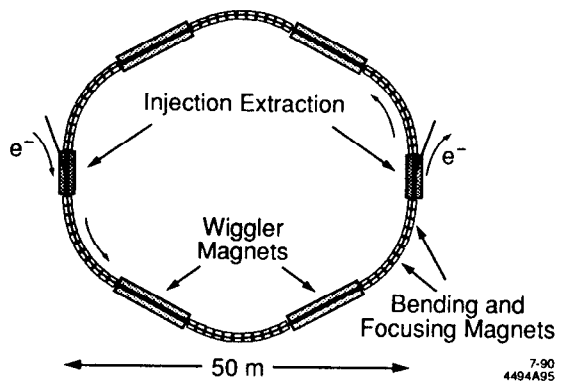

Fig. 11. A design of an NLC damping ring.

Because the bunches forget their origins in the damping ring, their conditions upon emerging are entirely determined by their behavior in the damping ring. This places special emphasis on the stability of the magnets in the damping ring and extraction system.

3.5 Bunch Compression/Pre-Acceleration[22,23]

Although the longitudinal emittance obtained in the damping ring is small enough, the bunch is still much too long for acceleration in a linac. In the SLC and NLC, this problem is solved by a technique called bunch compression, which shortens the bunch while increasing its energy spread. Each bunch passes through an RF accelerating structure phased so that the trailing particles emerge with lower energy than the leading particles. Then the bunch passes through a sequence of magnets that disperses the beam so that particles of different momenta travel on different paths. Particles with higher momentum (at the head of the bunch) travel a longer path than those of lower momentum (at the tail). The tail of the bunch can therefore catch up with the head, producing a shorter bunch—but at the cost of a greater energy spread.

This type of bunch compression has been used routinely in the SLC, where bunches 5 mm long are compressed to 0.5 mm for acceleration in the linac. Much shorter bunches will be required in the NLC. Short bunches will suffer less from transverse wakefields in the linac, and they permit a smaller depth of focus at the IP (about 100 microns for the NLC). In principle, another order of magnitude in compression could be obtained in a single stage; in practice, however, this approach would lead to other deleterious effects due to the large energy spread that would be induced in the beam. For this reason, the extra compression is provided by a second bunch compressor operating at a higher energy.

In the NLC, the bunch is first compressed as in the SLC to 0.5 mm in length, after which the beam is accelerated to about 16 GeV. The longitudinal spread of the beam is unaffected by this acceleration, but the relative energy spread decreases linearly with energy. The compression is then repeated, resulting in a bunch length as low as 50 microns. By separating the compression process into two discrete steps, we can keep the relative energy spread small throughout.

3.6 Linac Emittance Preservation[24]

The linac is the heart of the linear collider. As the beam is almost continuously accelerated, it is also focused transversely. During this process various effects conspire to dilute the emittance unless special care is taken. Because the linac is so vital and the potential for emittance dilution and beam size increase is so great, we will discuss various contributing factors in the next few subsections.

3.6.1 Injection Errors

After the bunch is compressed in length and as it enters the high-gradient linac, the bunch is about 2 μm high, 20 μm wide and 100 μm long. To obtain the necessary luminosity, the beam must be demagnified to the size shown in Table 1, $\sigma_y \times \sigma_x = 4$ nm $\times$ 320 nm. All of the offsets or angular kicks of the beam which occur upstream of the final focus system, however, get demagnified right along with the beam size. This means that the local beam size sets the scale for any offset and the local beam divergence sets the scale for any angular kick. If we examine the beam at some location along the accelerator, and if the beam motion from pulse to pulse is large compared to the beam size, then the beams will miss at the interaction point. In order to avoid this problem, these pulse-to-pulse offsets must be small compared to the local beam size. Equivalently, if a particular magnetic component has a varying amplitude, the variation of the angular kick must be small compared to the beam divergence at that point.

The emittance can also be destroyed by initial errors in beam size at the entrance to the linac. The beam size in an accelerator was discussed in Section 3.3. If there is bending or if the beam is offset in quadrupoles, the beam is dispersed with different momenta occupying different positions. In this more general case, the beam size is

$$\sigma^2 = \epsilon\beta + D^2\delta^2 \,, \tag{19}$$

where D is called the dispersion function and δ is the momentum variation in the beam prior to the bending field. For example, at the end of the compression section, $\delta \simeq 0.01$. At the entrance to the linac D should vanish. If not, this error in beam size results in emittance dilution in the acceleration process. For typical flat beam parameters the tolerance on dispersion D given by

$$\begin{aligned} D_y &< 0.2 \text{ mm} \\ D_x &< 2 \text{ mm} \,. \end{aligned} \tag{20}$$

The dilution caused by residual dispersion is additive. There are also multiplicative effects due to the mismatch of the beta function of the magnetic focusing lattice. If the beam were mono-energetic, these mismatches would not filament and could be compensated at any point along the linac. Since there is a significant energy spread, this mismatch must be avoided. For a small error in β at the entrance to the linac, and provided the filamentation is complete, the emittance dilution is given by

$$\frac{\Delta\epsilon}{\epsilon_o} \simeq \frac{1}{2}\left(\frac{\Delta\beta}{\beta}\right)^2 \,. \tag{21}$$

For incomplete filamentation, the emittance dilution will be somewhat less.

3.6.2 Wakefields and BNS Damping

Wakefields are a key problem not only for linear colliders, but for all accelerators and storage rings. The standard solution to this problem is to first reduce the wakefield forces until they are small compared to the applied external fields. Then compensation can be used, either feedback or modification of beam parameters or we can simply live within the limits by keeping the number of particles in the bunch sufficiently small.

In Sections 2.2 and 2.3 we discussed the effects of the long range wakefield. The multibunch beam breakup can be controlled by damping the undesirable modes in the RF structure; this reduces the long range wake at the second bunch but has little effect within the first bunch. Now we examine the effect of the short range wakefield on the stability of a single bunch.

The short range wakefield can be expressed again as a sum of modes; however, in this case it is necessary to include modes at very high frequency. A typical short range wake is shown in Fig. 12. It rises from zero, has a large peak and then oscillates with a frequency determined by the dominant mode. The bunches which will be in an NLC are so short that they fall on the initial rise of the wakefield. This is sometimes approximated as a linear rise (shown as the dotted line in Fig. 12).

The transverse wakefield increases rapidly with increasing frequency. If all dimensions are scaled, then

$$W_\perp(z) = \left(\frac{\lambda_o}{\lambda}\right)^3 W_\perp^o(z\lambda_o/\lambda) \,, \tag{22}$$

where λ is the scaled wavelength and λ_o is a reference wavelength. The initial slope varies inversely with the fourth power of the wavelength. Most of this variation comes only from the proximity of the iris hole to the beam. (By causality the short range wakefield must be independent of the distance to the outer wall of a structure.) It is, therefore, possible to reduce the short range wakefield by increasing the iris hole size relative to the wavelength. This reduces the short range transverse wakes, but it also decreases the effectiveness of the accelerating structure. Therefore, one must balance the transverse benefit of increasing the iris size with the increased RF power necessary to achieve a given acceleration field with the larger iris size.

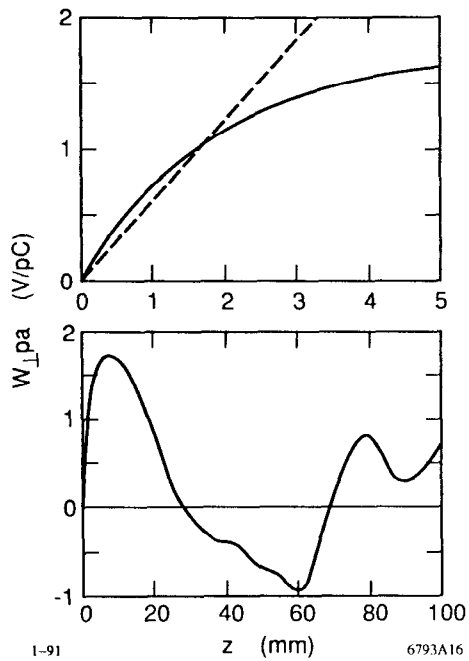

Fig. 12. The short range transverse wakefield at the SLC. The upper graph shows a detail of the behavior 5 mm behind a point bunch.

Even with the reduced wakefield within the bunch, there is still an instability induced within a bunch due to the coupling of the head and tail by the wakefield. The situation is completely analagous to that for multibunch instability discussed in Section 2.3; the same two-particle model suffices. In this case the head of the bunch, particle one, drives the tail of the bunch, particle two, on resonance. The growth is initially linear with distance but becomes exponential as the simple model breaks down.

Fortunately, there is a technique, called BNS damping, which can be used to compensate the instability.[25] The problem and solution are illustrated in Fig. 13 where a two particle model is shown. If the two particles are offset to one side of the structure, the wakefield force deflects the tail particle away from the axis. We add to this the external fields due to the focusing magnets; on the average there is a focusing force in the opposite direction. If we reduce the energy of tail of the bunch by inducing an energy correlation along the bunch (this occurs naturally and is controlled by the phase offset discussed earlier), then the tail particle experiences a stronger force than the head particle. Finally, if the additional force can be adjusted to cancel the wakefield force, then the two particles, the head and tail, move coherently together, and the growth is completely eliminated. The BNS correlated energy spread is given by

$$\left(\frac{\Delta E}{E}\right)_{\rm BNS} \equiv \delta_{\rm BNS} = \frac{e^2 N W_\perp(\sigma_z)\beta_o^2}{4E_o} , \qquad (23)$$

where N is the number of particles in a bunch, $W_\perp(\sigma_z)$ is the transverse wakefield evaluated at σ_z, and β_o is the β-function at energy E_o.

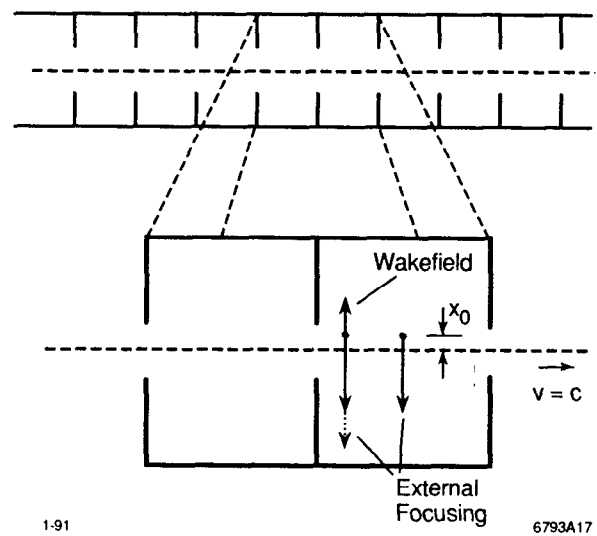

Fig. 13. Illustration of BNS damping. The additional focusing force on the lower energy trailing particle (dotted line) exactly cancels the wakefield force of opposite sign.

If the wakefield is large, then one can still satisfy Eq. (23) by using RF focusing rather than energy variation to vary the focusing strength. In this case, however,

trajectories can filament rapidly. To avoid emittance dilution with strong wakes, the alignment and trajectory tolerances are less than the beam size. This leads to 1 μm alignment tolerances.[26,27] As we shall see in the next sections, these tiny tolerances can be avoided by keeping the wakefields weak. With weak wakefields, tolerances are dominated by chromatic effects.

In the weak wakefield regime, BNS damping has been tested at the SLC linac.[28] In this case the tail growth due to a coherent oscillation was reduced by an order of magnitude. BNS damping has since been adopted as the normal running configuration for SLC.

3.6.3 *Chromatic Effects*

Upon injection into the linac, the compressed bunch has about a 1% uncorrelated energy spread. As the beam is accelerated, this relative spread decreases inversely with energy. At the same time a correlation between energy and bunch position is introduced due to the longitudinal wake and the curvature of the RF. Thus, the distribution in phase space becomes a wavy line which, when projected on the energy axis, yields an effective energy spread. At any location along the accelerator, the overall energy spread is a combination of the damping injected energy spread and the variation of energy along the bunch. After the bunch emittance is sufficiently damped, the relative energy spread remains constant unless deliberately increased by phase changes. For this reason it is useful to consider two models; one with constant energy spread and one with damping energy spread. In all cases considered below, we give not only the formula but also the value for the first design from Table 1 with energy 0.5 TeV in the center of mass.

The first chromatic effect to consider is that of a coherent betatron oscillation. If the variation of the phase advance with momentum (chromatic phase advance) is much greater than unity, the oscillation filaments. In this case the oscillation amplitude must be less than the beam size to avoid emittance dilution. If the chromatic phase advance is small ($\delta\psi_{\rm tot} < 1$), then the tolerance on a coherent oscillation of size $\hat{x}_o$ is

$$\begin{aligned} \hat{x}_o &< \frac{\sigma_\beta}{\delta_o \psi_{\rm tot}} = \frac{\sigma_\beta}{\delta_o \psi_{\rm cell}} \frac{2}{N_q} , \\ \hat{x}_o &< 2\sigma_\beta , \end{aligned} \tag{24}$$

where $\delta_o = 2\times10^{-3}$ is the constant relative momentum, $\psi_{\rm cell}$ and $\psi_{\rm tot}$ are the phase advance per cell and total phase advance respectively, and N_q is the number of quadrupoles. For the case of a damping energy spread with initial value $\delta_i = 0.01$, the tolerance is

$$\begin{aligned} \hat{x}_o &< \frac{\sigma_\beta}{\delta_i \psi_{\rm cell}} \frac{2}{N_q} \left(\frac{\gamma_f}{\gamma_i}\right) , \\ \hat{x}_o &< 5\sigma_\beta . \end{aligned} \tag{25}$$

For the case of a corrected trajectory let us consider the model of a sequence of random trajectory bumps. In this case the tolerance on the trajectory or alignment is

$$\begin{aligned} (\Delta x)_{\rm rms} &< \frac{\sigma_\beta}{\delta_o \psi_{\rm cell}} \left(\frac{3}{N_q}\right)^{1/2} , \\ (\Delta x)_{\rm rms} &< 30\mu{\rm m} \end{aligned} \tag{26}$$

for a constant energy spread δ_o. For an initial damped energy spread δ_i, we have

$$\begin{aligned} (\Delta x)_{\rm rms} &< \frac{\sigma_\beta}{\delta_i \psi_{\rm cell}} \left(\frac{1}{N_q}\right)^{1/2} \left(\frac{\gamma_f}{\gamma_i}\right)^{3/4} , \\ (\Delta x)_{\rm rms} &< 30\mu{\rm m} . \end{aligned} \tag{27}$$

3.6.4 *Misaligned Accelerator Sections*

BNS damping only cures the growth and filamentation of coherent oscillations in the linac; it is an average compensation rather than a local one. In an actual linac, the wakefield kicks are not cancelled locally by adjacent quadrupoles. This leads to an incoherent growth of wakefield tails due to the random sequence of misalignments between the trajectory and the accelerator structure. If we parameterize the strength of the wakefield kick by $\delta_{\rm BNS}$ as defined in Eq. (23), the tolerance on random accelerator misalignments is given by

$$\begin{aligned} (\Delta x_{\rm structure})_{\rm rms} &< \frac{\sigma_\beta}{\delta_{\rm BNS} \psi_{\rm cell}} \left(\frac{3}{N_q}\right)^{1/2} , \\ (\Delta x_{\rm structure})_{\rm rms} &< 25\mu{\rm m} , \end{aligned} \tag{28}$$

for $\delta_{\rm BNS} = 2.5 \times 10^{-3}$. From Eqs. (26) and (28) above, we see that the structure tolerances and quadrupole alignment tolerances are comparable provided that

$\delta_{BNS} \sim \delta_0$, that is, provided that the energy correlation needed for BNS damping is equal to the minimum energy spread in the linac.

3.6.5 *Compensation of Chromatic/Wakefield Effects*

The alignment tolerances shown above assume that the trajectory is a random sequence of bumps. There is no particular reason that it has to be random. Let us for the moment neglect wakefields. Then it is possible to measure the trajectories for particles of *different* energy and choose a trajectory which yields a small difference. Such a difference trajectory can be generated by scaling all the magnetic fields in the linac by a small amount so that the entire beam has an effective energy which is changed. By choosing the corrector sequence to minimize this difference trajectory (as well as the actual trajectory), the dispersion generated by misalignments is cancelled locally.

This technique is called dispersion-free correction. Provided that the beam position monitors have a *precision* the order of 1 μm, it is possible to essentially decouple the quadrupole misalignments from the dispersive effects.[29,30] This increases the tolerances given in Eqs. (26) and (27) by an order of magnitude.

When we include wakefields, the coherent motion is BNS damped and the incoherent motion gives rise to a random tail growth which can be controlled by tight tolerances. All that really matters for this effect is the value of the offset of the bunch within the structure. The offsets can be caused by two effects: misalignments of structures and trajectory offsets in structures. The trajectory is under our control; therefore, it is possible to use a trajectory which cancels the wakefield effects locally. Recently, it has been shown that by modifying the dispersion-free trajection technique, we can obtain a trajectory which cancels both the wakefield effects and the energy variation of the trajectory.[31]

Finally, we are left with the misalignments of accelerating structures. The most straightforward technique is to simply align the structure to the beam by using a BPM which is geometrically linked to the structure center. Such a BPM could consist of simply measuring the transverse wakefields induced by the beam.[32] One can use this information to either move the structure or move the trajectory to minimize the wakefield effects. Alternatively, for weak wakes, it is possible to deliberately move the beam or the structure to add a wakefield which cancels the effect of the rest of the accelerator.[33]

3.6.6 *Beam Tilt*

If there are RF kicks due to construction errors in the accelerator sections, the tail of the beam receives a different kick than the head. This can give a tilt to the beam. If we assume a random uncorrelated sequence of RF kicks, and compensate the center of the bunch with dipole correctors, the tilt tolerance is given by

$$(\Theta_{rms})\left\{\beta_o < \sin\phi_o >_{rms} \frac{2\pi}{\lambda_{rf}} \sqrt{N_s} \left(\frac{\gamma_o}{\gamma_f}\right)^{1/2}\right\} < \frac{\sigma_y}{\sigma_z} , \qquad (29)$$

where Θ_{rms} is the rms RF kick angle for a beam with energy γ_o, N_s is the number of accelerator sections, ϕ_o is the phase angle of the transverse kick relative to the bunch, and σ_z is the bunch length. For the parameters for collider number one in Table 1, we have

$$\Theta_{rms} < 2\mu\text{rad} . \qquad (30)$$

If such a kick is caused entirely by the systematic tilting of irises in an accelerating section (the bookshelf effect), then the tolerance on the systematic tilt angle of all the irises is given by

$$\Theta_{iris} < 0.3 \text{ mrad} . \qquad (31)$$

3.6.7 *Jitter and Vibration: Motion Pulse to Pulse*

Feedback is essential to handle the "slow" drift of x, x', y, y', E. In practical cases it is possible to feedback at $f \lesssim \frac{f_{rep}}{6}$. This sets the scale for what we consider slow. Time variation has many sources in linear colliders, for example: damping ring kicker jitter, power supply variations and ground motion. The jitter of the kicker in the damping ring must be kept small compared to the natural divergence of the beam at the kicker. Tolerances in power supply variations are also set in many cases by the beam divergence. The effects of ground motion depend upon the design and assumptions for the motion. If the wakes are weak and chromatic effects are kept small, there is no filamentation, and the beam moves coherently from pulse to pulse. If wakes are strong, and there is a large spread of betatron wave number, there is filamentation so that the beam size varies from pulse to pulse with a smaller centroid motion.

If we assume coherent motion, then for random magnet-to-magnet jitter the tolerance is

$$(\Delta x_{\rm jitter})_{\rm rms} < \frac{\sigma_\beta F}{\beta}\left(\frac{3}{N_q}\right)^{1/2}, \qquad (\Delta x_{\rm jitter})_{\rm rms} < (0.04)\sigma_\beta \ , \tag{32}$$

where F is the focal length of a lens. If, on the other hand, there is magnet-to-magnet correlated motion of amplitude Δx_λ, then the dominant effect occurs when the wavelength λ is equal to the betatron wavelength. However, since in the planned designs the betatron wavelength changes $\propto \gamma^{1/2}$, the resonance is only temporary. If $2\pi\beta_i < \lambda < 2\pi\beta_f$, then the tolerance is given by

$$\Delta x_\lambda < \sigma_\beta \frac{2}{(\pi\psi_{\rm cell})^{1/2}}\left(\frac{\gamma_f}{\gamma}\right)^{1/2}\left(\frac{2}{N_q}\right)^{1/2}, \qquad \Delta x_\lambda < (0.1 \text{ to } 0.4)\sigma_\beta \ , \tag{33}$$

where γ is the energy at which $2\pi\beta = \lambda$.

3.7 The Final Focus[34-37]

At the end of each linear accelerator is a final focus system whose purpose is to compress the tiny bunches to sub-micron dimensions. To obtain the luminosity desired, the cross-sectional area of each bunch must be only a few hundred square nanometers. In addition, we must focus it to the shape of a flat ribbon (rather than a string) in order to minimize the radiation emitted as the particles in the bunch encounter the intense electromagnetic field of the opposing bunch. These goals are accomplished by the use of a complex magnetic focusing system analogous (in reverse) to an optical telescope used to magnify distant objects. This system uses quadrupole magnets as focusing elements in a unique combination that provides a very large demagnification.

A major problem is the so-called "chromatic" effect of the final quadrupole magnets. Two parallel electron beams with different momenta entering a perfect quadrupole magnet are brought to a focus at slightly different longitudinal positions because the lower energy beam is bent slightly more than the higher energy beam by the magnetic field (see Fig. 14). For it not to affect the spot size, this shift of focal point must be smaller than β^*, the depth of focus of the beam. Due

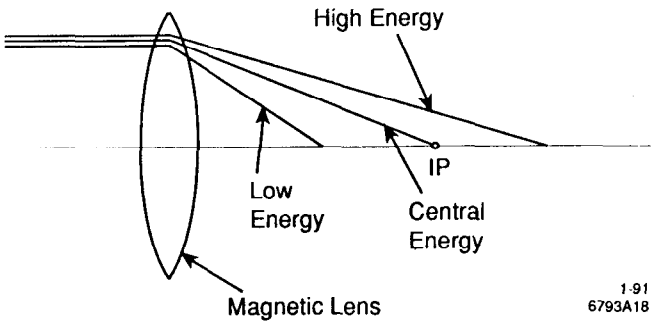

Fig. 14. The "chromatic" effect of the final quadrupole focusing magnet. Particles of differing energy are focused to different locations.

to the requirement of flat or ribbon-shaped beams, this depth of focus is about 100 microns in the vertical dimension.

Such a small depth of focus makes the chromatic effects particularly serious. The chromatic correction of the final quadrupoles is in fact the key to the final focus. Upstream of these quadrupoles, a combination of bending magnets that disperse the beam combined with nonlinear sextupole magnets ensures that higher energy particles get a bit *more* focusing than lower energy particles. When a bunch arrives at the last quadrupole, the chromatic effect of the magnetic field upon it is exactly canceled.

The basic principles of the chromatic correction for particle beams have been known for about 30 years. Their first application in a linear collider was in the SLC, where the beams are demagnified by about a factor of 30, yielding spot sizes of about three microns. Because the demagnification necessary in the NLC is about a factor of 300, however, the design of its final focus system will be substantially different from that of the SLC.

In order to test such a next-generation final focus experimentally, an international collaboration including SLAC, INP, KEK, Orsay and DESY has been formed to design and construct a Final Focus Test Beam (FFTB) at SLAC.[38] This facility will use the SLC beam emerging straight ahead from the linac as its source of electron bunches.

. Figure 15 shows a schematic of the location and layout of the FFTB. It is a scaled version of an NLC final focus, and as such, is qualitatively similar to NLC designs. A special feature of the design is that the chromaticity-correcting sextupoles are grouped in separate pairs, one for the horizontal dimension and one for the vertical. This pair of magnets is arranged so that the nonlinear aberrations introduced are cancelled, while the chromatic effects add. The bends shown in Fig. 15 horizontally disperse the different momenta in the beam so that the sextupoles give somewhat more focusing to the higher-energy particles. This additional focusing is arranged so as to cancel the lack of focusing of the higher-energy particles in the final quadrupoles.

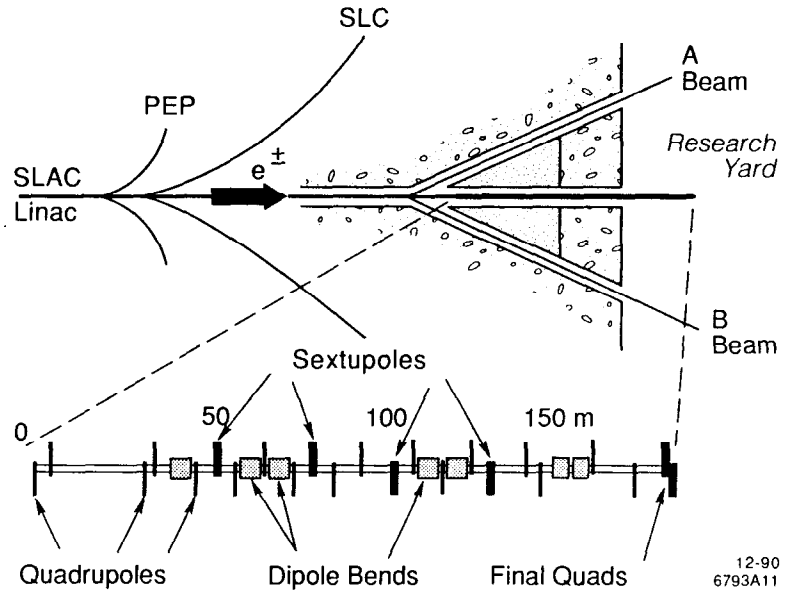

Fig. 15. The location and schematic layout of the Final Focus Test Beam.

The goal of the FFTB is to produce bunches with transverse dimensions of 60 nanometers high by 1 micron wide. Figure 16 shows the vertical beam size plotted versus the vertical β^* at the IP. In an ideal linear system, as discussed in Section 3.3, the beam size is just proportional to the square root of β^*. This is shown as the dotted line in Fig. 16. If the bunch has finite energy spread and with no correction, this linear decrease is modified by chromatic aberrations so that the beam size reaches a minimum of about 1 μm (the solid line in Fig. 16). Finally, if the chromatic-correction sextupoles are powered and if the system is properly tuned and adjusted, the vertical beam size follows the linear optics down to a size of about 60 nm before other high-order effects spoil the compensation (the dashed line in Fig. 16).

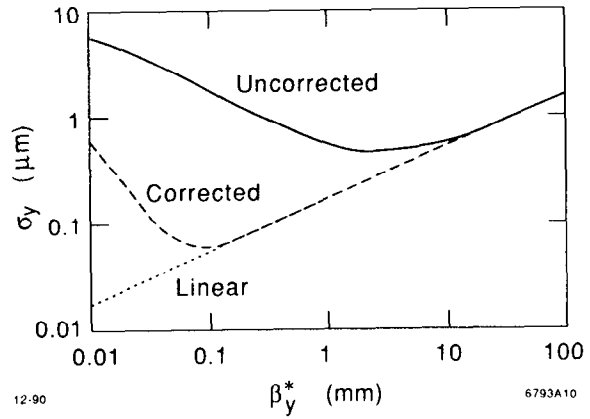

Fig. 16. Beam size versus optical tuning for uncorrected optics (solid), corrected optics (dashed) and linear optics (dotted).

The FFTB will not achieve the beam size necessary for NLC due to the lack of a suitable low-emittance source. In fact, to achieve such low emittance, we need the NLC damping ring and linac. The FFTB *will*, however, test the optical demagnification necessary for an NLC. In fact, the design β^* for FFTB is identical to that for NLC. In addition to this primary goal, the collaboration will use this facility to test the alignment, stability and instrumentation requirements needed to achieve such small spots. The Final Focus Test Beam is a key component of the worldwide research effort towards the NLC.

3.7.1 *Beam-Beam Effects*[39-44]

When two oppositely-charged bunches collide at the IP, the intense electromagnetic field generated by the bunches tends to mutually focus them. This leads to disruption of the bunch and to a pinch enhancement of the luminosity. The enhancement factor H_D was given in Eq. (12) for the luminosity. For round bunches, this enhancement can be quite large ($\gtrsim 5$); for flat bunches, however, it is considerably reduced ($\lesssim 2$) because the pinch only occurs in one dimension. If, in addition, the bunches are misaligned relative to each other, the centroids are attracted during the bunch passage. This leads to a two-stream instability which for moderate disruption actually helps the collision process; if the bunches are misaligned, they bend toward each other and collide partially anyway.

The combination of very high electromagnetic fields and high particle energy yields substantial amounts of synchrotron radiation known as beamstrahlung. The average energy loss due to beamstrahlung ranges from 1 to 30 percent in various NLC designs. In extreme cases, many of these photons can subsequently generate electron-positron pairs in the intense electromagnetic fields present. The radiated photons or charged particles can strike detector components, causing undesirable backgrounds.

The train of bunches on each cycle also presents a problem at the final focus. In order to have a separate channel for the outgoing disrupted bunch, collisions take place at a small angle. As the bunches approach the collision point, they feel the field from those bunches which are exiting and have already collided. This sequence of bunches can induce a multibunch kink instability which can cause trailing bunches to miss the interaction point. This effect can be controlled by the charge per bunch or by the crossing angle.

In practice, these beam-beam effects are what impose the ultimate limits on the possible charge per bunch—and thus on the luminosity. In the design described above, the luminosity limit is bypassed by using a short train of bunches, each with moderate total charge. This approach allows us to maintain the desired luminosity while keeping beam-beam effects under control.

4. Summary and Concluding Remarks

In the previous sections we have outlined the basic issues important in the design of a Next Linear Collider. The energy can be obtained by essentially conventional means, with the use of RF accelerating structures combined with high peak power RF sources—klystrons—which are similar to those used presently in the SLC. The key difference is the change of frequency by a factor of 4. For the structures this change of frequency presents no problem. Structures at 11.4 GHz have been constructed; damped and detuned structures have been built or are being designed. The power source is very close to realization. The klystron discussed in Section 2.5.1 could easily provide enough power for the lower gradient option (option 2) in Table 1. The RF pulse compression necessary to achieve the proper pulse length has been tested and has behaved as theory would indicate.

The luminosity of the Next Linear Collider is perhaps the more difficult problem. To reach the desired levels of $10^{33} - 10^{34}$ cm^{-2}sec^{-1}, it is necessary to compress the beam spot to a few hundred square nanometers. This situation is forced upon us by conservation of energy; the wall-plug power must be kept within reasonable bounds. In spite of the small size required, many of the tolerances can be brought to conventional values when compensation techniques are applied. Many of the issues of producing small spots will be addressed by the Final Focus Test Beam.

The second major component to the luminosity increase is the acceleration of many bunches on each machine cycle. This increases the efficiency of the collider but also introduces many complications throughout all the subsystems. Experience has been gained at the SLC which accelerates three bunches on each cycle, and also at all long-pulse linacs which accelerate sometimes thousands of bunches on each cycle. Thus far, no fundamental problems have been discovered which would preclude the acceleration of trains of bunches.

To conclude, the outlook for obtaining both the energy and luminosity of a Next Generation Linear Collider is bright. Provided that the engineering effort on the power source is successful, an NLC design could become a reality by the mid 1990s.

Acknowledgements

I would like to thank Tanya Boysen for help in preparing the manuscript and Perry Wilson for his careful reading and good suggestions. Finally, I would like to thank the members of the Accelerator Theory and Special Projects Department and the Accelerator Department at SLAC for their continuing work on SLC and NLC; without their effort this paper could not have been written.

References

1. Proceedings of the Workshop on Physics of Linear Colliders, Eds. L. Palumbo, S. Tazzari and V.G. Vaccaro, Capri, Italy (1988), available from INFN Frascati, Italy.
2. Proceedings of the Summer Study on High Energy Physics in the 1990's, Snowmass, Colorado, July 1988, World Scientific, Singapore (1989).
3. Proceedings of the International Workshop on Next Generation Linear Colliders, SLAC, Stanford, CA, Dec. 1988, SLAC-Report-335.
4. Linear Collider Working Group Reports From Snowmass '88, Ed. R. D. Ruth, SLAC-Report-334.
5. Proceedings of the 2nd International Workshop on Next-Generation Linear Colliders, Eds. S. Kurokawa, H. Nakayama and M. Yoshioka, KEK, Tsukuba, Japan, March 1990, KEK Internal 90-22.
6. R.D. Ruth, "Multibunch Energy Compensation," SLAC-PUB-4541 (1989), and in Ref. 1.
7. K.A. Thompson and R.D. Ruth, "Controlling Transverse Multibunch Instabilities in Linacs of High Energy Linear Colliders," Phys. Rev. D, 41, p. 964 (1990), and in SLAC-PUB-4801 (1989).
8. K.A. Thompson and R.D. Ruth, "Multibunch Instabilities in Subsystems of 0.5 and 1.0 TeV Linear Colliders," SLAC-PUB-4800 (1988), and in Refs. 2 and 4.
9. R.B. Palmer, "Damped Accelerator Cavities," SLAC-PUB-4542 (1988), and in Refs. 2 and 4.
10. H. Deruyter *et al.*, "Damped Accelerator Structures," Proceedings of the 2nd European Particle Accelerator Conference, Nice, France, (1990), and in SLAC-PUB-5263.
11. N.M. Kroll and D.U.L. Yu, "Computer Determination of the External Q and Resonant Frequency of Waveguide Loaded Cavities," Particle Accel., 34, 231 (1990), and in SLAC-PUB-5171.
12. M.A. Allen *et al.*, "High Gradient Electron Accelerator Powered by a Relativistic Klystron," Phys. Rev. Lett. 63, p. 2472 (1989).
13. M.A. Allen *et al.*, "RF Power Sources for Linear Colliders," Proceedings of the 2nd European Particle Accelerator Conference, Nice, France (1990), and in SLAC-PUB-5274.
14. Z.D. Farkas, "Binary Peak Power Multiplier and its Application to Linear Accelerator Design," IEEE Transcripts on Microwave Theory and Techniques, MTT-34, No. 10, p. 1036 (1986), and SLAC-PUB-3694.

15. P.B. Wilson, "RF Pulse Compression and Alternative RF Sources," SLAC PUB-4803 (1988), and in Refs. 2 and 4.

16. Z.D. Farkas, G. Spalek and P.B. Wilson, "RF Pulse Compression Experiment at SLAC," Proceedings of the 1989 IEEE Particle Accelerator Conference, Chicago, Ill. (1989), and in SLAC-PUB-4911.

17. T.L. Lavine *et al.*, "Binary RF Pulse Compression Experiment at SLAC," Proceedings of the 2nd European Particle Accelerator Conference, Nice, France (1990), and in SLAC-PUB-5277.

18. P.B. Wilson, Z.D. Farkas and R.D. Ruth, "SLED-II: A New Method of RF Pulse Compression," Proceedings of the Linear Accelerator Conference, Albuquerque, NM (1990), and in SLAC-PUB-5330.

19. E.D. Courant and H.S. Snyder, "Theory of the Alternating-Gradient Synchrotron," Ann. of Phys. 3, 1 (1958).

20. T.O. Raubenheimer, L.Z. Rivkin and R.D. Ruth, "Damping Ring Designs for a TeV Linear Collider" SLAC-PUB-4808 (1988) and in Refs. 2 and 4.

21. T.O. Raubenheimer *et al.*, "A Damping Ring Design for Future Linear Colliders," Proc. of 1989 IEEE Part. Acc. Conf., Chicago, Ill., p. 1316, and in SLAC-PUB-4912.

22. S.A. Kheifets, R.D. Ruth, J.J. Murray and T.H. Fieguth, "Bunch Compression for the TLC. Preliminary Design," SLAC-PUB-4802 (1988), and in Refs. 2 and 4.

23. S.A. Kheifets, R.D. Ruth and T.H. Fieguth, "Bunch Compression for the TLC," Proc. of Int. Conf. on High Energy Acc., Tsukuba, Japan (1989) and in SLAC-PUB-5034.

24. R.D. Ruth, "Beam Dynamics in Linear Colliders," Proceedings of the Linear Accelerator Conference, Albuquerque, NM (1990), and in SLAC-PUB-5360.

25. V. Balakin, A. Novokhatsky and V. Smirnov, Proc. of the 12th Int. Conf. on High Energy Accelerators, Fermilab, p. 119 (1983).

26. J. Tuckmantel, "Beam Tracking With RF-Focussing in CLIC," CLIC-Note 87 (1989).

27. G. Guignard, "Status of CERN Linear Collider Studies," Proceedings of the 1990 Linear Accelerator Conference, Albuquerque, NM.

28. J. Seeman et al., SLAC-PUB-4968, to be published.

29. T.O. Raubenheimer and R.D. Ruth, "A New Trajectory Correction Technique for Linacs," Proceedings of the 2nd European Particle Accelerator Conference, Nice, France (1990), and in SLAC-PUB-5279.

30. T.O. Raubenheimer and R.D. Ruth, "A Dispersion-Free Trajectory Correction Technique for Linear Colliders," SLAC-PUB-5222 (1990), submitted for publication.

31. T.O. Raubenheimer, "A New Technique for Correcting Emittance Dilutions in Linear Accelerators," SLAC-PUB-5355 (1990), Submitted for publication.

32. This technique was suggested by W. Schnell and V. Balakin.

33. J. Seeman, "New Control of Transverse Wakefield in a Linac by Displacing Accelerating Structures," SLAC-PUB-5337 (1990).

34. K. Oide, "Final Focus System for TLC," SLAC-PUB-4806 (1988), and in Refs. 2 and 4.

35. J. Irwin, "The Applications of Lie Algebra Techniques to Beam Transport Design," to be published in Nucl. Inst. & Meth., and in SLAC PUB-5315 (1990).

36. J.J. Murray, K.L. Brown and T.H. Fieguth, "The Completed Design of the SLC Final Focus System," Proceedings of the 1989 Particle Accelerator Conference, p. 331, and SLAC-PUB-4219 (1987).

37. K. Oide, "Synchrotron Radiation Limit on the Focusing of Electron Beams," Phys. Rev. Lett., 61, 1713 (1988).

38. J. Buon, "Final Focus Test Beam for the Next Linear Collider," Proc. of the 2nd European Part. Acc. Conf., Nice, France, 1990.

39. P. Chen, "Disruption, Beamstrahlung, and Beamstrahlung Pair Creation," SLAC-PUB-4822, 1988, and Refs. 2 and 4.

40. R. Blankenbecler, S.D. Drell and N. Kroll, "Pair Production From Photon Pulse Collisions," Phys. Rev. D, 40, p. 2462, 1989, and SLAC-PUB-4954.

41. P. Chen and V.I. Telnov, "Coherent Pair Creation in Linear Colliders," Phys. Rev. Lett., 63, p. 1796, 1989, and SLAC-PUB-4923.

42. M. Jacob and T.T. Wu, "Pair Production in Bunch Crossing," Phys. Lett. B, 221, p. 203, 1989.

43. V.N. Baier, V.M. Katkov and V.M. Strakhovenko, Proc. 14th International Conf. on High Energy Particle Accelerators, Tsukuba, Japan, 1989.

44. M.S. Zolotarev, E.A. Kuraev and V.G. Serbo, "Estimates of Electromagnetic Background Processes for the VLEPP Project," Inst. Yadernoi Fiziki, Preprint 81-63, 1981; English Translation SLAC TRANS-0227, 1987.