SLAC-PUB-8402
March 2000
LCC-009
CLIC-Note-418

Using Octupoles for Background Control in Linear Colliders - an Exploratory Conceptual Study

Rainer Pitthan

CERN
Geneva, Switzerland

and

Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94309 USA

Presented at 8th International Workshop On Linear Colliders (LC 99) ,
10/21/1999—10/26/1999, Frascati, Italy

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309

Work supported by Department of Energy contract DE–AC03–76SF00515.

Using Octupoles for Background Control in Linear Colliders - an Exploratory Conceptual Study

Rainer Pitthan [1]
CERN, Geneva, Switzerland and SLAC, Stanford, California, USA

Abstract. If one adds a suited Octupole (or an even higher multipole) lattice to linear collider Quadrupole FODO lattices, the amplifying properties of the combined lattice drive particles in the tails, but not those in the core, into resonant losses. This approach is quite different in concept and beam dynamics impact from past proposed use of non-linear elements for collimation.

This non-traditional scheme for background control has the added advantage that most, or maybe all, of the Halo collimation can be done using the lever arm of the real estate of the main accelerators, thus reducing the costly length of a separate dedicated collimation section and also unifying machine protection and background control. Simulations of particle distributions are presented. This approach requires cooperation by the designers of the accelerators, the beam delivery system, and the Detector, because a careful balance between sometimes conflicting requirements has to be found.

As a second component of this approach the use of Octupoles right before the final focusing Quadrupoles is proposed in order to enlarge the effective beam stay clear by a factor of 2-3, thus reducing the requirements for collimation. This concept would reduce the requirement for collimation but simulation have not been carried out here in detail.

To further explore and implement this concept will require a considerable effort in manpower, possibly comparable to, although less in scope, than the effort to develop the NLC RF or the CLIC RF schemes.

Executive Summary
For the hurried executive reader, the punch lines are summarized in Figures 14, 20 and 24.

I THE PROBLEM

The background in the SLAC Large Detector (SLD) [1] has been the biggest obstacle to achieving higher luminosity with the Stanford Linear Collider (SLC), by decreasing the spot size through increasing the divergencies of the colliding beam. Increasing the divergencies increases the beam size in the final focusing Quadrupoles and, therefore, makes more particles in the tails hit those Quadrupoles, producing lethal background in the detector. The background problem will not be any easier to solve in future colliders, rather the contrary because of the higher energy. Similarly, in LEP, the useful luminosity has been limited by background in the detectors [2].

[1] permanent address: Stanford Linear Accelerator Center, Mail Stop 12, P.O.Box 4349, Stanford, CA 94309, USA. Work partially supported by the Department of Energy, contract DE-AC03-76SF00515.

The work shown here grew out of the simple idea that if Halo particles, as is believed, are the problem in collider background, they should be removed early in the system when produced and not after they have been accelerated to maximum energy. If classical passive scrapers are failing to do the job, or only at great expense to length and luminosity, then active non-linear machine elements should be applied which work only on the Halo and not the core. A simple schematic explanation is shown in Figure 1. The method is now given the fancy name "resonant-nonlinear-collimation", and the effects had been predicted by Chirikov [3], but the roots for the work shown here were much simpler in thought. Although the parameters chosen here are close in magnitude to NLC parameters, the concepts can be equally applied to any linear collider.

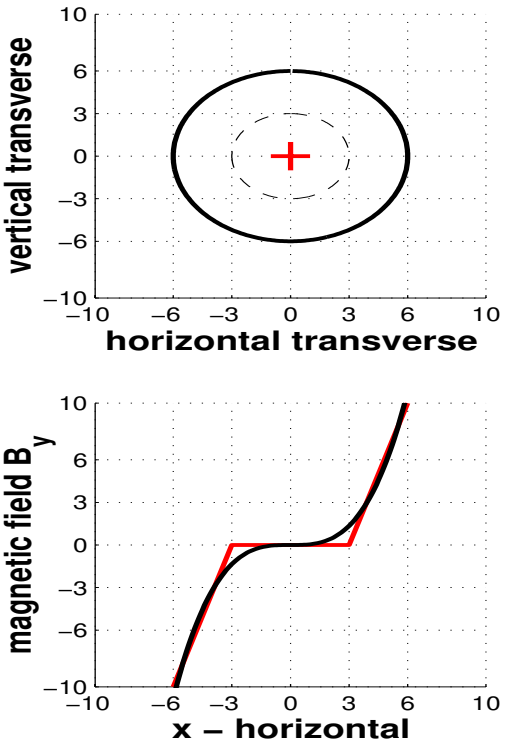

FIGURE 1. The top part shows the conceptual division of an accelerator in two concentric parts, with their separate lattices. Since this is physically not possible, ideally magnetic elements have to be found which are zero in the center and linear in the Halo region (light curve in the lower part). Maxwell's equations make this also impossible, so a compromise, here an Octupole field, is shown (dark curve in the lower part). But theoretically, see below, higher order multipoles have to be even better, although physically more difficult to build. It should strongly be emphasized that non-linear RF fields also should be considered.

Non-linear effects are tricky to develop because the normal common sense approach of the experimentalist, linear in nature, easily fails. In this study, I am indebted to many colleagues, in particular to: Franz-Josef Decker, who not only patiently listened to the ideas exposed here, but also helped me formulate a concept where a "stop band" (his phrase) for particles with large transverse amplitudes in space would be created; Dieter Walz, who was instrumental in helping with the problems of the interaction of relativistic electrons and matter, and who pointed out the importance of a large enough average incident angle of the collimating particles; David Whittum, who helped me to proceed when for a long time the method did not seem to work, by remaining convinced that the basic idea was correct, and that I just not yet had found the right solution; and last but not least Frank Zimmermann,

who always is able to rapidly make an analytical estimate of new ideas and show if it would work in principle or not. Some of his calculations are shown in reference [4] and on the NLC Beam Delivery Website.

Beam Halo particles are not the only danger for the Detector in the Interaction Point (IP), as there is also the synchrotron radiation produced in or close to the Final Doublet. This type of background one has learned to suppress after years of hard work with better synchrotron masking.

II GENERAL INTRODUCTION TO COLLIMATION

A Why yet another collimation system?

The NLC Zeroth Order Design Report [5] had a traditional design of a collimation section using a system of mechanical spoilers and collimators which had as a goal preventing background particles in the beam Halo beyond 6 σ [1] from hitting the material of the final focus insertion Quadrupoles, where the beam size is biggest and the aperture smallest. So why do this extra work?

The NLC collimation system length of several kilometers is determined by the condition that spoilers and absorbers must physically survive the impact of an entire errant bunch train (about 10^{12} electrons), else they be destroyed on impact. This requires a minimum spot size at point of impact, in order that the collimator surface does not fracture or that the collimator does not melt inside its volume at the point of maximum shower development [6]. For the NLC beam parameters, fracture and melting conditions give about the same spot-size limit, approximately $10^6 electrons/\mu\mathrm{m}^2$ for a copper absorber at 500 GeV [5].

This survival requirement leads to the necessity to have the beam first hit a thin spoiler, to ensure it has the minimum spot size needed when it interacts with the thick absorber. A schematic of such a classical, "hard edge", collimation system, from the NLC design [5], is shown in Figure 2.

As becomes clear from the foregoing, one of the detrimental points of the mechanical approach is the need to make the beam large at the places where the collimation takes place. Large here is relative, and in all collider projects the beams have to be enlarged from just a few μm to a few dozen μm vertically to be able to collimate them at the 10 *sigma* level. Still, absorbers have be as close as about 500μm to the beam to make the collimation work at the desired level.

This relatively large beam size is needed both for the mechanical collimation to work, and for the protection of the collimators themselves, in case of an aberrant beam. While the surface fracture does not depend on the beam energy, the melting limit does, since the energy of an electromagnetic shower deposited per unit length increases in proportion to the beam energy. Therefore, the beam area at the absorbers must increase linearly with energy. Since, in addition, the emittances decrease inversely proportional to the energy, the beta functions must increase not linearly but quadratically. Assuming that the system length L scales in proportion to the maximum beta function at the absorbers, this results in a

---

[1] This very nomenclature indicates that one assumes a Gaussian distribution of the beam. While this may be true for the center, the SLC/SLD assumption was that there is a flat Halo of a total strength of 1% beyond 3-4 $sigma_{Beam}$. In this paper an estimated width of about 10 σ_{Beam} [1] is used for the width of the Halo, where the width directly enters the calculations.

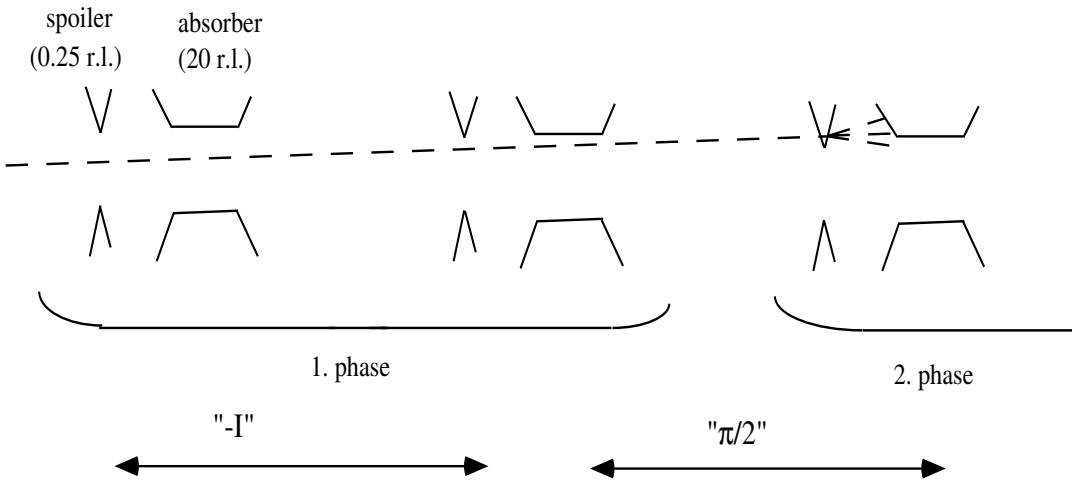

FIGURE 2. Schematic of a conventional collimation system taken from [5], consisting of a series of spoilers and absorbers which collimates in two planes and two lattice phase advances. The size of the spoilers and absorbers is approximately 1/4 and 20 radiation lengths (r.l.), respectively. A beam coming from the left hits first a spoiler, where its size is enlarged, before it interacts with an absorber.

quadratic dependence, $L \propto \gamma^2$ ($\gamma \equiv E_{beam}/m_e c^2$) [7]. Counting both sides of the IP, the NLC collimation system is 6 km long. Using the classical collimator approach, at 5 TeV the length of the collimation system for a linear electron collider could easily require 50 km.

In addition to all this, there is also an uncertainty in the calculations of the wakefields produced by the spoilers and absorbers. These need to be brought close to the center of the beam, as outlined above, thus creating strong wakefields diluting its emittance. The calculation of the wakefields for parallel plate geometry was underestimated by a factor of 6 in the NLC Design Report [5]. Consequently the effects on the emittance of the beams is much larger than planned for. Thus, research on an alternative collimation scheme which avoids these weaknesses is needed.

In summary, the classical "hard edge" absorber approach leads

- to a lengthy system because the beam has to be enlarged four times, at two phases 90^o apart in both x and y planes
- to large chromaticities with all their negative consequences when one corrects those with strong sextupoles (it limits the beam size achievable due to chromatic effects), and
- to general emittance dilution due to the wakefields of the collimators which need to be close to the beam to make the mechanical collimation system work.

While seemingly the "hard edge" approach of the classical collimation system leads to a well defined cut-off in real space and in phase space for background, experience has shown that in practice there is no such hard edge. This is presumably due to edge scattering, and due to close-to-wall wakefields, which in effect diffuses the hard edge. Moreover, it has been extremely difficult to correctly simulate these soft edge effects, which are believed to be the highest source of background in Collider Detectors.

One historical problem is that accelerators, beam delivery systems (BDS), and the experiments have generally been designed separately, each system being maximized by its inventors for it's own performance. This was done without consideration for the independent needs of the other systems, on the one hand, and the possible contributions the other

systems could make to the solutions of one's own problems, on the other hand.

This note, therefore, lays out a different approach to background control in linear colliders, which incorporates a collimation system in the front accelerator systems, and does not wait with collimation until particles have been accelerated to the maximum energy. This may not completely pre-empt the need for collimation in the BDS, but it will at least greatly relieve it. The recent development of having central injector systems and therefore, long transfer lines to the main linac, makes such an approach even more possible.

One could say that the new approach carries the background control for the Interaction Point (IP) backwards to the accelerators by using non-linear magnetic elements and non-linear beam dynamics with-in a linear lattice instead of, or at least supplementing, stand-alone mechanical collimators. A stand-alone Octupole system in the BDS is also investigated. This study focuses on Octupole magnetic elements, but higher order magnetic elements like Dodecapoles must be suitable in general and should be investigated. These higher order elements would have a more sharply defined collimation depth compared to Octupoles, although it would be more cumbersome to build them to the required strength.

In addition, RF focusing with RF Quadrupoles [8] should be re-visited, and background control with higher order RF fields should be considered. [1]

This RF focusing with Quadrupoles, which has been shown to work in principle and has the nice feature of being proportional to wakefield effects, was given up for the CLIC project after BNS damping was successfully implemented with magnetic Quadrupoles at the SLC. The RF Quadrupoles required more space than BNS with magnetic Quadrupoles. However, with the high energy colliders now under consideration, the total length including collimation seems to be shorter with RF focusing and without classical BNS. In particular one has to consider that some of the worrisome Halo background in the beam delivery system is produced by wakefields in the last part of the main linac when one has to go off BNS damping with magnetic Quadrupoles.

B The Approach Chosen, its Advantages and Limitations

Following the classical line of thinking about collimation, it seems that collimation with non-linear magnetic (or RF) elements has the disadvantage that the cut-off for backgrounds is by its nature is soft and gradual. That means, that even in theory in the transverse transition region of 8-10 σ some particles get through.

However, the non-linear element approach is more suited to proper simulation, thus in effect giving the experimentalist more realistic numbers to build the detector to, and less unavoidable surprises later on, which are difficult to cure in hindsight.

In addition there is another important advantage, in as much as the magnetic element approach protects the machine against off-momentum or off-center beams, which would pass the linac FODO lattice but could wreck havoc on the BDS (errant beams) in a natural way by diffusing the beam gradually and depositing it in special aperture elements over an extended length of the machine. This will be shown in detail later. Off-momentum beams can be produced by klystrons cycling on and off in the pre-accelerators; off-center beams can result from Quadrupoles partially shortening out.

And finally, it reduces the number of μ particles produced by the absorbers in the

[1] With higher order RF fields is meant higher than Quadrupole, e.g., RF Octupoles. I am thanking Dieter Quitmann for asking the question which made me look into RF (de-)focussing in general.

classical collimation section. The number of μ can be considerable: the probability of muon production per primary electron is several 10^{-4}, so if the collimation system at full energy collimates even only a Halo of 10^{-3} of a beam of 10^{12}, several 100,000 μs are being produced. These μs again have to be removed from the beam before the IP.

The conceptual approach chosen consists of two parts:

- peeling off Beam Halo particles with non-linear magnetic elements before the beam enters the final focus area [4], and then
- folding in the remaining tails with a doublet (or maybe triplet) of Octupoles in the final focusing area.

The second part is an added independent boon which augments the effectiveness of any collimation system without replacing it. Such a "folding in" was described first by Mead [9], and later implemented in an actual beam line by Enge and collaborators [10].

Using non-linear elements to peel off the beam resonantly specifically does not require enlarging the beam to enable mechanical collimators to work, rather it is the repetitive action of Octupole fields at the point where the betatron functions are at their **normal maximum** in the FODO lattice which does the work. It should not be surprising that this approach works best, with the weakest Octupoles and the least change required to the lattice, in the main linac where there are many FODO cells to do the work. For absorbers, thick copper pipes, water cooled if needed, captured in a Quadrupoles-Octupole assembly, with a circular inner radius of larger than 3-4 mm will suffice. They will have smooth tapered transitions to the design MPS spoilers at the end of each accelerator structure. Since the inner radius forseen for these spoilers is a few mm only, the absorbers with a 3-4 mm radius will create no additional problems with respect to wakefields.

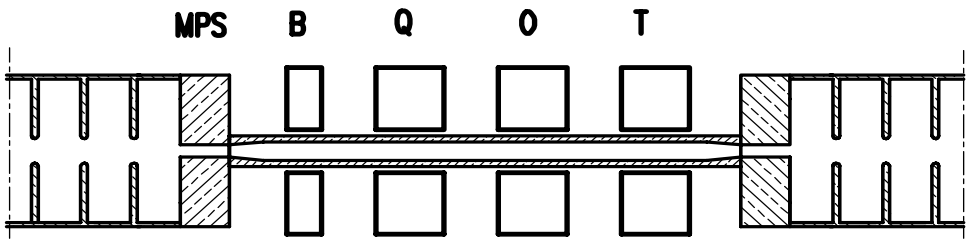

FIGURE 3. Schematic of the collimation assembly proposed: a circular copper pipe between acceleration sections, captured in the quadrupoles and Octupoles will suffice even for collimation of errant beams. The inner diameter of the collimator is thought to be 6mm, the outer diameter (OD) 10-13mm, depending on the pole tip radii of the magnetic elements. The pipe is tapered to match the aperture of the Machine Protection end-plugs (MPS) of the acceleration structures. The beam comes from the left, the copper pipe is captured in sequence by a BPM (B), a Quadrupole (Q), an Octupole (O), and finally a Tantalum block (T), if the latter is required at higher energies.

In areas of the collider other than the main linac this approach requires some modest change (20%) in the lattice length of the pre-accelerator, and a drastically different beam line in the post-accelerator area. There are still some open questions to answer, such as wide is the resonance, how different can the tune be between the two planes, etc.

The limitation of this paper is that it uses simulations for an "equivalent energy" FODO lattice, i.e., the simulations were performed for one energy only and the Quadrupole and Octupole lattices would have to be scaled to take actual acceleration into account. Since the

energy change in a linear accelerator is indeed linear, such an approach does not pose any limitations for the conceptual results of this study. It was checked that Octupoles for the small beam size at full energy can actually be built with manageable lengths.

C Concepts of Non-linear Elements

Before getting into the conceptually more involved simulation using 10,000's of particles under the influence of non-linear elements coupled in two planes, it is useful to look at the single particle motion in just one plane. Figure 4 shows the position of a particle on the x-axis as it goes through a combined Quadrupole-Octupole lattice.

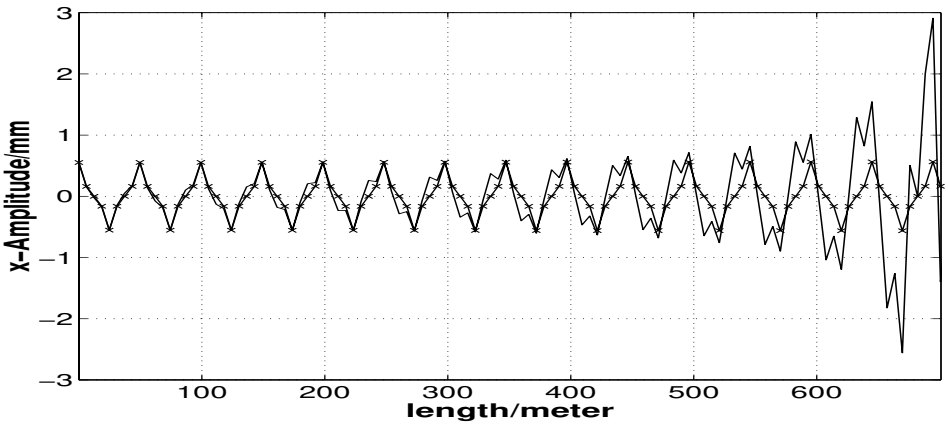

FIGURE 4. Although only one coordinate is shown here, this calculation was done for a realistic Octupole field with the required cross coupling terms in x and y. The single particle motion in the linear (Quadrupoles only) and non-linear (Quadrupoles and Octupoles) case are compared. The starting value was x= 600μm (corresponding to 9 σ in the NLC main linac) and $x' = 0$. The positions of the Quadrupoles is indicated with x's in the linear case. This curve shows the very regular even pattern of a constant 90^o FODO lattice. The added kicks of the Octupoles have the effect of changing the effective phase advance to higher values. The amplitude changes, slowly at first, then rapidly, until the particle gets lost at about 3 mm of amplitude. Note that this combined phase advance depends strongly on the position in x,y space, making predictions of when a particle will get lost close to impossible in general. This example was calculated with the standard NLC main linac lattice cell length of 12.5 m and Octupoles with a 3kG pole tip field at a pole tip radius of 5 mm ($k_3 = 43300m^{-3}$ in MAD notation).

The amplitude grows, gradually at first, until the particle is abruptly lost. This type of instability has been described by Chirikov [3] as "a motion of the (non-linear) system as if the latter were influenced by a random perturbation even so, in fact, the motion is governed by purely dynamical equations".

Frequently the question arises: are there not islands of stability for particles with large amplitude which could counteract collimation? In this work such islands have only been seen in uncoupled idealized calculations in one dimension (Figure 5), never in the coupled case in the full phase space of x,x' and y.y'. The absence of such islands is probably due to Arnold diffusion, but this is beyond the scope of this work.

In summary, as outlined above, the work described here evolved out of the basic idea of conceptually constructing two concentric beam dynamics lattices in one, covering the same beam line. One lattice was designed to work linearly on the core, while another lattice

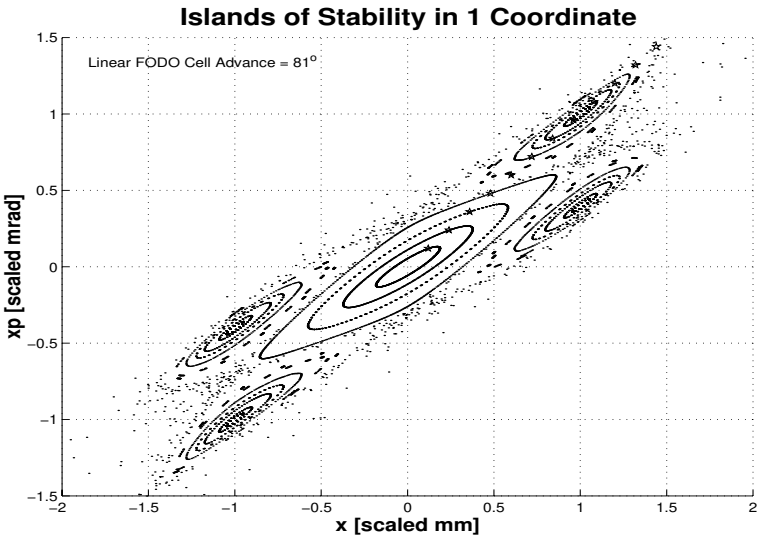

FIGURE 5. One coordinate only was used in this calculation without the cross coupling terms in x and y, and, therefore, islands of stability appeared. A FODO lattice with 81^o linear phase advance and 500 FODO cells, Quadrupoles and Octupoles, was used. The way this figure has to be interpreted is as a sequence of separate single particle motions. The particles are started in the diagonal of phase space, as indicted by the star symbols. They are plotted for each quadrupole. For the particles with a small amplitude in the center nothing happens. The Octupole fields are to weak to influence the motion and so these particles stay on their closed ellipse in phase space. For particles with larger amplitude the phase space gets deformed, and finally they get unstable (e.g., follow the particle with x=0.62). But then the system recovers, and particles are stable again; but not on closed ellipses, they hop from island to island. Figure 4 did show how the effective phase advance in the system changes; for the islands it should be assumed that this phase was close to, and passing through, 90^o. Particles even further out then quickly get lost (collimated).

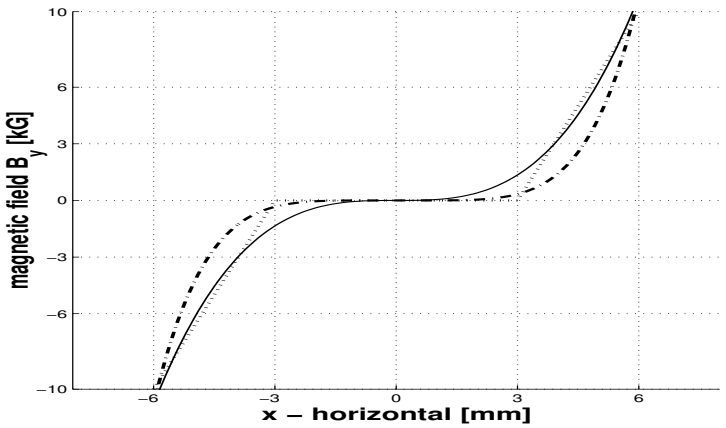

FIGURE 6. Comparison between the B-fields of an ideal magnetic element which leaves the center of the beam alone and influences only the tails (dotted curve), with actually buildable elements which will follow Maxwell's equations, here an Octupole (solid curve) and a Dodecapole (dash-dotted). For comparison the curves were normalized to reach 10 kG at 6 mm. They clearly show the theoretical advantage of a 12-pole.

with non-linear elements would only operate on the tails outside the core of typically about 5 σ. This approach has to be regarded as quite different from previous proposed uses of non-linear elements [11], [12]. These techniques used the non-linear elements to enlarge the beam over a shorter distance than classical telescopic Quadrupole systems and could thus save costs due to the reduced length of the collimation system. But since they then used classical collimators, the requirements on beam size at the collimation point and the incipient chromatic and wakefield problems stayed the same.

The ideas converged on using Octupole fields, although clearly Dodecapoles (12-poles) might be considered, too[1]. Figure 6 shows how the B-fields of such elements in the mid-planes of the magnets might look like. However, one must recognize that higher multipoles have strong cross-coupling terms in the diagonals which change the picture considerably, as shown for Octupoles in Figure 7.

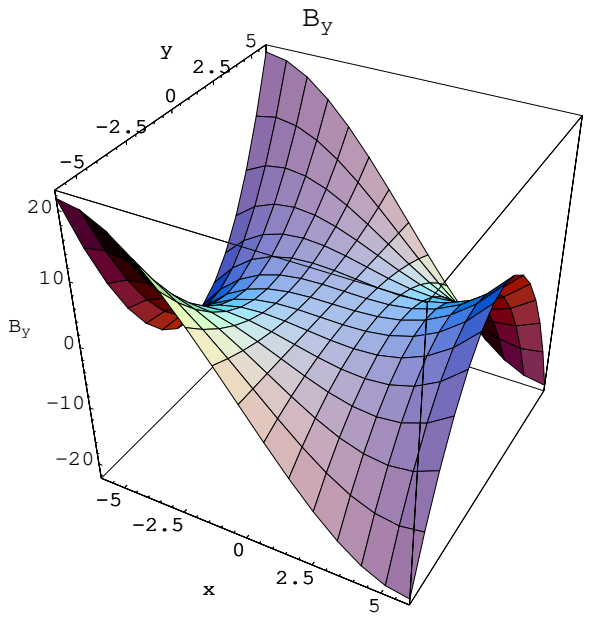

FIGURE 7. The 3-dimensional display of the B_y-field of the Octupole field of Figure 6. The B_x-field looks the same, with x and y rotated. The figure shows that for large amplitudes the cross-coupling terms, off the x and y-axis, dominate the on-axis fields.

D Tools used

For accelerator physics the conceptual difficulty in pursuing the approach outlined above apparently stems from the unusual requirement of a 1^{st} order application of 3^{rd} order elements, with effects targeted toward the tails of the approximate order 10 σ, which must not influence the core more than a few percent at most. Consequently, in order to not impact the core it is important to place the Octupoles in regions without design dispersion and to tightly control any spurious dispersion. While 2^{nd} and 3^{rd} order elements have been used extensively to perform higher order chromatic corrections on the core, an application of this kind apparently has never been used in a 1^{st} order application on the tails. In the course of

[1] Interestingly enough, dodecupole "tail collimation" was observed in the *SLC Positron Return Line* a decade ago, see [13].

this study it was found that well known beam dynamics programs, like MAD, seem to use approximations applicable to central orbit particles only [14], resulting in misleading, even wrong, results for the physics of the tails.

It is conceivable that the simulations shown here were successful only because they were performed using a simple thin-lens approximation program written in MATLAB (a pre-cursor was written in collaboration with Paul Emma and used for [4]), rather than the elaborate existing beam dynamics codes like DIMAD or MAD. Using MATLAB has the advantage, in addition to enabling direct and transparent control over the physics and its approximations, to have easy use of graphics capabilities. Good graphics are important in such an exploratory undertaking, where it is not clear which properties are important to quantify numerically.

Within MATLAB two different mathematical approaches were used: (A) a complete matrix multiplication for FODO lattices [15] and (B) simply adding Octupole kicks to the analytical equations for FODO lattices. Both formalisms agreed in the results; the latter implementation was faster on the computer and was, therefore, used for the bulk of the simulations.

E Some Basics of Octupoles

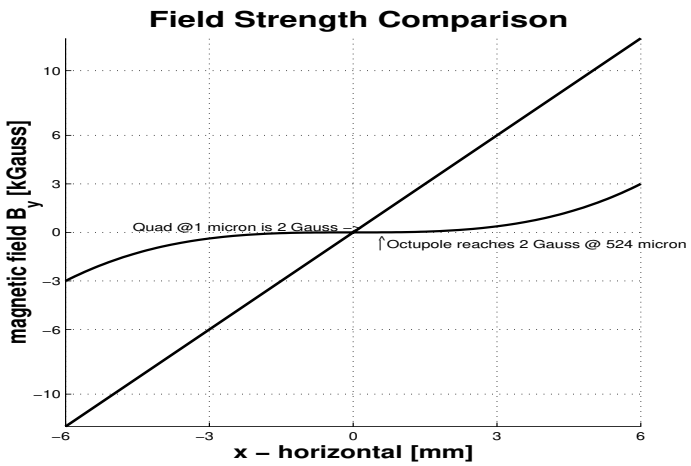

FIGURE 8. The magnetic field a standard collider Quadrupole reaches at 1 μm is about 2 Gauss, an appropriate Octupole reaches these 2 Gauss at a much larger off-set: 500 μm.

Some basics of Octupole physics, important for large amplitude particles, are dealt with first. To get a back-of-the-envelope feeling for the magnitude of the fields involved, the field strengths for a typical collider design Quadrupole, and the field for an Octupole, are compared in the region close to the x-axis. For the NLC design-Quadrupole of 1.2 Tesla with a pole tip radius of 6 mm, the field at 1 μm offset would be 2 Gauss. The Quadrupole alignment tolerances are about the same, 1 μm. In turn, for an Octupole with a pole tip field of 0.3 Tesla and the same pole tip radius, 2 Gauss field strength on axis is reached at the much larger distance of 100s μm (Figure 8) [1].

[1)] Figure 7 shows that the region close to the center of an Octupole is also flat off-axis, so that a field strength with impact on alignment tolerances is only reached at several 100 μm.

One concludes that for Octupoles of this strength the alignment tolerance then has to be in the many 100's of micron. And in general, this exercise should show that even the strongest conceivable Octupoles will have a negligible effect on the core of the beam, and should neither constitute an alignment nor a luminosity dilution problem. A simulation of the effects of alignment errors of Octupoles on emittance dilution shows this to be true [4].

To understand multipoles one has to understand their B-fields and the kicks those fields impart on particles traversing them, more precisely their B_x and B_y components. Following the complex notion of Reference [16], where

$$B^* = (B_x - iB_y) = -ik_1(x+iy)^n$$

(with $k_1 = b_1/r_0$) one finds by simple arithmetic the x and y components of the magnetic fields (and therefore the kicks) [1] For a quadruple (n=1) these are thus defined as

$$B_x = +k_1 y$$
$$B_y = +k_1 x.$$

The kicks then follow from the right hand 3-finger rule as:

$$\Delta\text{x'} = -k_1 y$$
$$\Delta\text{y'} = +k_1 x.$$

Physically a QF (Q_x) is changed into a QD (Q_y) by reversing the magnetic poles (or rotating the magnet by 90^o).

With the complex notation one easily sees that for Octupoles the sign of the leading term changes from the Quadrupole case. Here we have from $B^* = -i(x+iy)^3$ and with the common notation of $k_3 = b_3/r_0^3$:

$$B_x = -k_3(y^3 - 3x^2 y)$$

and

$$B_y = +k_3(x^3 - 3y^2 x).$$

The kicks then follow from the right hand 3-finger rule as:

$$\Delta\text{x'} = +k_3(x^3 - 3y^2 x)$$
$$\Delta\text{y'} = +k_3(y^3 - 3x^2 y).$$

A three dimensional plot of B_y (and, therefore, the x-kicks) was shown in Figure 7 and showed that for off-axis particles the cross-coupling terms become can larger than the on-axis fields. Note, since B_x and B_y have opposite signs, that in the Octupole case the kick experienced by a particle by either field (close to the axis) has the same sign. This property may be important for the effectiveness of resonant background collimation, because kicks imparted by one kind of Octupole are not partly negated by the next which is the of the other kind, as is the case for Quadrupoles (see more discussion, below) [2] . It may also be the reason that efforts to construct focusing channels with Octupoles, along the lines of a Quadrupole FODO lattice, have failed.

Dodecapoles do not have this same sign property. Possibly they will not be as effective for resonant collimation. However, since these are nonlinear processes, impossible to predict just so, a full simulation using 12-poles should be performed. The advantages of using 12-poles (see Figure 6) would, naturally, be a sharper onset of collimation in transverse tail space, and even less impact on the core than with 8-poles. The difficulties would mainly

---

[1] I thank Phil Bryant for pointing out the many possibilities for confusion of signs between European, Engineering, Physics, and US notations. His write-up in Reference [17] was most helpful.

[2] One can also see this through a basic geometric consideration, namely, that if one rotates Octupoles by 90^o to go from x to y one gets to the identical local magnetic North-South pole geometric configuration.

be to actually build strong enough magnets, so the onset of collimation would be later in σ-space.

The equivalent of focusing and defocusing Quadrupoles would be Octupoles with the poles interchanged (or the whole magnet rotated by 45^o). We will call them OF and OD, although they do not focus or defocus in the same, or even a similar, way as Quadrupoles. The most effective pattern for placement of Octupoles with a 90^o linear Quadrupole lattice has been investigated in [4]. It was found to be a repeating pattern of 2 OF's followed by 2 OD's. Most lattice simulations shown here have used this pattern at a linear (Quadrupole) phase advance of 90^o.

Some early simulations (Figure 9) showed that, with a different pattern of Octupole placement, different phase advances will be possible, but this avenue of research has not yet been followed in more detail.

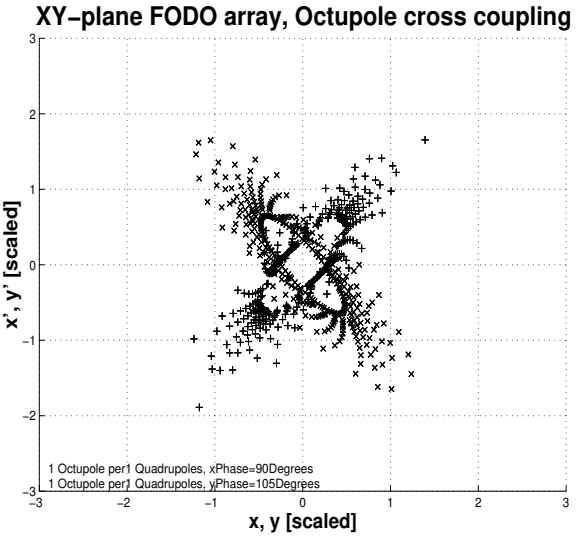

FIGURE 9. Phase space plot of one particle in x and y phase space. The particle goes through a 100 cell FODO lattice with a phase advance of 90 and 105^o in x (x's) and y (+'s), respectively. This simulation was performed with a full coupled matrix multiplication. The figure has to be interpreted as the (one) particle initially being on its unperturbed phase ellipse(s), then slowly spiraling out under the relentless kicks of the Octupoles, until it is abruptly lost.

In summary, the most effective phase advance for resonant collimation is determined by the pattern of placing the non-linear elements, and vice versa. In Ref. [13] it was found that the "tail collimation" (positron loss) was most severe for 90^o because the Dodecapole component of the fields were due to the way the Quadrupoles were designed, thus the pattern was fixed and attached to the Quadrupoles.

III COLLIDER SPECIFIC PROBLEMS

A General

Collider collimation systems have to serve two different functions: they must be able to (A) protect downstream systems from the full charge of errant bunch trains, whether

produced by large betatron excursions or energy errors, and (B) remove the beam Halo. Octupole systems can deal with both problems.

For the **errant beams**, an Octupole based collimation system is "self-protecting" as will be seen. If the beam goes through the centers of the Octupoles, nothing happens to the core. If the core of the beam is off-set, for the Octupole strength used here, the beams are blown up over many betatron wavelength (FODO cells) and the beam is lost over a length of system in different places horizontally and vertically. Thus, neither the local activation by nuclear processes nor the direct local structural damage to the material is as severe as in classical systems.

B The Errant Beam Problem

Although not "background" in the classical sense, the errant beam problem has to be treated carefully in any context of collimation. Since it is not immediately obvious why the errant beam needs to be treated at all, because without an Octupole lattice such an off-set would have no impact while the beam is going through the accelerator(s) and the strong focusing FODO lattices, an explanation might be in order.

Like a garden hose carries water, the strong focusing lattice will transport the beam through, nearly no matter what the off-set. The danger arises later in the bends of the collimation and final focus sections where, at TeV energies and μm beam sizes, the beams will destroy any beam line elements in their way. So it is important to collimate, or absorb, an errant beam before it leaves the strong-focusing FODO lattice of the main accelerator(s), or to have such a strong focusing system after the accelerator, or both. The best way to collimate is gradually, so as not allow any damage to the collimating elements. Octupole collimation does precisely this in a natural way as will be seen. Even if a good stand-alone solution for collimation in the BDS should be found, a low strength Octupole system in the linac as a defense against errant beams is recommended.

Two main causes are foreseen to off-set a beam from the nominal central trajectory:

- klystrons cycling on and off in the pre-linac which accelerates the beams between the damping rings and the 180^o arc of the second bunch compressor. This will produce a horizontal off-set.
- one coil of a Quadrupole shortening out and effectively creating a bend magnet. This can produce off-sets in the horizontal and vertical.

We deal here exclusively with the horizontal case, which is the more probable and the more dangerous. More probable because of the cycling klystron effect, and more dangerous because the blow-up in the vertical in this case will be small, i.e. the beam will act as a horizontal knife, slicing material it encounters. If one can show that the horizontal off-set can be handled, any off-set can be handled.

Of the two main causes for off-set beams identified here, the klystron problem is the most severe, because it may happen from one pulse to the next. The shortened Quadrupole coil problem can be mitigated through the construction of the Quadrupole iron core from extended solid steel pieces, rather than laminations, leading to a magnetic field decay time of the order of 100-200msec. This is long enough that the beam can be switched off or reduced by other means, such as interpulse intensity comparators.

There are several important parameters for the protection of the machine:

- (A) The *collimation radius*, that is, the radius of the narrowest repeating aperture in the FODO lattice with respect to the beam size. To be most effective, this defining aperture should be in the Quadrupoles of the FODO lattice, where the beams are largest. And to be protective of the diskloaded wave guides, it has to be smaller than the smallest iris' of those.
- (B) The *transverse off-set* of the incoming beam. Depending on other constraints on choices for the strength for the Octupoles, off-sets up to 10-15 sigma can be handled.
- (C) There is, finally, the *strength of the Octupoles* required.

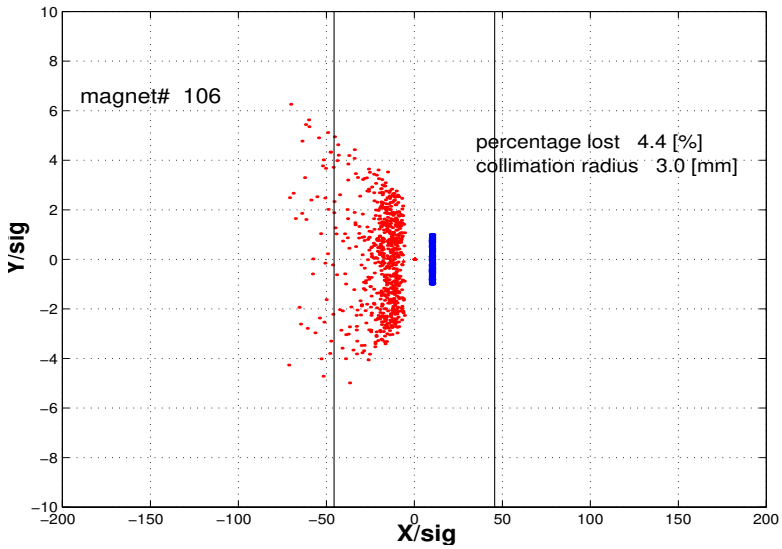

FIGURE 10. Pattern of loss of particles in cell number 53 down the main accelerator, for a 3 mm collimation radius, and an original off-set of 10 σ. Note the very different scales in x and y. The black rectangle on the right side is the original beam distribution with 1 σ each in x and y and a corresponding 1 σ distributions in x' and y' (not shown). The left side shows the distribution of the beam after it has been submitted to kicks of the Octupoles in each previous FODO cell. The particles to the left of the black line at 3 mm are the ones which are going to be lost in this cell. See also the next Figure for a 4 mm collimation radius and the large difference in penetration depth, indicative of the non-linear effects.

This distributed absorber system of the collimation radii should consist of a circular, [1] several mm thick, water cooled copper beam pipe which goes from one accelerator section through the BPM-Quadrupole-Octupole assembly to the next accelerator section. If needed, at the higher energies (0.5 TeV) that can be enhanced by a tantalum block just before the next accelerator section. The next 2 figures (Figures 10 and 11) show the local situation in a main-accelerator linac for a beam with 660 μm horizontal off-set for two different assumed collimation radii.

A typical off-set foreseen would be in the 10 σ region, which corresponds approximately to 600 μm - 700 μm in the front end of the machine. In this case the Octupoles, as described in the first section, will do to the off-set core what they will normally to the tail: they will slowly remove the beam. But more important, they will only do that after they have blown up the size of the beam by a large factor, thus reducing the destructive power when it interacts with material. Equally important is the average angle $\langle x' \rangle$ of the lost particles,

[1] Circular collimators produce the least wakefields.

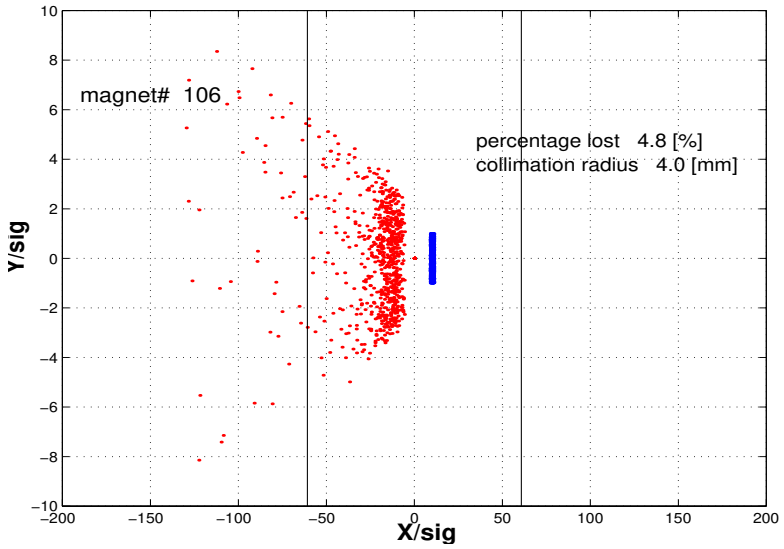

FIGURE 11. Same as Figure 10, for a 4 mm collimation radius. Note the large difference in penetration depth as compared to the case of a 3 mm collimation radius in Figure 10, indicative of the non-linear effects. Also note that the percentage loss is about the same.

typically having grown to more than 100 micro radian when the collimation starts. That means, as an added benefit, that a beam originally even as small as $\sigma = 10\mu$m will be spread out over 20 cm longitudinally, increasing the effective beam size by another large factor and making it non-destructive.

Figures 10 and 11 show that a larger collimation radius leads to a larger penetration depth and that, therefore, the collimation radius has to be carefully optimized. If it is too small, the errant beam will not spread out enough. But if it is too big, particles will not be absorbed in the defining aperture, and instead will reach the copper acceleration structures.

The biggest concern of the effects of errant beams is a loss which is too rapid in the longitudinal dimension. If the cumulative effect of the non-linear fields is such that a large percentage of the beam is lost in one spot, than even the widening of the beam as described will not prevent damage. On the other hand, the simulations, shown below in the linac section, show that the Octupoles cannot be too weak or the background will not be removed rapidly enough; in other words, more particles will be transported from the few sigma tail to the region within the 3 mm, but not removed. A strength between $k_3 = 15000m^{-3}$ and $k_3 = 50000m^{-3}$ was found to be needed for effective removal in the 5 σ region and below, but such strength limits the off-sets which can be handled to 15-10 σ. Thus, a balance has to be struck and methods have to be devised to either prevent beams with an offset larger than 10-15 σ from entering the FODO lattice, or to collimate such beams with different means.

Figure 12 illuminates this point [1] . It was again calculated with an Octupole strength of $k_3 = 43200m^{-3}$, as used before, corresponding at 10 GeV to an Octupole 10 cm long, 1 cm in diameter for the pole tips, and a pole tip field of 3 kG. The three distributions correspond to an initial off-set of 5, 10, and 15 σ; that is, in real dimensions a beam would enter the

[1)] In as much as the errant beam in Figure 12 corresponds to one macro particle of the Halo, it shows very clearly and in a simplified way that the Octupole collimation will remove the Halo nicely down to about $4-5$ σ, and, importantly, that it will distribute the collimation over some length. The question still to be answered is: what will Octupole collimation do to the core which contains 99% of the beam. Or, will the cure be worse than the disease?

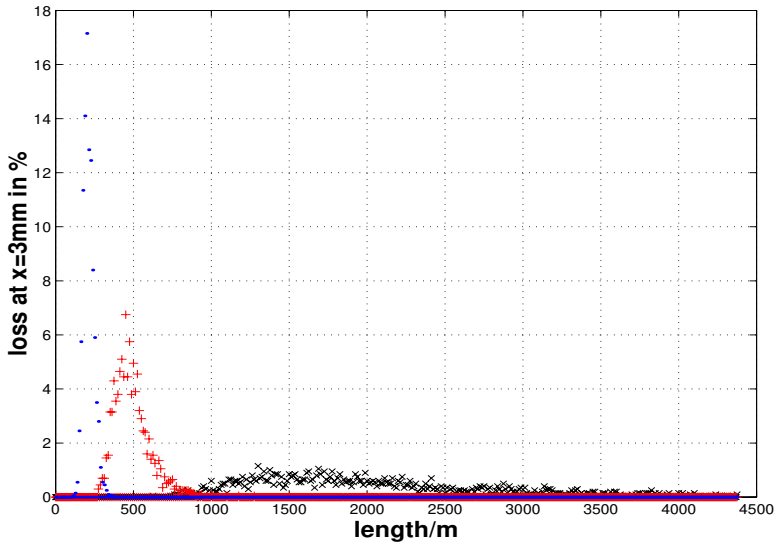

FIGURE 12. Pattern of loss of an errant beam along the linac with an Octupole strength of $k_3 = 43200m^{-3}$ and a 3 mm radius absorber. The 3 distributions correspond to an initial offset of 15, 10, and 5 sigma, as shown from left to right. As will be shown in the treatment of the linac case below, the strength of the Octupoles has to be sufficiently high to prevent new background beyond the beam stay clear, but within the collimation radius, to be created from tails close to the core. But is has to be sufficiently weak to not transport particles from the core into the tails. Therefore, beams with off-sets larger than about 10 σ will have too large a loss in too small a length of linac and do damage independent of the enlarged size. Beams with an off-set larger than 10 σ, therefore, have to be prevented by other means.

main linac with about 0.4, 0.7, or 1 mm offset. It is clear from this figure that 10 σ is about the limit the system could handle with Octupoles of this strength, because at 10 σ there is already one cell where 6% of the beam is being lost.

Figure 13 shows dramatically that the collimation can be tuned very well by changing the strength of the Octupoles. Using weaker Octupoles, as explained in the caption, protection against errant beams with much larger deviations is possible but at the price of collimation efficiency of the tails.

And finally the last figure in this series (Figure 14) makes yet another point, showing how the beam size grows progressively along the FODO lattice, and how the loss of the beam is distributed over many Copper absorber-Quadrupole-Octupole assemblies, as described above. Note that the losses are (should one say naturally?) largest where the beam is largest, reducing the destructive impact because in locations with sizeable losses the beam area is $5000\mu\text{m}^2$ or larger, a growth of 50 or more over the initial condition.

IV COLLIMATING THE HALOS

A Origins of Halo

For the **beam Halo** one must investigate where and by which processes the Halo is most likely produced, if the Octupole based systems are effective in removing the Halo, and if Octupoles can be placed in locations where the Halos should be removed.

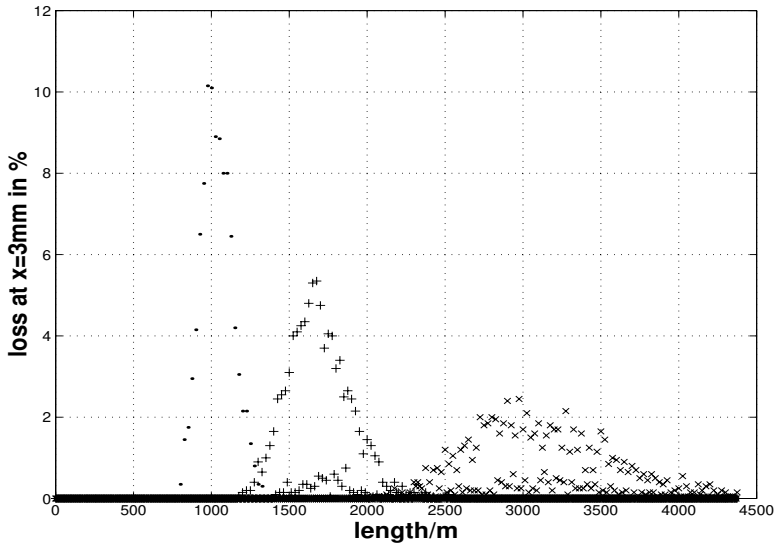

FIGURE 13. Pattern of loss of an errant beam along the linac with a 3 mm radius absorber. The 3 distributions correspond to an initial offset of 25, 20, and 15 sigma, as shown from left to right. As will be shown in the treatment of the linac case below, the strength of the Octupoles has to be sufficiently high to prevent new background to be created from tails close to the core. This example has a strength $k_3 = 4320m^{-3}$. Here, beams with off-sets larger than about 20 σ will be lost in too small a region.

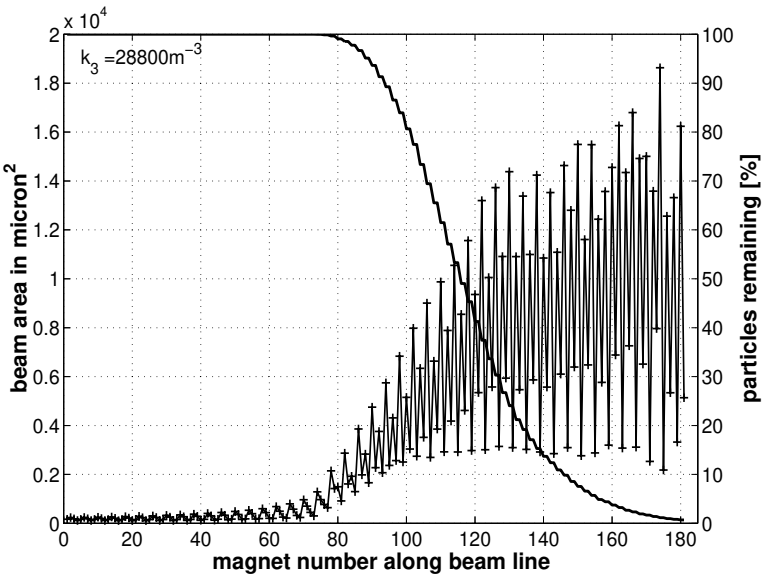

FIGURE 14. Development of the size, and the corresponding loss, of an errant beam along the linac with a 3 mm collimation radius as defined above. This example has been calculated with an Octupole strength of $k_3 = 28800m^{-3}$. The smooth curve describes the beam percentage still remaining, and the ragged curve the corresponding beam area in μm^2. One realizes from Figures 10 and 11 that most of the size enlargement happens in the horizontal. So in principle the beam has the shape of a horizontal knife slicing into material it encounters, however, now the average angle $\langle$x'$\rangle$ becomes important, spreading the interaction area over a large distance.

Experience with LEP and SLC has identified the following sources for Halo background:

- 1. The Damping Rings
- 2. Bunch Compressors
- 3. Misalignments, including steering errors in the linac
- 4. Mistakes in the multi-bunch energy compensation
- 5. Dark Current Acceleration in the linacs
- 6. Wakefields in the linac, especially after BNS, if no RF Quadrupoles are used
- 7. Scattering of Beam Gas anywhere
- 8. Compton scattering of black-body radiation anywhere [18]

Clearly, items 1 and 2 can be collimated, with or without Octupole collimation, in the pre-linac and /or the transfer lines from the injector complex to the pre-linacs. Good steering and control will minimize item 3 and 4; item 5 is assumed to be small because off energy mismatch (over focusing in the Quadrupoles sine the main linac starts with 10 GeV); item 6 is also minimized in the front end with BNS, but all of the Halo produced in the first 3/4 of the main linac (items 3 to 6) will be reduced with Octupole collimation in the linacs. This leaves Halo produced in the last part of the linacs, when one goes off BNS damping (and in the absence of RF Quadrupoles or "lossy RF cells" (see reference [5], section 7.4 [1]), and finally Halo produced by sources 7 and 8 at the end of the linac(s) and in the beam delivery systems. Consequently, the mechanisms which produce 7 and 8 should be minimized by minimizing the length of the BDS system itself. And the length of BDS can be reduced if the Halo transferred form the main linac can be reduced.

B The initial condition chosen

In Ref. [4] simple Halo Rings in real transverse space x and y, without a core, were chosen to be transported through the FODO lattices and the final focus.

The x' and y' variables were randomly chosen and attached to a x,y pair, as shown in Figure 15. There was much discussion and criticism of the "Halo Ring" concept used to explore the FODO lattices of the NLC linacs and beyond. The argument and criticism is that by using circles in x,y one chooses automatically a certain (and not random) set of phases in phase space. It is not clear that this really makes a difference for the Octupole collimation case which, in contrast to passive mechanical systems, collimates in all phases and both planes.

The choice of the precise shape of the initial background being transported naturally is critical for the quantitative results in any collimation scheme, but this choice is even more critical for magnetic elements where one needs to quantitatively evaluate what percentage

[1] The "classical" BNS requires a larger energy spread in the linac than the momentum acceptance of the final focus can handle, and, therefore, the energy spread and with it the BNS transverse damping has to be tapered down toward the end of the linac. This is when the tails are (re-)populated from wakefields in the absence of damping. This classical BNS can be replaced by BNS damping through "direct" head-to-tail energy compensation within the bunch with RF Quadrupoles, which method does not require to forego damping at the end of the linac(s). Do not think this is a free lunch: the alignment tolerances of the RF Quadrupoles are stringent with respect to the magnetic Quadrupoles.

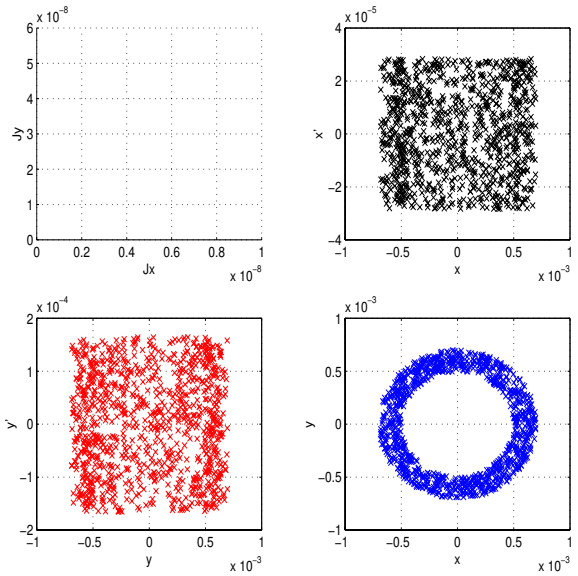

FIGURE 15. Patterns of Halo particles created from a concentric Halo Ring in real space, combined with random x' and y'. Subfigure a is empty but carried along to allow direct comparison with action variable plots later. Subfigures b and c show the particles used in the x,x' and y,y' space. Subfigure d finally shows the Halo ring, from which the phase space distributions are derived, in real space. Since the phases are randomly chosen, the phase space distributions are filled in. The next Figure shows that the magnitude of the distributions are the same, but that with action variables the phase space envelope is elliptic (as it should be) and not a rectangle as here.

of the core distribution might leak into tails from the 3 or 4 σ region of the core. The SLC experience [1] made the experimenters conclude that the background was "flat", but flat was not quantitatively defined [1] .

For the simulations discussed further below amplitude and action angle variables (J, ϕ) of the linear system were used. These are related to the physical transverse coordinates via

$$x(s) = \sqrt{2\beta(s)J_x}\cos\phi(s) \tag{1}$$

$$x'(s) = -\sqrt{2J_x/\beta(s)}\,(\sin\phi(s) + \alpha\cos\phi(s)) \tag{2}$$

$$y(s) = \sqrt{2\beta(s)J_y}\cos\phi(s) \tag{3}$$

$$y'(s) = -\sqrt{2J_y/\beta(s)}\,(\sin\phi(s) + \alpha\cos\phi(s))\,, \tag{4}$$

where s is the position along this model linac.

The Figure 16 shows the distributions which were constructed in analogy to the rings of Figure 15, but starting from action variables. Indeed the dimensions of the phase space distributions are very similar in both cases, Figures 15 and 16, albeit more rounded in Figure 15. This has the important consequence that the core collimation, at the same Octupole strength, starts later in σ-space.

The equivalent of "Halo Rings" (this time with a core indicated) in the x and y space of Ref. [4] then will be "stripes in J_x, J_y" space, as shown in Figure 17.

[1] The background, e.g., could be "flat" in dx dy or in dr dϕ. But for quantitative purposes a Gaussian width of the background of 10 times the beam width was used [1].

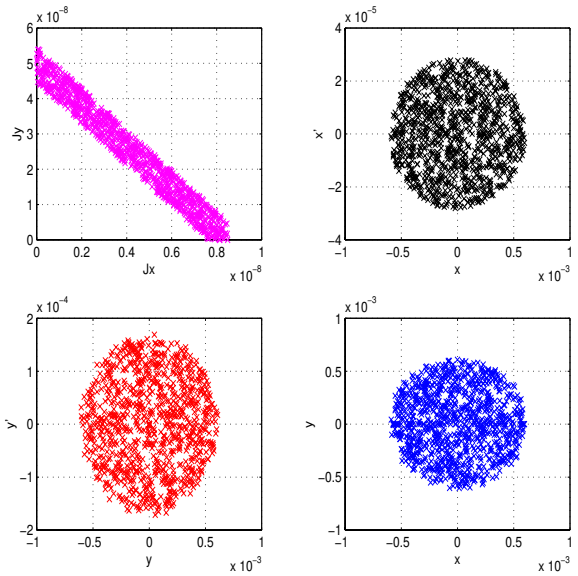

FIGURE 16. Patterns of Halo particles created with random phases and constant sum of amplitudes, designed to replace the "ring" of Figure 15. The $J_x + J_y = const$ particle stripe in subfigure a corresponds to the Halo Ring explained before. Corresponding to the increase in radius of a Halo Ring would be the movement of the stripe diagonally to the upper right, as "const" is increased. Subfigures b and c show the relation of cores and Halos in the x,x' and y,y' space. Subfigure d finally shows the equivalent of Halo Ring in real space. Since both amplitudes and phases are randomly chosen, the x,y distribution is also filled in, and no longer a ring.

The stripes correspond to a constant sum of amplitudes. The phases in the above equations were randomly generated. This makes sense because Octupole collimation works in all phases.

In summary, the change from "Halo Rings" in real space to stripes in "J_x, J_y" space did not make much difference in principle, but it shifted the on-set of collimation further out. It also allowed a cleaner parameterization of the tails. Namely, in a further step looking at stripes (and Halo Ring) was given up altogether and the focus was put on calculating the effect on the tail as a whole.

This total tail was made 1% of the central distribution. Following the SLD experience, it was given a Gaussian distribution with a 10 times wider σ. It starts at σ=3.29 (=0.1%) in σ_x units for both x and y. In radial distributions the data are shown in Figure 18, and with the phase space templates used before, in Figure 19.

V POSSIBLE SOLUTION FOR THE DIFFERENT LOCATIONS

A General

The demands on Octupoles as collimation elements are different for the S-band pre-accelerator, the X-band main accelerator, and, if it is needed, a post collimation section. What is determining the possibilities is simply the radius of the Octupoles allowed, and what

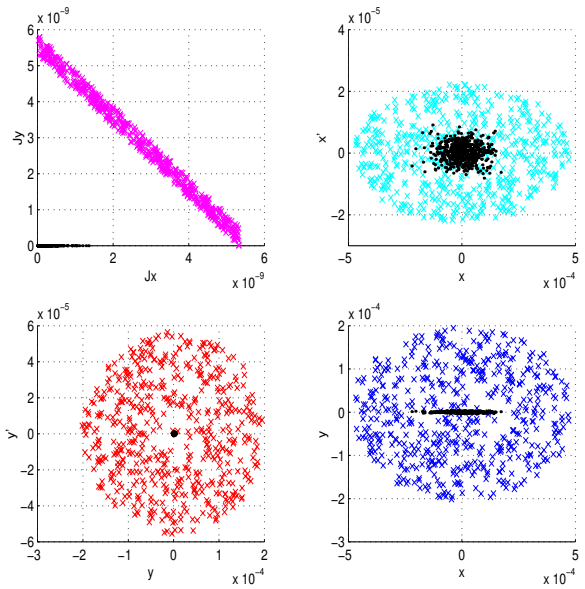

FIGURE 17. Patterns of core and Halo particles created with random phases and constant sum of amplitudes. Subfigure a shows the flat "core" in the lower left corner. The $J_x + J_y = const$ particle stripe corresponds to the Halo Rings explained before. Corresponding to the increase in radius of a Halo ring is the movement of the stripe diagonally to the upper right, as "const" is increased. Subfigures b and c show the relation of cores and Halos in the x,x' and y,y' space. Subfigure d shows the equivalent of Halo Rings in real space. Since both amplitudes and phases are randomly chosen, the "rings" are filled in (and are no longer rings). Since the "Halo" is constructed to be symmetric in space, but the core beam is flat, the design Halo to be collimated touches the x-core, but stays well clear of the y-core distribution.

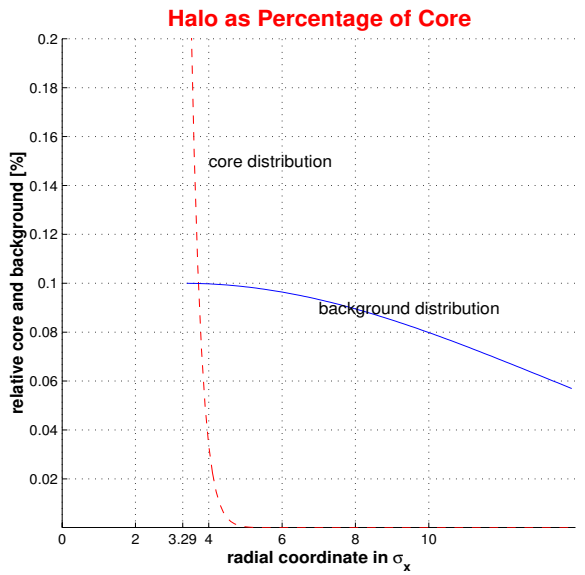

FIGURE 18. The construction of the SLC/SLD-like Halo attaches a radially symmetric 1% Halo to the 100% core at a point where the core is 0.1% of its maximum value.

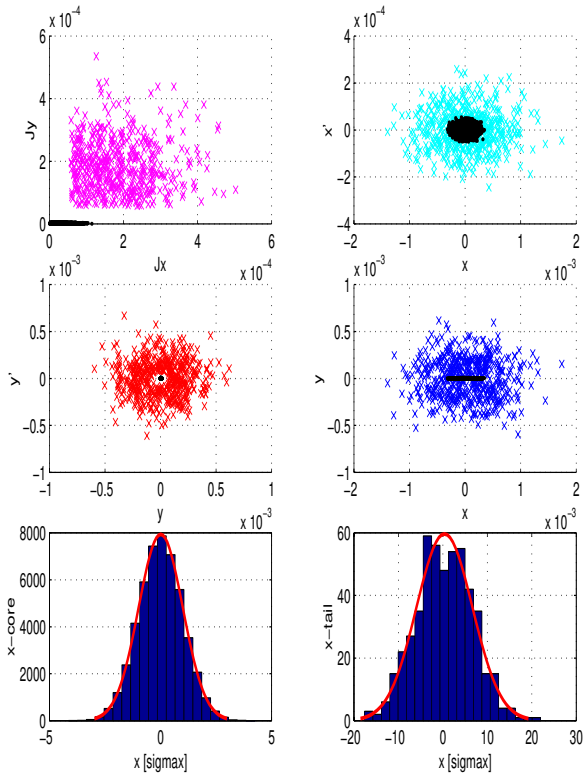

FIGURE 19. Patterns of core (50,000) and Halo particles (500) created with the method explained above, for the parameters of the pre-linac case. Subfigures a to d correspond to Figure 17. Subfigure a shows the flat "core" in the lower left corner and the Halo in $J_x + J_y$ space above it. Note again, as described in the text, that the Halo corresponds to an exponential distribution starting at (the equivalent of) 3.29 σ in both x and y, and that the width of the Halo distribution is 10 times the width of the core. Subfigures b and c show the relation of cores and Halos in the x,x' and y,y' space. Subfigure d shows the equivalent of Halo Ring in real space around the flat core beam. Since the amplitudes and phases are randomly chosen, the "rings" are filled in, as before. Since the "Halo" is constructed to be symmetric in space, but the core beam is flat, the Halo phase space touches the x-core in b, but stays well clear of the y-core distribution in c. Subfigure e shows a histogram of the core in transverse x-coordinate, and f finally shows the tail in x. As expected the core has a width of 1 σ and the tail of 10 σ. Note further, that the tail as a projection of a circular distribution onto the x-axis seemingly has a dimple in the middle, as it should.

is driving the demands is the size of the beam to be collimated. Some of these parameters are, within limits, arbitrary.

Since the Octupole fields in a magnet go down with the third power of the radius clearly a small beam is much more difficult to collimate. In the front, at 10 GeV, a 10 cm long Octupole was found to easily do the job, at a length comparable to the Quadrupoles foreseen. If one would use strict scaling, Octupoles at the TeV location would have to be several times longer than the quadrupoles in these locations.

However, two mitigating strategies were found to work:

- (A) The system worked well with Octupoles not placed at every Quadrupole location. Even though one has to watch out for non-linear scaling laws, the Octupole strength integrated over many elements does play a role.

- (B) The system also worked well with Octupoles progressively weaker along the linac (see table). For the main linac a reduction in just 1/2 of one percent from magnet to magnet ($B_R = B_{n+1}/B_n = 0.995$) yields an overall reduction of a factor more than 5. This method naturally requires limiting the production of Halo toward the end of the accelerator (see discussion about BNS and RF Quadrupoles above).

Here are the parameters used for the simulations shown below:

Parameters	Pre-linac	Main-linac	Post-linac
# of Cells	52	350	24-36
L_{Cell}/m	3.6	12.5	30-50
E_o/GeV	2	10	500
σ_x / μm	80	66	14-19
R_{oct}/mm	8	5	5
B_{oct}/kG	4	3	≈ 10
L_{0ct}/m	0.4	0.1	10-20
$B_R = B_{n+1}/B_n$	0.99	0.995	1
$B_R^{\#Cells}$	0.6	0.17	1
R_{coll}/mm	3	3	1.5

Table of Octupole Parameters: Emittances used were $\gamma\epsilon_x = 4 \cdot 10^{-6}$m-rad in all cases and $\gamma\epsilon_y = 6 \cdot 10^{-8}$m-rad for pre- and main accelerator, and $= 10 \cdot 10^{-8}$m-rad for the post-collimation after the main linac.

The rationale for some of the parameters were existing choices or designs, they are not necessarily optimal for physical implementation of an Octupole Collimation scheme.

For the pre-accelerator, e.g., the inscribed radius of the Octupole, 8 mm, was chosen to agree with the then assumed radius of the Quadrupoles. Since, on the other hand, the simulations were done with a collimation radius of 3 mm, this Octupole radius might be unnecessary large.

For the post accelerator collimation, if such a collimation is found to be necessary at all, the integrated strength has to be considerable, because of the small beam size and the high energy. Cold bore superconducting magnets should be considered, if they can be built strong enough, and because that would eliminate the background produced by Compton scattering of thermal photons [18]. The rule of thumb already for Quadrupoles is that below a radius of 8 mm higher gradients can be achieved with permanent magnets than with superconducting

ones. This rule must be even more stringent for higher multipole magnets. So because of the large strength needed, permanent magnets might be required.

But the fine details of the parameter choice is not essential for the validity of the results and has to be fine-tuned in any case if this concept is to be implemented.

For the purpose of this study, which is, repeat, only an exploratory conceptual investigation of the validity of such a collimation scheme, the demonstration of effectiveness of collimation through either showing expanding stripes in J-space, or showing the effect on the total 1% tail, is solely a matter of graphical representation. They are redundant. In the following section, therefore, only halo stripes are shown for the main linac, total tails for the pre-linac and post-linac. They are equivalent.

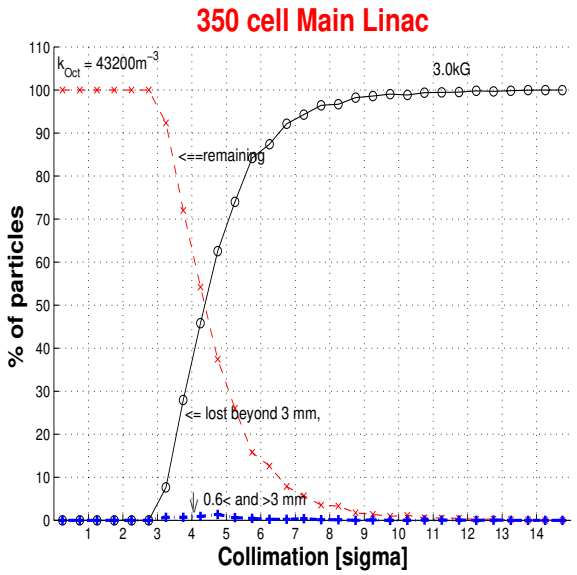

FIGURE 20. This main linac figure, as the next two, shows three curves. The calculations were done with a granularity of 0.5 σ in x, that means distributions were constructed which correspond to amplitude sums (rings) centered at 0.25, 0.75, 1.25 σ etc., with a width of 0.5 σ each. These distributions of particles then were passed through the modified FODO lattice. The rising curve shows the percentage of particles lost out of each stripe (=ring). A very important aspect is that most the particles "lost" out of the "ring" were brought to a position beyond 3 mm, the "collimation radius". The ones which were not lost are shown on the second, the bottom, curve. These latter particles are potentially dangerous, because they could hit material in the BDS and produce background. The third curve, of remaining particles, falling from 100% at the left, finally shows another important aspect: the inner 3 σ of the distribution are not touched. Even with 10^6 particles in the area between 2.5 and 3.0σ, not a single one was removed from this area. This is even more pronounced in the next two Figures, which show results with weaker Octupoles.

B The Main Linac

After this lengthy "introduction", a description of how the collimation in various parts of a collider actual works, and what it requires on accommodation in the design, is relatively simple. The main linac is dealt with first, although the pre-linac might be as good or better a place for application of this scheme. Because of it's large length and the many FODO elements, resonant FODO-Octupole collimation works especially well, and with low strength

Octupoles, in the main linac. The calculations used the constant energy principle explained previously (10 GeV).

The next 3 Figures (Figures 20 to 22) will show 3 graphs for "Halo Ring"-equivalents in action variable space. Although the total tail method is more instructive regarding bleeding out of the core into the tail, the "ring" representation is more obvious. The collimation was simulated with Octupoles with a length comparable to the Quadrupoles used (10 cm in the beginning of the linac) and pole tip fields of 0.3, 0.1, and 0.03 Tesla. Such magnets are technically straightforward to build. Equally important, these Figures show that the collimation is highly tunable with the strength of the Octupole fields. [1]

Along the linac(s) two technical solutions come to mind: (a) electromagnetic iron pole magnets and (b) permanent magnets. Since tunability is important, (a) seems to be the preferred solution, except, that by designing the Octupoles such that they can be rotated (b) is also tunable. This choice needs to be investigated more closely.

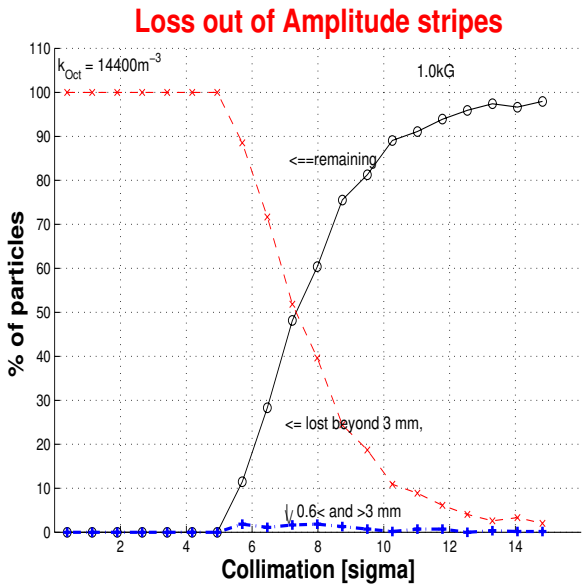

FIGURE 21. This main linac figure for $k_3 = 14400m^{-3}$, as the one before and the one after, shows the three curves described before. Again the rising curve shows the percentage of particles lost out of each ring. Important is, what percentage of the "lost" particles were brought to a position beyond 3 mm, the "collimation radius", and what percentage was not, as shown on the bottom curve. These latter particles are dangerous, because they could hit material in the BDS and produce background, but compared to the previous Figure there are only slightly more of them, even in the relative plot. But since we have calibrated the figure to percentages, and it is assumed that out beyond 5 σ far fewer particles exist in the Halo, this should not matter. Most important is that last curve, falling from 100% at the left, shows the inner 5 σ of the distribution not touched. That means that at this Octupole strength only a σ space is collimated where one assumes that there are no core-like densities. In other words: the core does not bleed into the tails.

[1] The only Octupole installed in the linac system of the SLC (in the North Damping Ring to linac line (NRTL)), even though installed for other reasons, brought the background collimated at the end of the linac in sector 30 down by a factor of 10 [19].

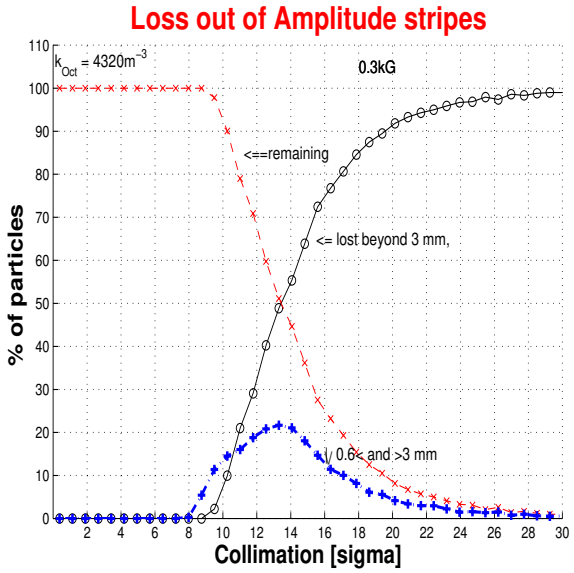

FIGURE 22. This main linac figure for $k_3 = 4320m^{-3}$, as the two figures before, shows three curves. It is important to realize is here that only particles outside of 8 σ are being collimated, a σ-space where there are no core-like densities.

C The Pre-linac

The 52 FODO cell pre-linac is possibly the most logical choice for an Octupole collimation. Collimation there would remove background from the damping rings when the energy is still relatively low. The calculations used the constant energy principle explained before (here 2 GeV); the NLC standard FODO lattice length of 3.1 m for the pre-linac at 2GeV had to be enlarged to 3.5 m to make space for the Octupoles. The space needed for the Octupoles can be reduced by making some other choices for the diameters of the Octupoles, collimator beam pipes, and other vacuum chambers between the acceleration sections.

The initial conditions chosen were identical to Figure 19, but the calculation were done using the pre-linac parameters. The situation after collimation is shown in Figure 23. This is a busy figure, but captures the quintessence of the method in one graph, so careful study of the picture and the caption is recommended.

The original design report [5] had the injector complex close to the beginning of the main linacs. More recently it has been recommended that the injectors be centralized. This then requires long transport lines which are ideal for Octupole collimation, because they will require for emittance preservation relatively short, and therefore many, FODO lattices which, as outlined before, makes the non-linear process easy.

In addition, the Dodecapole "collimation" discovered accidentally in the SLC positron transport line [13], could be turned into a virtue. The Dodecapole moment of a Quadrupole can be enhanced by symmetry plane shimming, and the Halo collimation could be nearly cost free if it works with Dodecapoles.

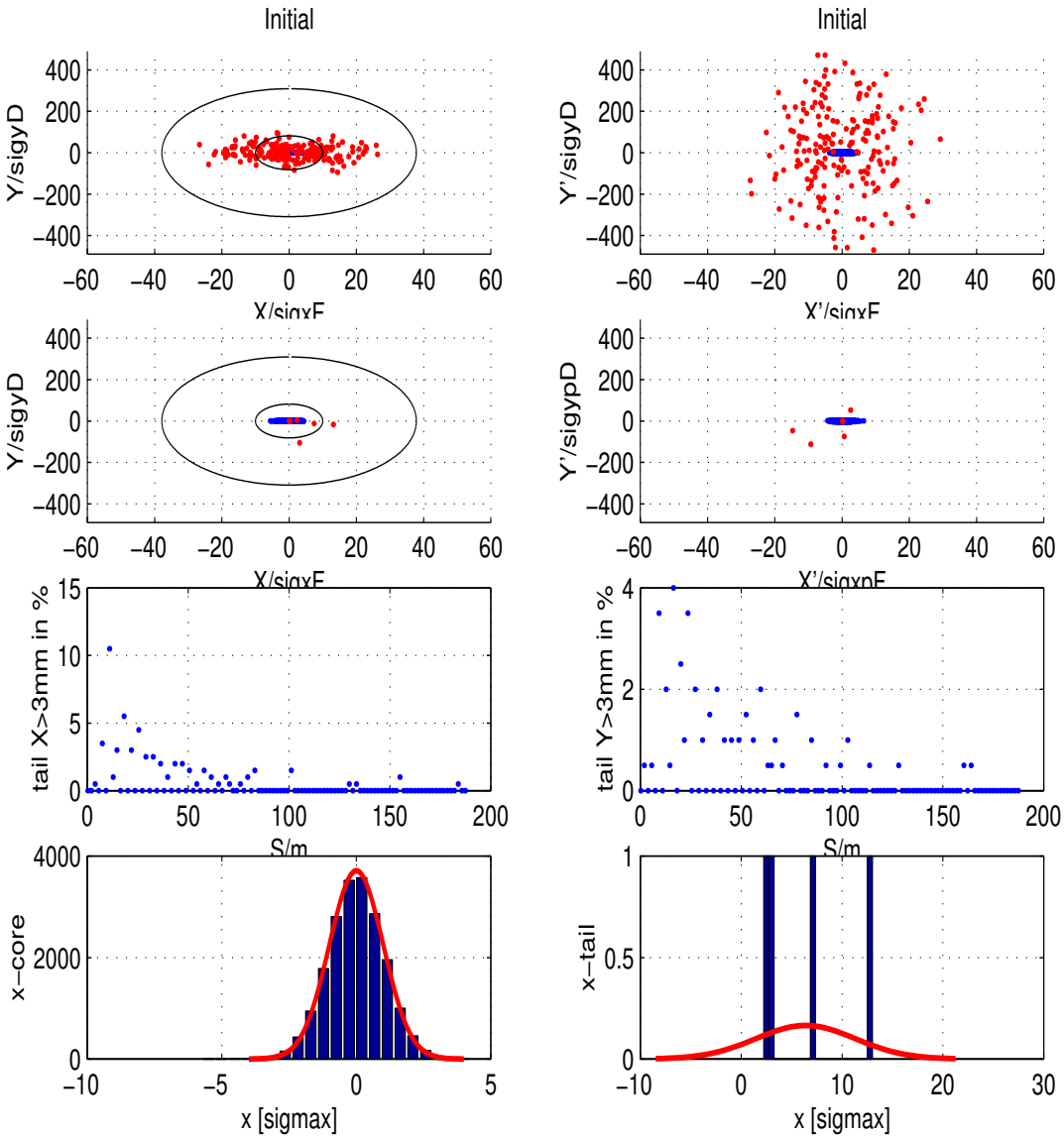

FIGURE 23. This Figure for the pre-linac is different from the ones shown before, in as much as it shows the initial distribution in x-y and x'-y' space, (subfigure a and b), the same parameters after collimation (subfigures c and d), the loss process along the accelerator in both x and y (subfigures e and f), and the x-core and x-tail after collimation (subfigures g and h). The core (20000) and Halo particles (200) were created with the total tail method explained above. The circles in subfigures a and c are the beam stay clear (inner circle, here a 10 σ radius) and the collimation radius used (3 mm radius, outer circle). As explained earlier in the text the main reason for doing the simulation with a total tail approach, rather then the Halo Ring approach, is to test graphically if there is bleeding from the core into the tails. None did. One also has to test if particles get beyond the BSC without being removed beyond 3 mm. 2 out of 200 (2%) did. Note again, as described in the text, that the Halo corresponds to an exponential distribution in J-space starting at (the equivalent of) 3.29 σ in both x and y, and that the width of the Halo distribution is 10 times the width of the core.

D The Post-linac

From all the systems treated in this study, the post-collimation is the one which has least converged towards a unique solution. There are several reasons for this. One reason is that the beams are very small, in the μm region only, so that one needs very strong integrated Octupole strengths, which could lead to the core leaking into the tails. In addition, the real estate here is not quasi-parasitically used for collimation, as it is in the pre-linac and main linac scenarios. So there is pressure to build a system as short as possible, which again points to an Octupole system as strong as practical.

The picture evolving from using different Quadrupole lattices is not yet clear, but a solution along the situation shown in Figure 24 seems possible. This figure shows that in principle an effective collimation can be reached with a system length of about 1800 m. What is worrisome, and something were no clean solution has yet been found, is that with the large integrated Octupole strength required, the collimation starts to eat into the core distribution, as shown in subfigure g of Figure 24, although not badly yet. This needs more investigation.

Further studies will have to attempt to play the resonant effects in a longer system vs. the sheer strength needed in a shorter system This approach, after initial frustration, led to success in the main linac.

Subfigure a of Figure 24 shows the initial distribution of the tail in x,y (real) space in the first magnet of the first FODO cell (the 0^{th}) magnet. The next two Figures (Figures 25 and 26) show the same distribution in x and y in the third FODO cell Quadrupoles (magnets named number 4 and 5). The magnet in Figure 25 is a Quadrupole focusing in the horizontal, the next in Figure 26 is focusing in the vertical. What is spectacular here is that already here Halo particles are shown to be outside the collimation radius; that is, they are being collimated.

And the final result is shown in subfigure b of Figure 24 demonstrating that there is very little left of the Halo.

The intense studies of the problems of a post linac system had one good consequence. It produced Figures 25 and 26 which can help answer the often asked question: how does this collimation "really" work?

So this is how Octupole collimation really works: the quadrupole lattice drags the Halo alternatively in the x and in the y directions. Each time the Octupole gives the particles in the tails a little kick. Eventually these kicks add up and the particles get into a more non-linear region in the Octupole field. Then the process accelerates, and eventually the particles hit, at a relatively large angle in 100s of μrad, the collimator beam pipes. These collimators are designed round and further away from the beam than classical collimators, so wakefields do no damage to the emittance of the core.

VI THE FINAL FOCUS

The efforts to establish a good collimation system in accelerators have one goal in mind: to prevent more particles than a detector can handle to reach the Detector(s) in the Interaction Point, or target area.

In addition to collimating the tails upstream, there is another method known in the literature to reduce unwanted background: folding in the tails in a one-time effort just

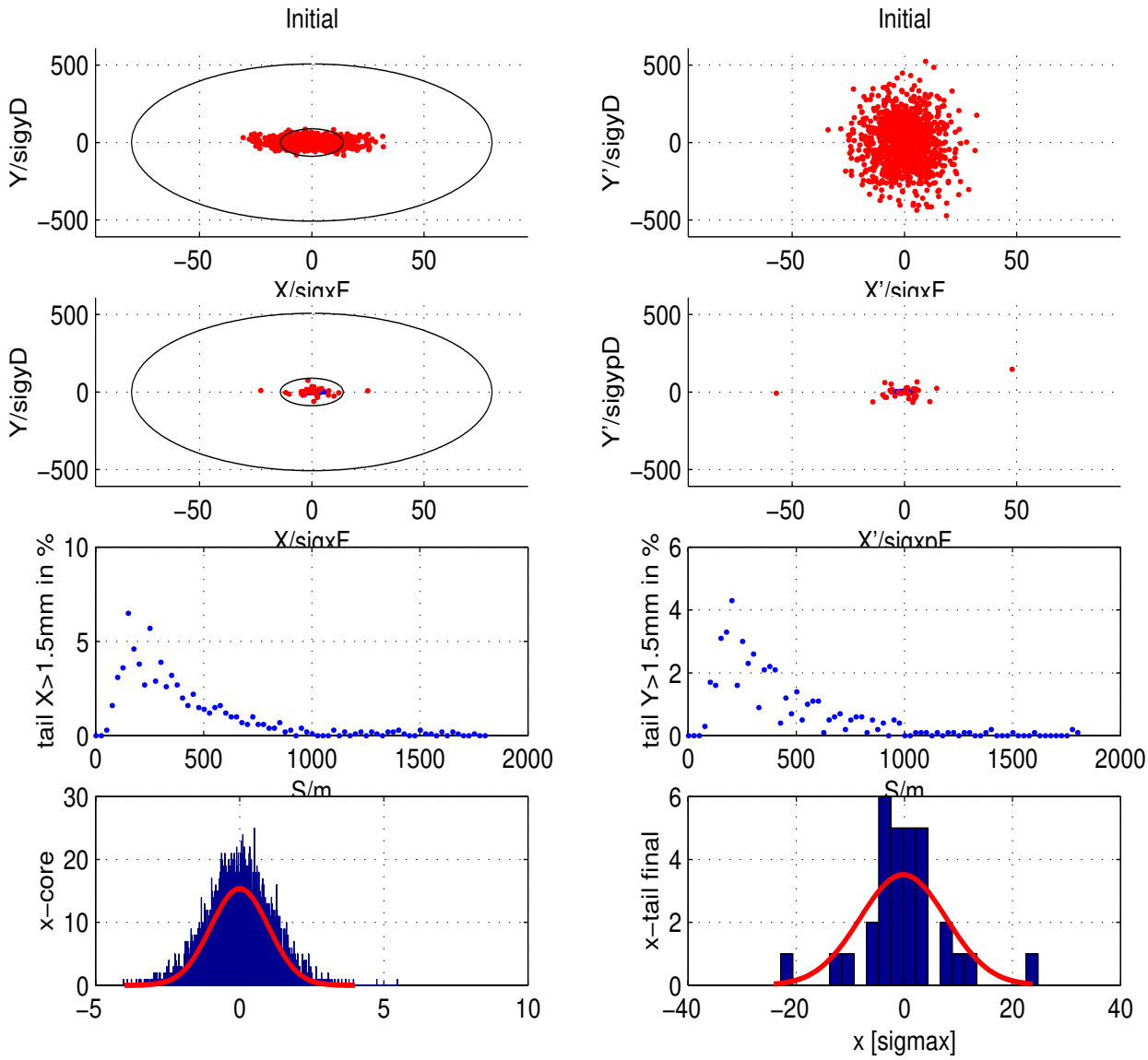

FIGURE 24. This Figure for the post-linac collimation is made with the same template as Figure 23, so the general explanations in the previous Figure apply. The core and Halo particles were created with the method explained above, but the number of core and tail particles was changed to 10000 and 1000, respectively, for better statistics with the tail simulations. The other parameters chosen were 36 FODO cells of 50 m length. The circles in subfigures a and c are the beam stay clear (0.25 mm, inner circle) and the collimation radius used (here a 1.5 mm radius, outer circle). Note again, as described in the text, that the Halo corresponds to an exponential distribution in J-space starting at (the equivalent of) 3.29 σ (equivalent to 0.1% in a Gaussian distribution) in both x and y, and that the width of the Halo distribution is 10 times the width of the core. The collimation of the post-linac beam is the most challenging for several reasons. The beams are small, the energy is high, and the space is not free, as in the other systems treated before.

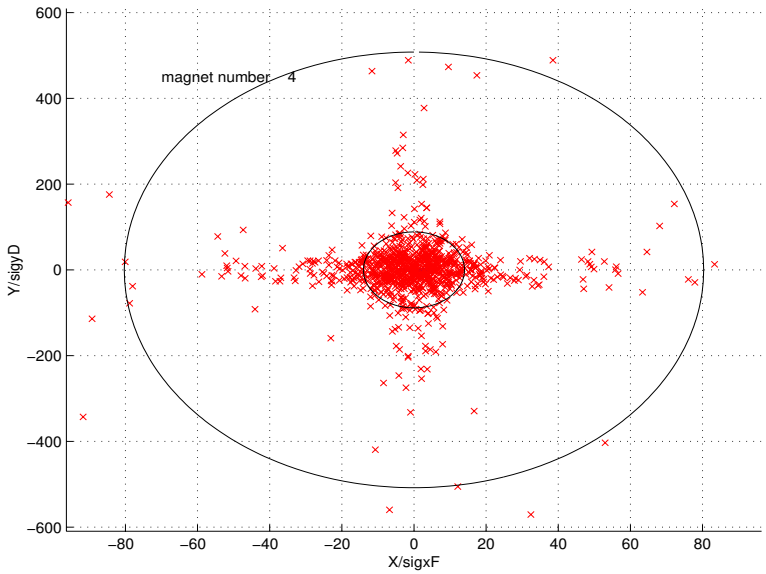

FIGURE 25. This Figure shows the real space distribution in the 1^{st} magnet in the 3^{rd} FODO cell, a horizontally focusing Quadrupole. Particles are spread out horizontally and vertically and are being collimated.

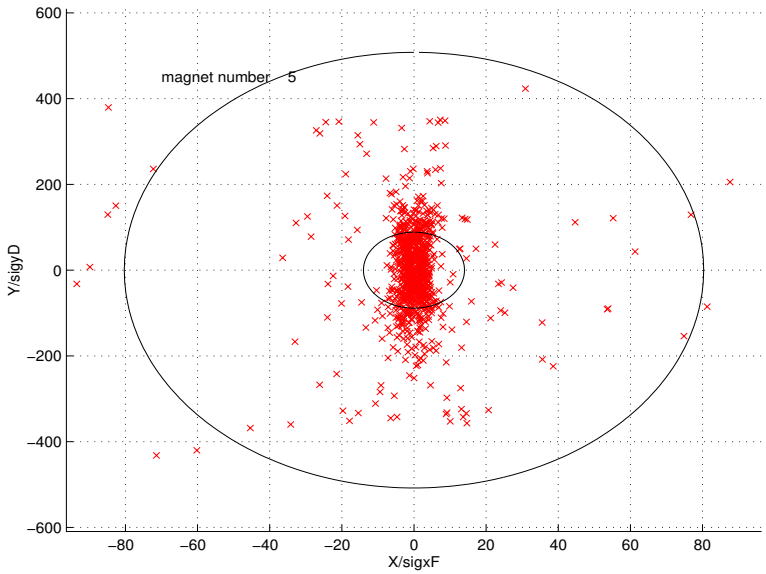

FIGURE 26. This Figure shows how the tail is being tugged into the vertical direction in the 2^{nd} magnet in the 3^{rd} FODO cell, and that the collimation continues. This progress is also evident from subfigures e and f of Figure 24.

before the target. This method was originally developed for a different goal, namely to illuminate a target uniformly (see [10] and references therein), shown in Figure 27. This was done successfully in both planes.

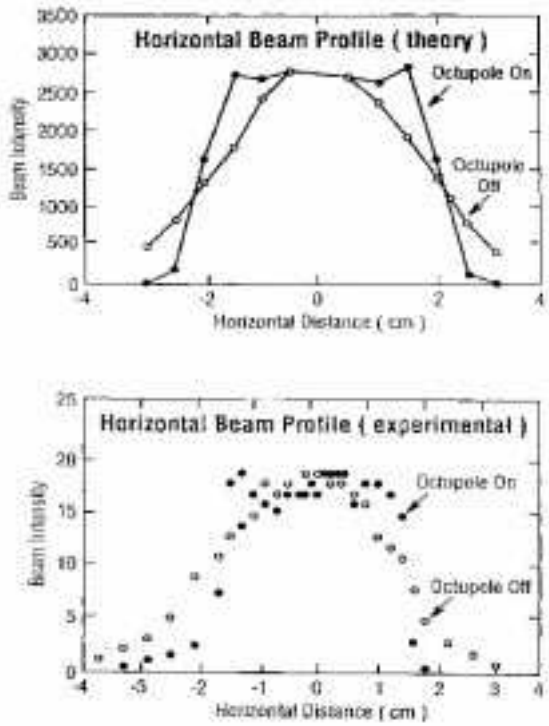

FIGURE 27. Comparison between theory and experiment for folding in the tails to illuminate a target uniformly from Reference [10].

If Octupoles are placed upstream of the target or the final doublet, as the case may be, it also can prevent tail particles from hitting apertures near or in the IP in both planes. This method increases the effective beam stay clear, which determines the required collimation depth up-stream. For the classical collimation a larger effective beam stay clear would imply that the spoilers and absorbers can be further away from the beam, reducing the wakefield effects, which are quite severe.

In the effort to illuminate a target uniformly the beam was focused in a special way to do something once, at one point only, in a beam line. Clearly the problem one has in a collider is more difficult: instead of focusing onto one point in z, the target, as in [10], one must minimize the tails in two different extended locations (the doublet quadrupoles) and the beams must be able to go through the opposite doublet on their way out. Nevertheless, analytical calculations [20] show that the effective beam stay clear can potentially be increased by a factor of 3. Ref. [20] found that the required Octupoles in front of the final doublet are very weak. This is understandable since the betatron functions (the beam sizes) in and around the final doublet are very large. The effects on beam spot size were shown to be negligibly small.

While this approach shows great promise, from the experience with the non-linear linac lattice, it will probably require a difficult and lengthy trial and error effort to find a valid solution.

VII THINGS NOT DONE (YET)

There are many things which could have been done, and should be done, to improve upon this study. The following should be regarded as a "to do" list if the non-linear element approach to collimation is to be continued.

Among things not done is introduction of acceleration instead of the constant energy principle, using thick lenses for the Octupoles instead of kicks, analyzing the collimated distributions in terms of J-distributions instead of just in real space, using Dodecapoles instead of or in addition to Octupoles. The role of the linear (Quadrupole) phase advance was not investigated in the depth required. The impact of spurious dispersion, and the limits this puts on geometrical straight line alignment, was not considered. And so on.

Because of the negative experience with MAD no attempt was made to transport the scheme to an existing beam dynamics program. This will have to be done, although it probably will require extensive modifications and trouble shooting of these programs. But the use of a standard program would make checks by others so much easier.

Acknowledgments
I want to thank David Burke and Tom Markiewicz from SLAC, and Jean-Pierre Delahaye and Kurt Hubner from CERN, for encouraging and indulging me in the pursuit of this endeavor.

REFERENCES

1. Stan Hertzbach, Tracy Usher, private communication 1998.
2. H. Burkhardt "Machine Background Common to all Machines," Sitges Conference, 1999.
3. B.V. Chirikov, "A Universal Instability of Many-dimensional Oscillator Systems", Physics Reports **52**(1979)263
4. P. Emma, R. Helm, Y. Nosochkov, R. Pitthan, T. Raubenheimer, K. Thompson, F. Zimmermann, "Nonlinear Resonant Collimation for Future Linear Colliders", proceedings of the 16^{th} ICFA Beam Dynamics Workshop on Nonlinear and Collective Phenomena in Beam Physics, Arcidosso, Italy, 1998; SLAC-PUB-7958.
5. C. Adolphsen et al., "Zeroth Order Design Report for the Next Linear Collider," *SLAC-Report* **474** (1996).
6. Dieter Walz, in "Linac Coherent Light Source", pp. 8-32, SLAC-Report 521, 1998, and private communication.
7. F. Zimmermann, "New Final Focus Concepts at 5 TeV and Beyond," 8^{th} Advanced Accelerator Concepts Workshop, Baltimore, 5-11 July 1998; SLAC-PUB-7883.
8. S. Fartoukh, "A Statistical Approach to Analyse the Efficiency of BNS Damping and Correction Algorithms in Linear Colliders", CLIC Note 323, 1997; G. Guignard, "Structure Scaling with RF Frequency and Betatron Scaling with Energy in CLIC", CLIC Note 188, 1993; W. Schnell, "Microwave Quadrupole Structures for the CERN Linear Collider", Proceedings of IEEE Particle Accelerator Conference, San Francisco (1991), p. 3237.
9. P.F. Meads, "A Nonlinear Lens System to Smooth the Intensity Distribution of a Gaussian Beam", IEEE Transactions on Nuclear Science, **NS30**, 1983, p. 2838.
10. N. Tsoupas, R. Lankshear, C.L. Snead,Jr., T.E. Ward, M. Zucker, H.A. Enge, "Uniform Beam Distributions Using Octupoles", Proceedings of IEEE Particle Accelerator Conference, San Francisco (1991), p. 1695.

11. N. Merminga, J. Irwin, R. Helm, R.D. Ruth (SLAC), "Collimation Systems for a TeV Linear Collider". SLAC-PUB-5165-Rev, Part.Accel. **48**(1995), 85-108.
12. N. Akasaka et al., "JLC Design Study", KEK-Report 97-1, 1997, p. 458pp.
13. Hans Braun, A.V. Kulikov, Rainer Pitthan and M.D. Woodley: "Nonlinear Effects in the SLC e^+ Transport Line", Proceedings of IEEE Particle Accelerator Conference, San Francisco (1991), p.1845.
14. A. Hofmann, private communication.
15. Karl L. Brown and Roger V. Servranckx, "First- and Second-Order Charged Particle Optics", SLAC-PUB-3381.
16. M. Cornacchia, W.J. Corbett, and K. Halbach, "Modified Octupoles for Damping Coherent Instabilities", Proceedings of IEEE Particle Accelerator Conference, San Francisco (1991), p. 1987.
17. P.J. Bryant, "Basic Theory for Magnetic Measurements", CERN Accelerator School on Magnetic Measurements and Alignment, Montreux, 1992.
18. I. Reichel, F. Zimmermann, P. Tenenbaum, T.O. Raubenheimer, "Thermal-Photon and Residual-Gas Scattering in the NLC Beam Delivery", ICAP 98, Monterey California.
19. P. Raimondi, private communication, 1999.
20. F. Zimmermann, "Octupoles in Front of the Final Doublet", NLC Beam Delivery Note dated July 14, 1998.