SLAC-PUB-3179
August 1983
(A)

BEAM-BEAM INSTABILITY

ALEXANDER W. CHAO*
Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94305

TABLE OF CONTENTS

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

Presented at the 3rd Summer School on High Energy Particle Accelerators
Brookhaven National Laboratory, Upton, New York, July 6-16, 1983

BEAM-BEAM INSTABILITY

ALEXANDER W. CHAO
Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94305

INTRODUCTION

The subject of beam-beam instability has been studied since the invention of colliding beam storage rings. Today, with several colliding beam storage rings in operation, it is not yet fully understood and remains an outstanding problem for storage ring designers. No doubt appreciable progress has been made over the years, but current knowledge is still rather primitive.

The subject of beam-beam interaction can be divided into two areas: luminosity optimization, and the dynamics of the interaction. The former is concerned mostly with the design and operational features of a colliding beam storage ring; the later concentrates on the experimental and theoretical aspects of beam-beam interaction. Although both areas are of interest, our emphasis will be on the second area only. In particular, we will be most interested in the various possible mechanisms that cause beam-beam instability. These have been reviewed more completely in Refs. 1-5.

1 THE STRONG-WEAK PICTURE

Figure 1(a) shows a storage ring in which two oppositely charged particle bunches circulate and collide at two opposite locations. As our first beam-beam picture, let us assume that one of the bunches consists of a single particle (weak beam), while the other bunch is a dense Gaussian charge distribution (strong beam). The strong beam is assumed to be a smooth cloud of charge rather than a collection of many point charges.

As the weak beam passes through the on-coming strong beam, it receives a transverse impulse. The strong beam, on the other hand, is unperturbed. In this "strong-weak" picture, the weak beam acts as a "probe" into the beam-beam force of the strong beam. The issue here is whether the weak beam motion will be stable under the beam-beam perturbation.

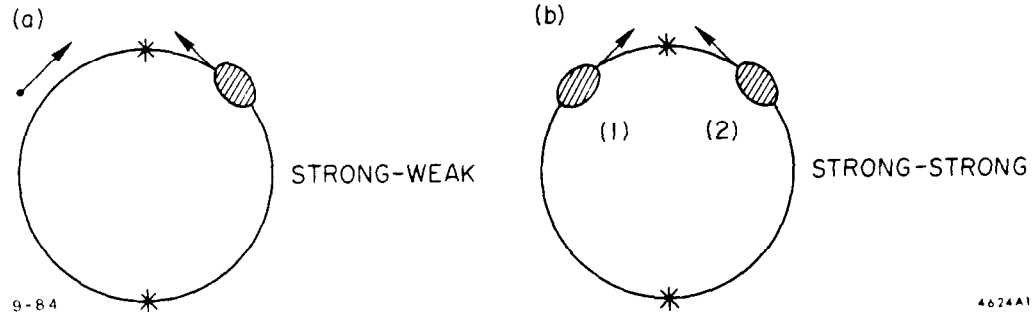


Fig. 1. The "strong-weak" and "strong-strong" cases of beam-beam interaction.

More specifically, let us specify the weak beam by its horizontal and vertical displacements, x and y, and their slopes x' and y'. Then, as the weak beam passes through the strong beam at the collision point, its displacements are unchanged but its slopes change by amounts related to x and y according to[6,7]

$$\Delta x' = -\frac{\partial U(x,y)}{\partial x} \qquad \text{and} \qquad \Delta y' = -\frac{\partial U(x,y)}{\partial y} \tag{1}$$

where $U(x,y)$ is an equivalent potential-well produced by the strong beam space charge and current,

$$U(x,y) = -\frac{Nr_0}{\gamma}\int_0^\infty dt \left\{\exp\left[-\frac{x^2}{2(\sigma_x^2+t)} - \frac{y^2}{2(\sigma_y^2+t)}\right] - 1\right\} \Big/ \sqrt{(\sigma_x^2+t)(\sigma_y^2+t)} \tag{2}$$

with N the number of particles in the strong beam, $r_0 = e^2/mc^2$ the classical radius of the particle, γ the Lorentz energy factor of the weak beam, and σ_x and σ_y the rms beam sizes of the strong beam at the collision point.

Problem 1. Show that when the beam is round, with $\sigma_x = \sigma_y = \sigma$, Eqs. (1) and (2) reduce to

$$\Delta r' = \frac{2Nr_o}{\gamma}\cdot\frac{\sigma}{r}\left[e^{-r^2/2\sigma^2} - 1\right]. \tag{3}$$

The problem is one-dimensional in nature. Eq. (3) can also be obtained directly from Gauss' law and Ampere's law.

After each collision, the weak beam executes a free betatron oscillation with its (x, x', y, y') being transformed linearly by the matrix[8]

$$\begin{bmatrix} \cos\mu_x & \beta^*_{x0}\sin\mu_x & 0 & 0 \\ -\frac{1}{\beta^*_{x0}}\sin\mu_x & \cos\mu_x & 0 & 0 \\ 0 & 0 & \cos\mu_y & \beta^*_{y0}\sin\mu_y \\ 0 & 0 & -\frac{1}{\beta^*_{y0}}\sin\mu_y & \cos\mu_y \end{bmatrix} \tag{4}$$

where μ_x and μ_y are the betatron phase advances from one collision point to the next, and β^*_{x0} and β^*_{y0} are the beta-functions at the collision point.

The weak beam motion is then described by a sequence of mappings on (x, x', y, y'), alternately representing free betatron oscillations and collisions — the former being linear and the latter nonlinear. The problem of the weak beam motion is thus equivalent to a mathematical problem of nonlinear mapping on the vector (x, x', y, y'), i.e.,

$$\text{beam-beam problem} = \text{nonlinear mapping problem} . \tag{5}$$

The question being asked is, again, whether the weak beam motion is stable under repeated application of the mapping procedure.

It should be emphasized here that the statement (5) is valid only if the strong-weak picture is adopted. Figure 1(b) shows the more realistic case of two strong beams. In this strong-strong picture, statement (5) is no longer valid. In this sense, (5) represents only a rather limited view of the beam-beam problem. We will postpone the discussions of the strong-strong case and this point until later.

2 LINEARIZED STRONG-WEAK MOTION

Some insight is to be gained by considering a weak beam that executes a small amplitude motion with $x \ll \sigma_x$, $y \ll \sigma_y$.[9] Equation (1) becomes

$$\Delta x' = -\frac{2Nr_0}{\gamma\sigma_x(\sigma_x+\sigma_y)}x \qquad \text{and} \qquad \Delta y' = -\frac{2Nr_0}{\gamma\sigma_y(\sigma_x+\sigma_y)}y . \tag{6}$$

In this case, the x and y motions are decoupled, and the problem is linear and can be readily solved by a matrix technique.

After linearization, we need to consider one dimension only. Let it be the y dimension. The transformation of the (y, y') vector through the beam-beam

collision is described by the matrix

$$\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \quad , \qquad \frac{1}{f} = \frac{2Nr_0}{\gamma\sigma_y(\sigma_x + \sigma_y)} \tag{7}$$

where f is the equivalent beam-beam focal length.

As a numerical example, a round beam with 10^{12} particles, 0.2 mm radius and 5 cm length produces at its edge a magnetic field of 10 kilogauss. If linearized, the beam-beam kick is equivalent to a quadrupole of gradient[1] 50 kG/mm, which is a strong gradient indeed. For a 10-GeV particle, the corresponding focal length is about 13 cm. This beam-beam quadrupole is of course focusing in both x and y planes.

For symmetry reasons, we will split the beam-beam kick in the middle and observe the weak beam there. The transformation from one collision to the next is then

$$\begin{bmatrix} \cos(\mu + \Delta\mu) & \beta^* \sin(\mu + \Delta\mu) \\ -\frac{1}{\beta^*}\sin(\mu + \Delta\mu) & \cos(\mu + \Delta\mu) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ -\frac{1}{2f} & 1 \end{bmatrix} \begin{bmatrix} \cos\mu & \beta_0^* \sin\mu \\ -\frac{1}{\beta_0^*}\sin\mu & \cos\mu \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2f} & 1 \end{bmatrix} \tag{8}$$

where β^* is the perturbed beta-function at the collision point, and $\Delta\mu$ is the perturbation on the phase advance.

As the weak beam circulates around, its (x, x', y, y') is transformed repeatedly by the matrix (8). If the net motion is stable, the matrix can be parameterized as Eq. (8) with the perturbed quantities related to the unperturbed ones through

$$\cos(\mu + \Delta\mu) = \cos\mu - \frac{\beta_0^*}{2f}\sin\mu \ ,$$

$$\frac{\beta^*}{\beta_0^*} = \frac{\sin\mu}{\sin(\mu + \Delta\mu)} \ . \tag{9}$$

---

[1]More correctly, the gradient should to doubled since the electric field also contributes. But then the length of interaction should be halved to 2.5 cm because both beams move and they move in opposite directions.

Note that in the linear approximation the entire problem is specified by the two scaling parameters μ (more specifically, μ modulus 2π) and

$$\xi_y = \frac{\beta_0^*}{4\pi f} = \frac{N r_0 \beta_0^*}{2\pi\gamma\sigma_y(\sigma_x + \sigma_y)} \tag{10}$$

where ξ_y is the well-known beam-beam parameter[9] that specifies the beam-beam strength. The beam-beam parameter in the x-motion is obtained from Eq. (10) by exchanging x and y.

Figure 2 shows the stable region in the (μ, ξ) space. Outside the stable region, the absolute value of the trace of matrix (8) is larger than 2; Eq. (9) then does not have a solution. The dividing boundary between the stable and unstable regions is

$$\xi = \frac{1}{2\pi} \cot\frac{\mu}{2} \ . \tag{11}$$

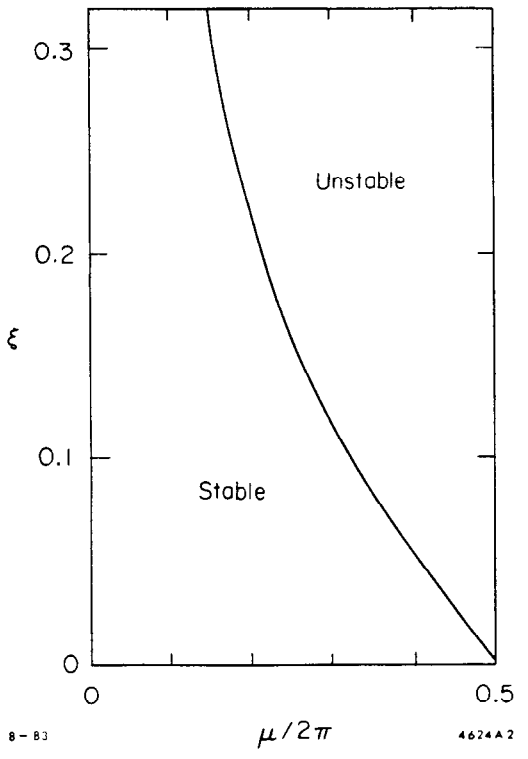


Fig. 2. Stability region for a weak beam executing small oscillations. μ is the betatron phase advance between collision points. The diagram repeats with period $\mu = \pi$ and shows a periodic sawtooth behavior which is typical of beam-beam models.

Figure 2 shows that the weak beam is most unstable if the tune advance $\mu/2\pi$ between collision points is slightly below a half-integer and is most stable if it is just above a half-integer. (Signs switch for two beams with same sign of charges.)

In case μ is not close to a multiple of π, the tune shift $\Delta\mu/2\pi$ is approximately equal to ξ if $\xi \ll 1$. This is why sometimes ξ is also identified (somewhat loosely) as the beam-beam tune shift per collision point.

The linear strong-weak model described above is our first beam-beam instability model. Unfortunately, it does not explain the observed beam-beam instability. It predicts much too high a value for the stability limit. Even worse, such linear instability can in principle be removed by simply readjusting the unperturbed storage ring optics and consequently can not constitute a fundamental limitation on the maximum beam intensity.

3 A STRONG-WEAK SIMULATION

In case the weak beam does not have small amplitudes, the linearization breaks down and the beam-beam force must be considered in its full glory. Such a nonlinear mapping problem is extremely difficult to handle analytically. We must then seek the help of the almighty computer. (See for example Refs. 10 and 11, which also contain interesting analytical techniques.) The hope is, of course, that by taking into account the nonlinear terms in the beam-beam kicks, we could readily explain the beam-beam instability.

To simulate the weak beam motion with a computer, we launch the beam with initial conditions (x_0, x_0', y_0, y_0') and apply the transformations (1) and (4) alternately and repeatedly. As an illustration, we assume a round strong beam (see Problem 1, above) and launch the weak beam with $x_0 = 0$ and $x_0' = 0$. The weak beam will then stay in the y plane. After each transformation the beam acquires a new set of values for y and y', which is then represented as a discrete point in the (y, y') phase space. Repeated application of the transformations then traces out the weak beam trajectory in the phase space. The motion is stable if the trajectory does not migrate away from the origin.

Figures 3(a) to 3(d) are the results of four strong-weak simulation runs. Figure 3(a) is the result if we ignore the beam-beam force. Not surprisingly, the weak beam traces out an elliptical trajectory in the phase space and the motion is stable.

In Fig. 3(b), we take into account of the linear term, i.e., the first term in the Taylor expansion of Eq. (3), of the beam-beam force. The weak beam still traces out a stable ellipse, although now the ellipse is distorted. Stability is

assured by the fact that we are in the stable region of Fig. 2.

Figure 3(c) takes into account the next octupole term in the Taylor expansion of the beam-beam force. We find that some trajectories trace out stable islands while some others show "stochastic" behavior in the phase space.[12] The trajectories that show stochastic behavior finally get outside the scope of the figure. Inclusion of the octupole term of the beam-beam force has caused the beam to become unstable.

Had we stopped here, we might have thought that we had found the explanation for beam-beam instability. But Fig. 3(d) shows the case when the complete beam-beam force is included. What is striking is that the stable

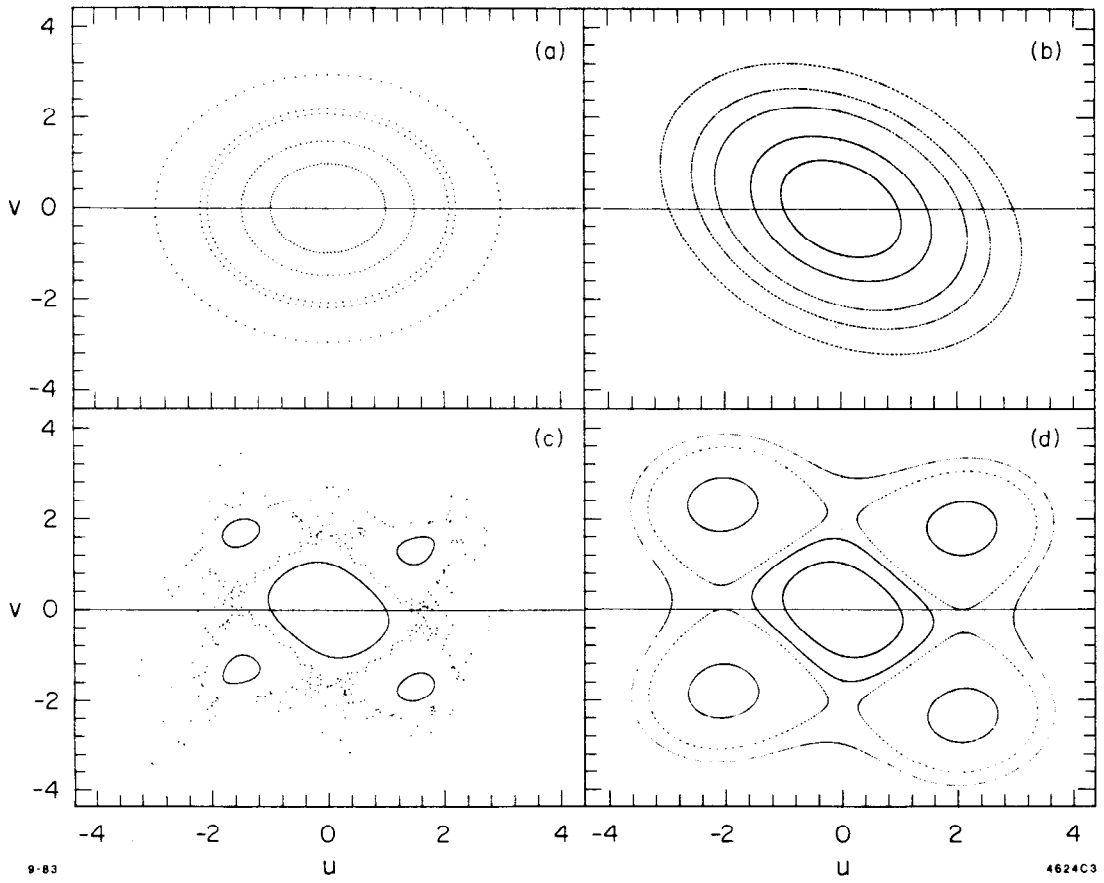


Fig. 3. Weak beam trajectories in the normalized phase space (u, v), where $u = y/\sigma$, $v = \beta_0^* y'/\sigma$. We assume $\mu/2\pi = 0.23$. (a) We ignore the beam-beam force. (b) We include only the linear term of the beam-beam force. (c) We include the linear and the octupole terms. (d) We take into account the complete beam-beam force. In each diagram, trajectories of the same five sets of initial conditions are followed. Note the qualitative difference between (c) and (d).

islands seen in Fig. 3(c) are still there, but the stochastic regions have basically disappeared!

This means that the nonlinear beam-beam force alone does not destabilize the beam, at least for the parameters considered. This is in sharp contrast to the resonance instabilities driven by magnetic field imperfections, as exemplified in Fig. 3(c). The reason for such behavior has been explained analytically.[13] The point is that the beam-beam force diminishes quickly once the weak beam acquires an amplitude larger than the size of the strong beam. In fact, in the limit of very large amplitudes, the weak beam acts as if unperturbed and is necessarily stable. The beam-beam force therefore produces islands in the phase space but not yet an instability.

4 TUNE SPREADS

Before proceeding to more sophisticated beam-beam concepts, let us first introduce one of the most prominent nonlinear effects of beam-beam interaction — tune spreads.[14,15]

We already know that, if a particle executes small oscillations, its tunes will shift by ξ_x and ξ_y in the horizontal and vertical motions. We have also learned that a particle with very large oscillation amplitudes experiences little beam-beam perturbation — and consequently small tune shifts — since it simply stays outside the range of beam-beam force most of the time. Therefore the tune shifts are functions of the oscillation amplitudes of the particle. (This behavior is called detuning.) If now a weak beam contains a distribution of particles of various amplitudes, it will end up with tune spreads.

The horizontal and vertical tune shifts are directly related to the slopes of the beam-beam forces $\partial f/\partial x$ and $\partial f/\partial y$.[2] For given x and y amplitudes, the tune shifts of a weak beam particle are calculated essentially by averaging the slopes of the beam-beam force over the range reached by the particle. The case of a particle executing a one-dimensional y-motion is illustrated in Figs. 4(a) to 4(d). In the general case, with a Gaussian strong beam, the tune shifts per collision point as functions of amplitudes are found by the averaging procedure to be[14–16]

$$\Delta\nu_x = \xi\left(\frac{1+\frac{1}{a}}{2}\right)\int_0^\infty \frac{du}{(1+u)^{3/2}(1+\frac{u}{a^2})^{1/2}}\, Z_1\left(\frac{\alpha_x}{1+u}\right) Z_2\left(\frac{\alpha_y}{1+\frac{u}{a^2}}\right),$$

---

[2] After all, a quadrupole magnet changes the tune because it produces a force with nonzero slope.

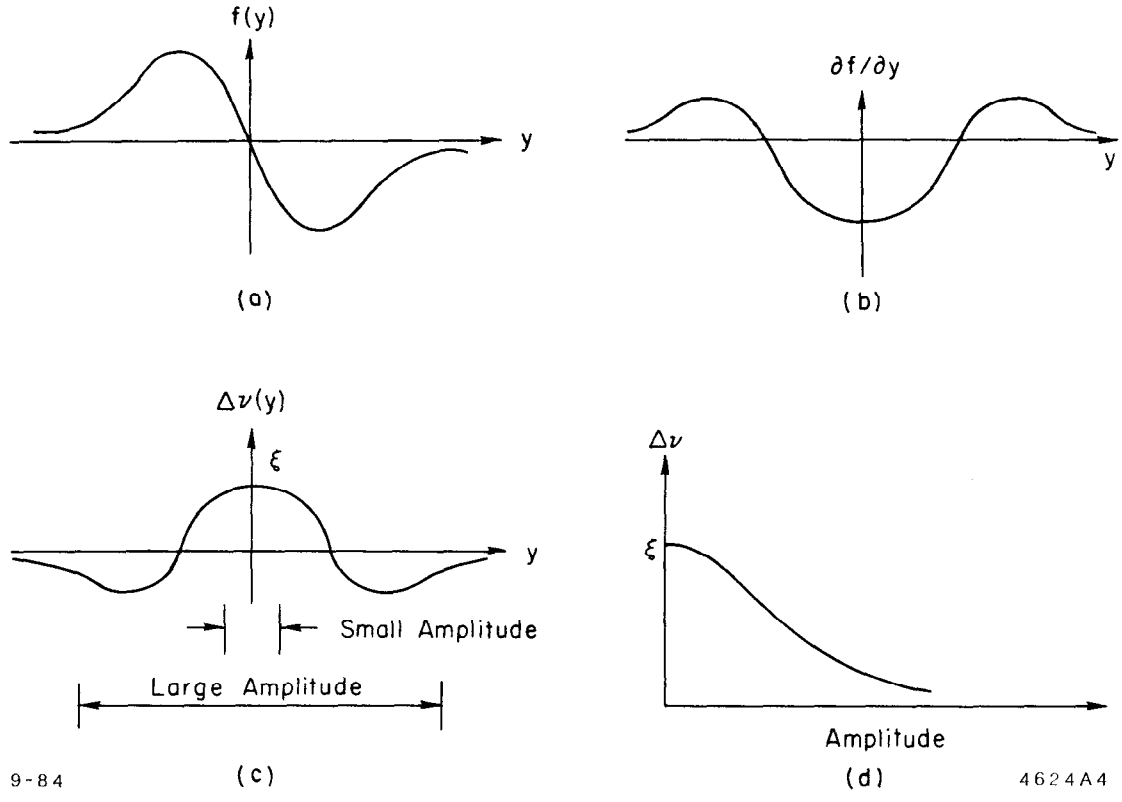


Fig. 4. Schematic illustration of the beam-beam tune shift mechanism: (a) shows the beam-beam force; (b) the slope of this force. Before averaging, the tune shift is proportional to $-\partial f/\partial y$ in such a way that $\Delta\nu = \xi$ at the origin, as shown in (c). This $\Delta\nu(y)$ is averaged over the range reached by a given amplitude; two such ranges are also shown in (c), one for a small amplitude particle and one for a large amplitude particle. The result after averaging gives the detuning curve, which looks like (d).

$$\Delta\nu_y = \xi\left(\frac{1+a}{2}\right)\int_0^\infty \frac{du}{(1+u)^{3/2}(1+a^2u)^{1/2}}\, Z_2\left(\frac{\alpha_x}{1+a^2u}\right) Z_1\left(\frac{\alpha_y}{1+u}\right), \qquad (12)$$

with functions

$$Z_1(x) = e^{-x}[I_0(x) - I_1(x)],$$

$$Z_2(x) = e^{-x} I_0(x),$$

where $\sqrt{\alpha_x}$ and $\sqrt{\alpha_y}$ are the amplitudes normalized by σ_x and σ_y, respectively, $a = \sigma_y/\sigma_x$ is the aspect ratio of the strong beam distribution, and I_0 and I_1

are Bessel Bessel functions. In Eq. (12), we have assumed that the beam-beam parameters in x and y are equal, i.e., $\xi_x = \xi_y = \xi$.

We have come across an important result in passing. The tune spread of the beam is equal to the tune shift of small amplitude particles, and both are equal to ξ:

$$\text{tune spread} \;=\; \text{small amplitude tune shift} \;= \xi \,. \tag{13}$$

In other words, the parameter ξ has now acquired a second physical meaning, namely, the strong-beam induced tune spread.

Figures 5(a) to 5(c) show the tune spreads in the $\nu_x - \nu_y$ space. Without beam-beam collisions, the weak beam has tunes ν_{x0} and ν_{y0}. When the beams collide, particles with small amplitudes have their tunes shifted to $\nu_{x0} + \xi_x$ and $\nu_{y0} + \xi_y$, while particles with large amplitudes keep their unperturbed tunes. The weak beam as a whole then occupies an area in the $\nu_x - \nu_y$ space. The "working point" thus extends in the upper right direction into a "working area" in the tune space. Note that in both x and y motions, Eq. (13) is independently satisfied.

Problem 2. Find the tune shifts as functions of amplitude if the strong beam has (a) a round Gaussian distribution, (b) a uniform disk distribution. Draw the working area in both cases. In case (b), the weak beam does not have tune spread until it goes beyond the boundary of the strong beam. Then, the weak beam motion is entirely linear.

5 THE SINGLE-RESONANCE MODEL

The calculation of tune spreads described above assumes there is no destructive resonance

$$p\nu_x + q\nu_y = n \quad (p, q, n = \text{integers}) \tag{14}$$

trespassing the working area. Otherwise particle motions will be seriously perturbed by the resonance. One then argues that these resonances — at least the lower-order ones — must be avoided. Note that the resonances can be driven by the beam-beam force itself as well as by the storage ring nonlinearities.

In order to avoid resonances, the tune spread must not be too large. One possible picture of the beam-beam instability then emerges: the tune spread ξ must be small enough that the working area can be fitted into a tight resonance-free region, as sketched in Fig. 5(d). This is the single-resonance model of the beam-beam stability limit.

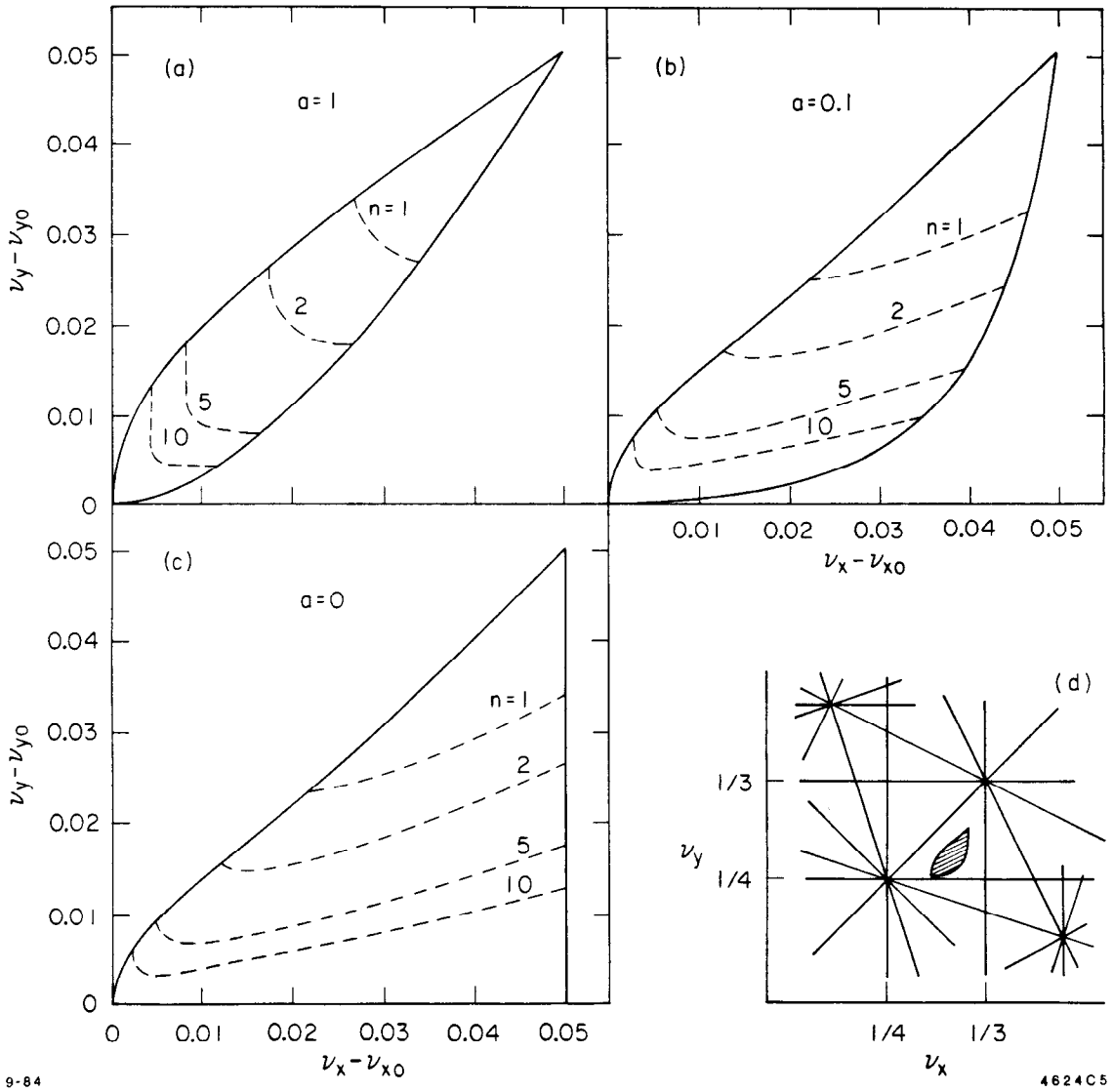


Fig. 5. Beam-beam tune spreads. We assume the two beams have opposite charges. (ν_{x0}, ν_{y0}) is the unperturbed working point. With beam-beam collisions, the working point extends into a working area. The dotted lines are the contours for particles with amplitudes satisfying $x^2/\sigma_x^2 + y^2/\sigma_y^2 = n^2$. We assume $\xi_x = \xi_y = 0.05$. Case (a) is when the aspect ratio is $a = 1$, i.e., a round beam. Case (b) is when $a = 0.1$, i.e., a flat beam. Case (c) gives the result in the limit $a = 0$. Fitting the working area (shaded region) into a resonance-free region in the tune space is shown in (d).

It is not clear how small ξ must be because it is not clear to what order the resonances must be avoided. The conventional wisdom has it that the maximum tolerable ξ is about 0.05 for electron rings and 0.005 for proton rings. In the single-resonance picture, the difference is attributed to the fact that there is

radiation damping in electron rings but not in proton rings; as a result, the available resonance-free region is larger for electrons than for protons because protons are vulnerable to resonances up to order, say, 10, while electrons need to avoid resonances only up to order, say, 5.

For e^+e^- storage rings, the aspect ratio tends to be small, say, $a = 0.1$. Inspection of the shape of the working area in Fig. 5(b) shows that the better choice is to have the unperturbed working point lie on the lower right side of the destructive resonances rather than on the upper left side. For example, when applied to the diagonal $2\nu_x - 2\nu_y = n$ resonance, this means that the unperturbed working point should be below the resonance line, as was first pointed out by Montague.[17] Note, however, that the principle applies to other resonances as well.

But there is a problem. As Fig. 3(d) showed, although a low-order resonance near the working area perturbs particle motion, its main effect is to produce a set of islands in the phase space and not really to cause any instability. The beam-beam instability is therefore still lacking a mechanism. To reconcile this apparent difficulty, several possibilities have been suggested. A few examples of these will be given in the next two sections.

6 TRAPPING MODEL AND ENHANCED DIFFUSION MODEL

The single-resonance model described above assumes that all parameters such as tune and ξ stay constant in time. In this section, we will first describe a trapping model in which the tune is modulated more or less sinusoidally in time with a certain slow frequency and, during this process, repeatedly crosses a resonance value n/q. Unlike the static single resonances, this provides a mechanism[18] which continuously brings particles from small to large amplitudes. A physical aperture limitation on the amplitude then potentially explains the observed lifetime limitation in colliding beams.

In the static model, a particle moves along a constant Hamiltonian contour and, as shown in Fig. 3(d), some contours form islands in the phase space. Since the distance of the islands from the phase space origin is proportional to $\nu - n/q$, islands in phase space move in and out as the tune oscillates. Phase space area elements, together with the particles enclosed in them, are distorted and relocated by this island motion. In particular, particles may be trapped by the islands as they move out in phase space (small oscillations around the island centers are stable.). This then causes particle loss to the aperture limit. The process is schematically shown in Fig. 6.

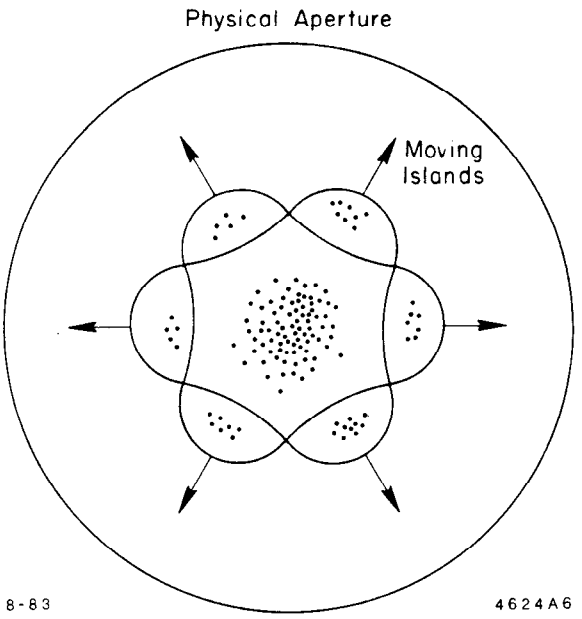


Fig. 6. The trapping model. Some particles are being trapped and moved outwards to an aperture limit. How many particles are trapped depends on the resonance strength and the island moving speed.

One possible source of tune oscillation is the synchrotron oscillation of a particle's energy coupled with a chromatic dependence of the tune. In this case, the tune is modulated at the synchrotron frequency.

We now turn to a second variation of the single-resonance model. This time the parameters are time-independent, but in addition to the beam-beam force there is a diffusion effect on particle motion. Although a beam-beam induced single resonance does not directly cause particle loss, it does enhance the diffusion process and cause particle loss indirectly.[19–21]

In electron rings, the diffusion caused by the synchrotron radiation noise and the radiation damping provided at the acceleration cavities counteract each other; an equilibrium beam distribution is reached when the two effects are in balance. The beam lifetime is then determined by the particle diffusion rate at the physical aperture limit. The larger the aperture, the longer the lifetime. Clearly the distortion of phase space by the presence of a single resonance will also distort the equilibrium distribution. As a result, the effective physical aperture limit is reduced by an amount of the order of the width of the phase space islands. The beam lifetime will then be shortened accordingly.

For proton rings, candidates for the diffusion effects are intra-beam scattering, residual gas scattering, power supply noise, etc. These diffusion effects will also be enhanced by the presence of single resonances.

7 OVERLAPPING RESONANCES

The enhanced diffusion and the trapping models described above are not the only possible explanations of beam-beam instability by way of strong-weak single resonances. We now describe still another possible alternative in this section.

To do this, we need to take into account resonances of not just the lower orders but all orders. The working area is then covered densely by resonance lines. Although the higher-order resonances have narrower widths, the fact that there is an infinite number of them may still result in a significant effect. Indeed, as suggested by Chirikov,[22] if these high-order resonances are wide enough that they overlap into a continuum in the tune space, particle motion will be unstable even if the working area is free of low-order resonances.

Computationally what we do is first calculate the widths of beam-beam resonances of all orders as if they were separated single resonances and then add up these widths to obtain a total width. If this total width is comparable with or larger than the available tune space, we will have reached a stability limit. This procedure is called the Chirikov criterion. Since the total width of all resonances is proportional to ξ, the Chirikov criterion sets a stability limit on ξ, which is referred to as the stochastic limit.

Figure 7(a) shows a detuning curve like that in Fig. 4(d), but here we have indicated two resonant tune values within the tune spread range, one of the 5th order and another of the 8th order. The 5th-order resonance occurs at a smaller amplitude than the 8th-order resonance.

Below the stochastic limit, particles move along closed smooth contours in the phase space, as in Fig. 7(b), and there is no instability. The two strings of islands correspond to the two resonances at two separated amplitudes. If ξ exceeds the stochastic limit, however, the two sets of islands overlap each other, as illustrated in Fig. 7(c). But particle trajectories do not intersect in the phase space; as a result, not knowing which set of islands to follow, particles can only move stochastically from one island region to another, yielding what is shown in Fig. 7(d). Although each resonance is stable if considered alone, overlapping resonances make it possible for a particle to gain amplitude rapidly within the stochastic region.

More quantitatively, the resonance width is obtained by first computing the width of the phase space islands, $\delta\alpha$ in Fig. 7(b), and then translating this width

into ν-units by Eq. (12).[22–24] For the one-dimensional case near the resonance $\nu = n/q$, the term $\delta\alpha$ is approximately given by

$$\delta\alpha \approx 4\left(2\xi\left|\frac{G_q(\alpha_0)}{\Delta\nu'(\alpha_0)}\right|\right)^{1/2} \tag{15}$$

where

$$G_q(\alpha) = \frac{2}{q^2-1}e^{-\alpha}\left[(1+2\alpha)I_{q/2}(\alpha) + 2\alpha I'_{q/2}(\alpha)\right] \ , \quad q = \text{even} \ ,$$

is a term that appears in the Hamiltonian that drives the resonance,[16] α_0 is the amplitude at which the island centers are located, and $\Delta\nu$ is the tune shift in the one-dimensional case with its derivative given by

$$\Delta\nu'(\alpha) = -\xi\frac{1}{\alpha}e^{-\alpha}I_1(\alpha) \ . \tag{16}$$

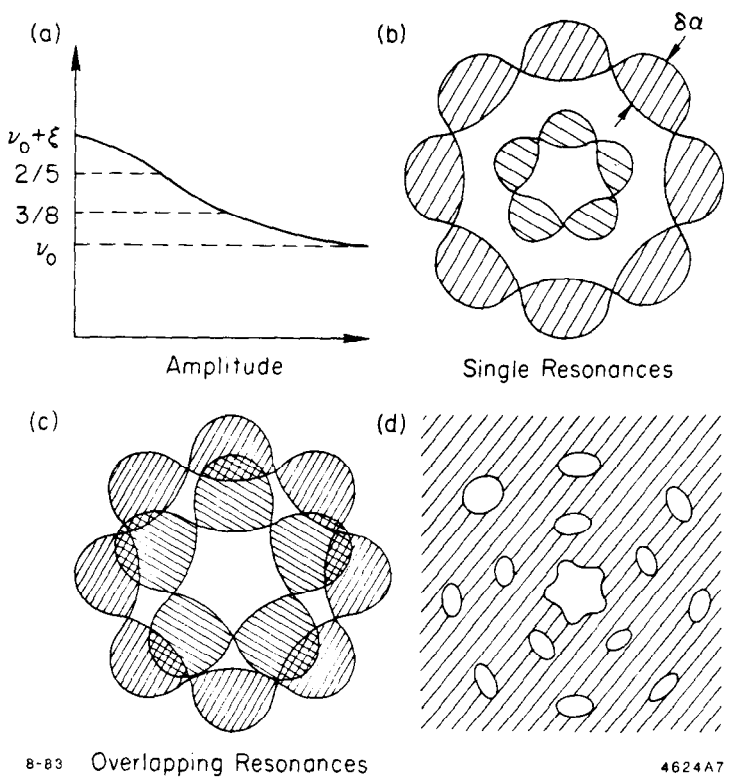


Fig. 7. Sketches of single resonances and overlapping resonances. (a) Detuning curve, showing tune spread covering two single resonances. (b) Two isolated separated single resonances in the phase space. (c) Two overlapping resonances which actually end up looking more like (d), where the shaded area represents the region of stochastic motion.

The corresponding resonance width in ν-units is therefore

$$\begin{aligned} \delta\nu &\approx \delta\alpha \cdot |\Delta\nu'| \\ &\approx 4\Big(2\xi\,|G_q(\alpha_0)\cdot\Delta\nu'(\alpha_0)|\Big)^{1/2} \end{aligned} \tag{17}$$

and the Chirikov criterion for beam-beam stability reads

$$\delta\nu_{\text{tot}} = \sum_{q=\text{even}} q\delta\nu < 1 \tag{18}$$

where $\delta\nu_{\text{tot}}$ is the sum of the widths of all resonances that occur in one unit range of tune space. Only even q's are summed over because the beam-beam interaction does not excite odd q's. The quantity $\delta\nu_{\text{tot}}$ is proportional to ξ.

Figure 8 shows $\delta\nu_{\text{tot}}/\xi$ as a function of α. It has a maximum at $\alpha \simeq 8$, where it has the value of about 10.6ξ. Figure 8 tells us that the most likely area for stochastic motion to occur is around an amplitude of about $\sqrt{8}\sigma$. In this region of the phase space, the beam-beam stochastic limit is found from Eq. (18) to be roughly

$$\xi_{\text{limit}} \approx \frac{1}{10.6} = 0.095\ . \tag{19}$$

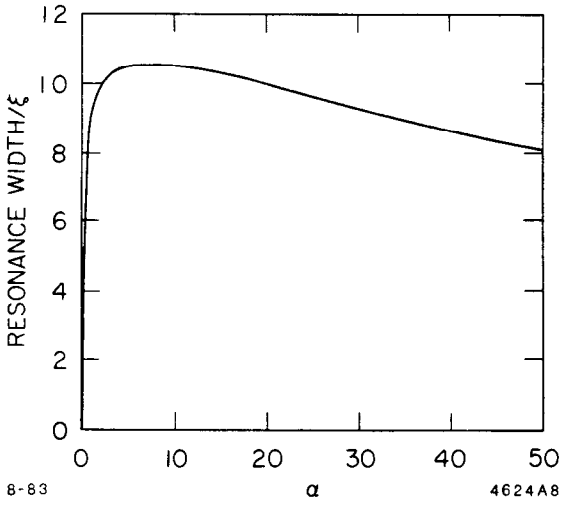


Fig. 8. The total beam-beam resonance width as a function of particle emittance.

The limit (19) is much higher than what has been reached experimentally. One perhaps could explain this discrepancy by adding a time modulation to the

tune or by adding a diffusion to particle motion, as we did on the static single-resonance model. Note also, however, that we have included resonances of a one-dimensional motion only; all coupling resonances have been ignored.

Figure 9(a) shows what happens to Fig. 3 if $\xi = 0.2$.[3] Particle motion exhibits clearly stochastic behavior, apparently leading to an instability. One difficulty with this, however, is shown in Fig. 9(b), which is the same as Fig. 9(a) but with an extended scale. Here we see that the stochastic region is limited to the region around a few sigma's, and that outside this region the motion is bounded again by smooth curves. The physical reason for this is, of course, that the beam-beam force diminishes at large distances, as explained when discussing Fig. 3(d). In other words, unless we are interested in the fine details of single-particle motion (dependence on the initial conditions, etc., to several digits), the gross beam behavior does not depend sensitively on whether the stochastic limit has been exceeded or not.

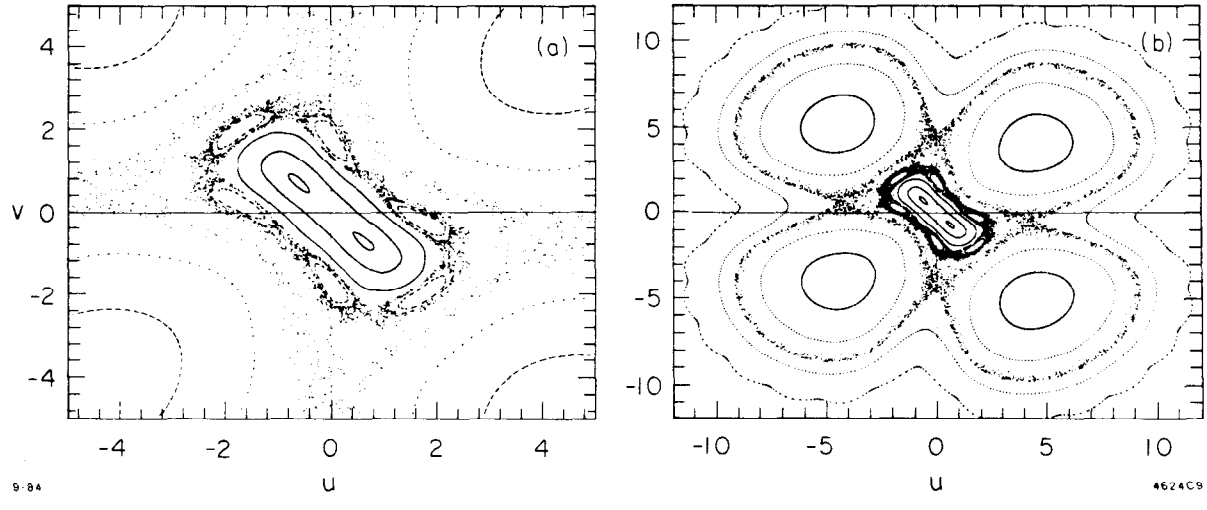


Fig. 9. Particle motion when the stochastic limit is exceeded. Diagram (b) is the same as (a) but with an extended scale. It could be misleading to look only at (a).

This difficulty with Fig. 9(b) aside, the stochastic instability causes rapid growth of particle amplitudes in the stochastic region. For electron storage rings, since the radiation damping would damp out any instabilty that does not grow much in 10^3 revolutions or so, we need not worry about other possible weaker instabilities. For proton rings, where there is no radiation damping, particles

---

[3] $\mu/2\pi = 0.23$ and $\xi = 0.2$ is in the unstable region of Fig. 2. This shows up in that the origin is an unstable fixed point in Fig. 9.

need to be stable for the lifetime of the beam, i.e., 10^{11} revolutions or so, and therefore instabilities much weaker than the overlapping resonances need to be dealt with. One such weak instability is called Arnold diffusion,[25–27] in which, long before resonances overlap in the (y, y') phase space, particles can acquire large amplitudes by slowly channeling through the very thin stochastic layers surrounding the islands in multidimensional phase space. This phenomenon requires sometimes long-term numerical trackings and is typically rather intricate to study.

8 THE INCOMPRESSIBLE FLUID MODEL

An interesting alternative view of the Chirikov criterion was suggested by Teng.[28] This view is based on the analogy between particle motion in the phase space of a Hamiltonian system and the motion of a viscous incompressible fluid. (After all, according to the Liouville theorem, phase space area is incompressible.) By writing the Hamiltonian equation on the one hand and the fluid equation on the other, it is possible to establish the analogy, as given in Table I. The Chirikov criterion is then equivalent to the Reynolds condition in fluid dynamics that the viscosity must be large enough to prevent turbulence from occurring. This offers an alternative derivation of the Chirikov criterion.

Table I Analogy between the overlapping resonances and an incompressible fluid

Overlapping resonances	Incompressible fluid
particle motion in phase space	fluid motion in real space
Hamiltonian equation	Navier-Stokes equation
smooth contours	laminar flow
stochastic behavior	turbulence
Chirikov criterion	Reynolds condition

9 DYNAMIC BETA

So far we have been talking about the strong-weak case. For the strong-strong case, beam-beam interaction becomes much more complicated because perturbation on one beam in turn influences the other beam. For instance, it would be invalid to preassume a Gaussian distribution since the distribution must come from solving self-consistently a dynamic system that contains both beams.

Nevertheless, if we consider a linearized strong-strong case, the beam distribution would still remain Gaussian. The only effect is that the rms beam size at

the collision point is now proportional to $\sqrt{\beta^*}$, where β^* is the perturbed beta-function. Since, according to Eq. (9), β^* depends on ξ, which in turn depends on the beam size, this means the beam size, β^*, and ξ depend on one another and need to be found self-consistently for any given value of beam intensity N. Such behavior is called the dynamic-beta effect.[29] It is the simplest of the strong-strong models.

One can also study the effect of dynamic-beta on the luminosity $\mathcal{L}$. For a Gaussian beam, $\mathcal{L}$ is given by

$$\mathcal{L} = \frac{N^2 f B}{4\pi\sigma_x\sigma_y} \tag{20}$$

where f is the revolution frequency, and B is the number of bunches per beam.[4]

In Figs. 10(a) to 10(c), we show the dynamic-beta behavior of β^*, ξ, and $\mathcal{L}$. In the range of N of interest, the beam-beam force pinches β^* — and consequently the beam size at the collision point — to a smaller value if $\mu/2\pi < 0.1$. For larger $\mu/2\pi$, the reverse is true. This behavior repeats with period $\mu/2\pi = 1/2$. This means the luminosity would benefit from having $\mu/2\pi$ slightly above a half-integer and would suffer if $\mu/2\pi$ is slightly below a half-integer. Note also that there is always a saturation behavior on $\mathcal{L}$ versus N.

In Fig. 10, we have assumed a round beam and $\mu_x = \mu_y$ for simplicity. In this simplified case, the self-consistent solution is given by

$$\begin{aligned} \beta^*/\beta_0^* &= \sqrt{1 + (2\pi\xi_0 \csc\mu)^2} - 2\pi\xi_0 \cot\mu \ , \\ \xi &= \xi_0\beta_0^*/\beta^* \ , \\ \mathcal{L} &= \mathcal{L}_0\beta_0^*/\beta^* \ , \end{aligned} \tag{21}$$

where β_0, ξ_0, and $\mathcal{L}_0$ are the quantities in the absence of dynamic-beta.

It is not clear whether there is experimental evidence of the dynamic-beta effect. There are some indications of increased luminosity as $\mu/2\pi$ is lowered toward slightly above a half-integer, see Fig. 11.[30] The fact that luminosity levels off at high beam intensities agrees with dynamic-beta, although almost any reasonable beam-beam model could have predicted the same.

---

[4] From the beam dynamics point of view, luminosity is not a very interesting quantity. It is simply a geometrical quantity representing the transverse beam area.

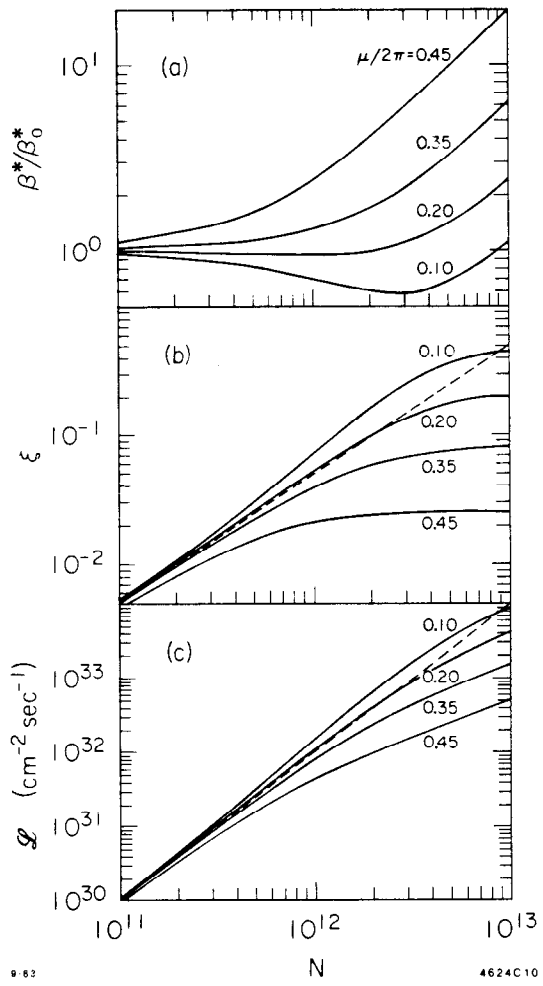


Fig. 10. Dynamic-beta behavior of β^*, ξ, and $\mathcal{L}$. The term $\mu/2\pi$ is the unperturbed tune advance between collision points (modulus 1/2). The dashed lines in (b) and (c) are the reference values when the dynamic-beta effect is ignored. We have chosen the normalization that $\mathcal{L}_0 = 10^{32}\ \mathrm{cm}^{-2}\ \mathrm{sec}^{-1}$ and $\xi_0 = 0.05$ at $N = 10^{12}$.

One may also ask the following question: suppose the strong-strong case is not linearized, what will be the equilibrium beam distribution now that it is no longer Gaussian? This is an important practical question since it relates directly to the luminosity, but unfortunately it is also a very difficult one. Some progress has been made on the weak beam distribution in the strong-weak case,[31,32] but as it stands now, more effort is needed in this research area.

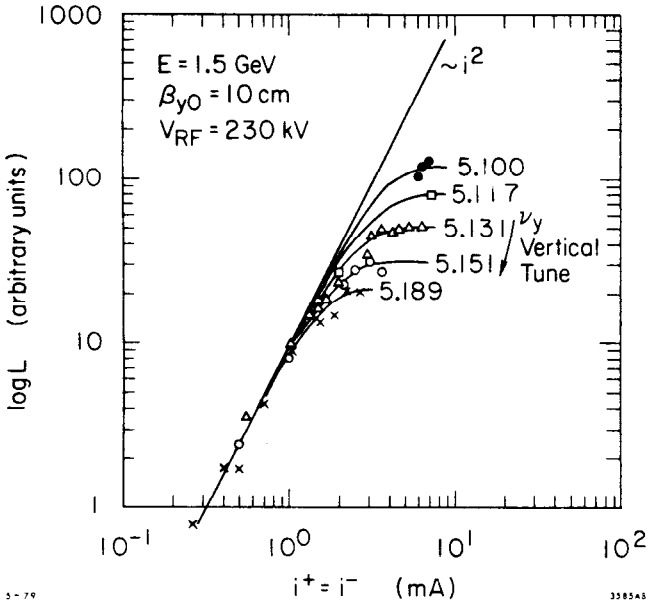


Fig. 11. SPEAR I data that show tune and beam intensity dependences of luminosity.

10 LOW-β^* INSERTION

The beam-beam parameter ξ, as explained in Eq. (13), has the meanings of the beam-beam induced tune spread and the small amplitude tune shift. In Eq. (19), it was used to set the stochastic limit. These studies, together with what we will see later, indicate that ξ has the meaning of simply being the dimensionless scaling parameter of the beam-beam problem. It is the parameter that specifies the linear as well as the nonlinear beam-beam strength.

This observation has one extremely important practical consequence — the invention of low-β^* insertions.[33,34] According to this scaling property, it would be beneficial to make ξ as small as possible, and inspection of Eq. (10) shows that minimizing β^* would do this.

Today, low-β^* insertions are implemented on all colliding beam storage rings. As a result, luminosities have increased by one to two orders of magnitude. And yet this is not the end. Ideas for making β^* smaller are still actively being developed, and they bring success every time they are tried, as evidenced by current talk about "mini-β^*" and even "micro-β^*" insertions.[35–37]

In a low-β^* insertion, a few strong quadrupoles are inserted in the interaction region to pinch β^* to a small value. Figure 12 illustrates the difference between a low-β^* insertion and a normal cell structure.

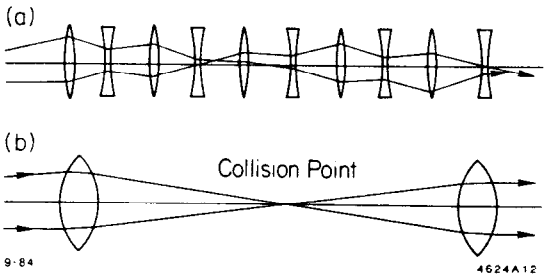


Fig. 12. Difference between (a) a normal cell structure and (b) a low-β^* insertion. In each case, two typical trajectories are drawn. In (a), the effects of focusing and defocusing magnets tend to cancel each other. The net effect is focusing, but the focal length is long, and the displacement of a particle changes sign only after the particle passes through several magnets. In (b), a strong quadrupole magnet "overfocuses" the particle trajectories so that all displacements change sign near the low-β point. Such an overfocused configuration is usually to be avoided in a normal cell structure.

The low-β^* insertion quadrupoles can not be too close to the collision point since the detector solenoid has compensating solenoids on both sides. This puts a limit on the smallest beta achievable at the collision point. In a mini-β^* insertion, the compensating solenoids are removed to make room for the insertion quadrupoles (at the cost of some complications in ring optics) so that β^* can be made smaller. One can even go one step further and contemplate the possiblity of having the insertion quadrupoles inside the detector to produce a micro-β^*. These micro-β^* quadrupoles need to be permanent magnets. The various small-β^* schemes are shown in Fig. 13.

It is incorrect to say that the benefit of low-β^* is due to pinching of the beam size at the collision point to a smaller value. Although the luminosity does increase when beam size is reduced, the idea of low-β^* actually tends to ask for a large beam size. The reason is basically given by

$$\xi_y \propto N/A^* ,$$

$$\mathcal{L} \propto N^2/A^* , \tag{22}$$

where A^* is the beam area at the collision point and we have assumed $\sigma_x \gg \sigma_y$. Since beam-beam limit is thought to be associated with a maximum allowed

value in ξ, the maximum beam intensity is given by

$$N \propto A^* \ . \tag{23}$$

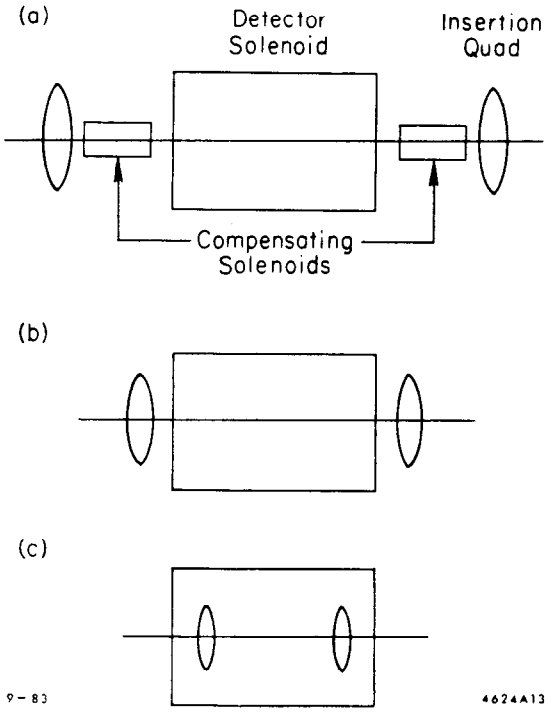


Fig. 13. (a) Low-β^*, (b) mini-β^*, and (c) micro-β^* insertions.

This leads to the conclusion that

$$\mathcal{L} \propto A^* \tag{24}$$

which means we want to have A^* as large as possible, in contrast to what one might have expected.[5]

The question is therefore how to insist on a small β^* and at the same time to ask for a large beam size at the collision point. There have been several clever ideas of how to do this. Some of them are listed below without much explanation:

---

[5]We are ignoring the subtlety that two ξ's, ξ_x and ξ_y, are involved in the beam-beam limit. The complete story is more involved, as always. See Ref. 38.

— weaker focusing, i.e., smaller tune

— wigglers[39]

— mismatched dispersions in the ring lattice[40]

— finite dispersion at the collision point

The first three of these involve artificially enlarging the beam emittance. Clearly such ideas are restricted by their requirement for imply large beam sizes — and therefore a large vacuum chamber — not just at the collision point but also everywhere else around the ring, which is very expensive. The fourth idea does not have this problem, but it does have the problem that the beam-beam force may excite harmful synchro-betatron resonances.[41]

11 OPTIMUM β^*

One might ask what sets the limit in going to smaller and smaller values of β^*. One limit results from the fact that the beams collide over a region of finite length, while β^* is the value of β only at the center of this collision region.[42,43] For bunched beams, the collision region is of course given by the length of the bunches. We will explain in this section how the finite bunch length introduces an optimal value for β^*. It would not be beneficial to make β^* smaller than this optimal value.

Let us adopt a simplified model in which the bunches have uniform longitudinal distribution with full length $2L$. We will assume a flat beam and study the effect of making β_y^* small, leaving β_x^* unchanged. We will show that in this model, the optimum β^* is about equal to 1/4 of the full bunch length.

As we move away from the center of the collision region by distance s, the β-function increases quadratically according to

$$\beta_y(s) = \beta_y^*\left[1 + \left(\frac{s}{\beta_y^*}\right)^2\right]. \tag{25}$$

We see that as β^* is pinched to a small value, the beams collide with a geometry like that shown in Fig. 14, keeping in mind that beam size at s is proportional to $\sqrt{\beta(s)}$. We see that too small a value of β^* is harmful because of two effects:

1. The parameter ξ is effectively increased since a particle has to traverse the collision region with $\beta > \beta^*$. This means a stronger limit on the collidible beam intensity.

2. For a given beam intensity, parts of the beams collide with large cross-sectional area, leading to degradation of luminosity.

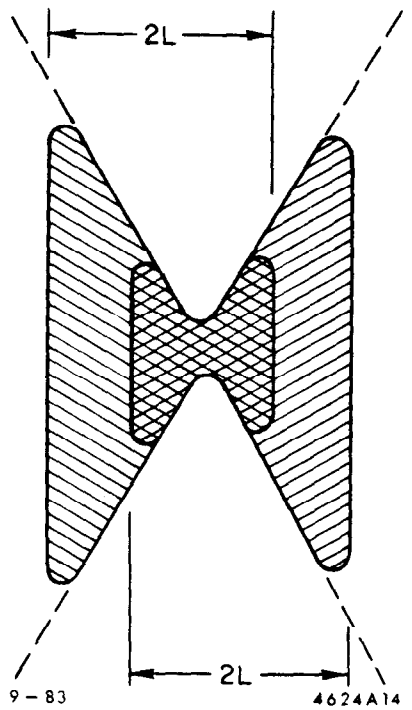


Fig. 14. Colliding beams when β^* is too small.

Since ξ_y is proportional to β_y/σ_y and σ_y is proportional to $\sqrt{\beta_y}$, the effective ξ is given by averaging ξ over the collision region, i.e.,

$$\xi = \xi_0 \frac{1}{2L} \int_{-L}^{L} ds \sqrt{1 + (s/\beta_y^*)^2} = \frac{\xi_0}{2} \left[\sqrt{1+u^2} + f(u)\right] \tag{26}$$

where $u = L/\beta^*$, $f(u) = (1/u)\ln\left(u + \sqrt{1+u^2}\right)$, and ξ_0 is the unperturbed value, which is proportional to $N/\sqrt{u}$. This means that if we demand that ξ be less than some constant value, then the maximum beam intensity allowed by beam-beam interaction satisfies

$$N \propto \frac{\sqrt{u}}{\sqrt{1+u^2} + f(u)} \ . \tag{27}$$

The maximum luminosity then behaves according to

$$\mathcal{L} \propto \frac{N^2}{\sqrt{\beta_y^*}} \frac{1}{2L} \int_{-L}^{L} \frac{ds}{\sqrt{1+(s/\beta_y^*)^2}} = \frac{N^2}{\sqrt{\beta_y^*}} f(u) \tag{28}$$

where N is given by Eq. (27).

In Fig. 15 we have plotted the beam-beam limited beam intensity and the luminosity as functions of L/β_y^*. We find that both N and $\mathcal{L}$ reach maximum when β_y^* is about equal to 1/4 of the collision region length $2L$. Further decreasing β^* does not help the luminosity.

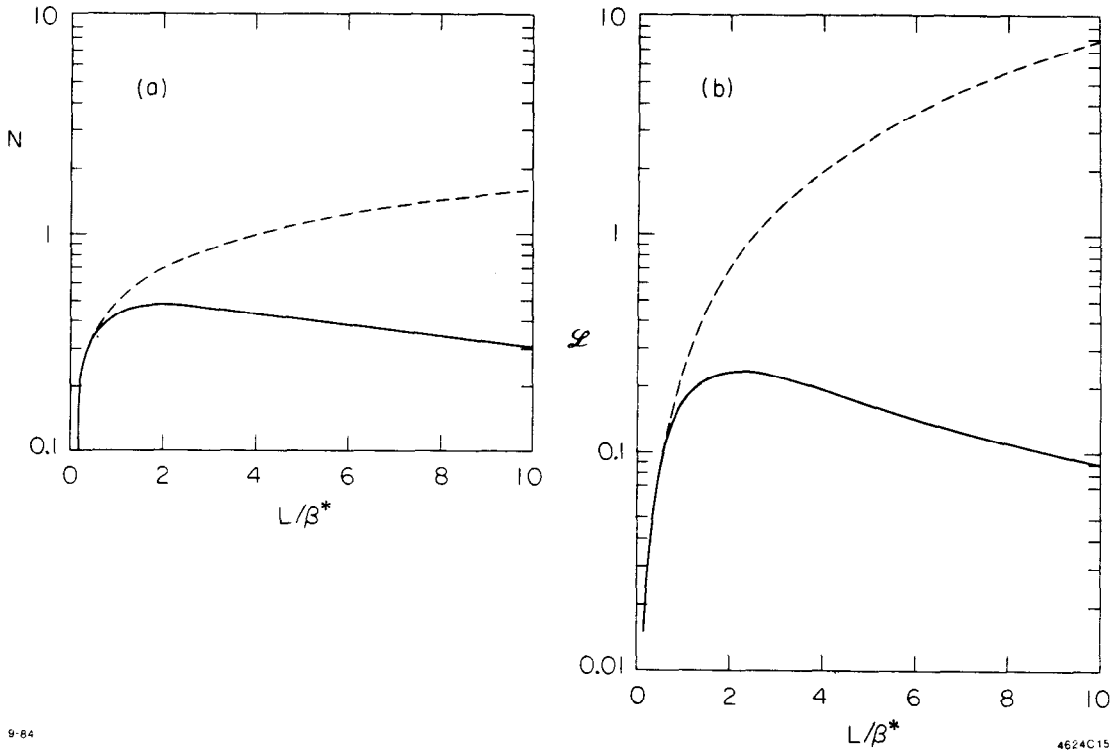


Fig. 15. Effect of finite bunch length on optimum choice of β^*. (a) Maximum beam intensity allowed by the beam-beam limit versus β_y^*/L. (b) Luminosity versus β_y^*/L. The units for N and $\mathcal{L}$ are arbitrary. The dashed curves are drawn with effect of finite bunch length is ignored.

It should be pointed out that we have considered only the geometrical implications of a small β^*. There are other considerations. One is that a small β^* necessarily requires a large β at the quadrupoles. This means these quadrupoles need to have large apertures to clear the beam, which in turn means they need to be very strong in order to produce the needed gradient. In addition, a large β at these quadrupoles means extreme sensitivity to errors in their construction and installation.

Another effect associated with small β^* is the possibility of exciting synchro-betatron resonances.[44] A particle executing synchrotron oscillation sees the beam-beam kicks away from the collision point. The kicks are applied to the betatron motion of the particle. If β^* is too small, this leads to strong modulation of the kicks at twice the synchrotron frequency, which then becomes a source of synchro-betatron coupling.

12 COHERENT OSCILLATION OF RIGID BEAMS

The next strong-strong model to be considered is coherent oscillation of the bunches. As a first step,[45–47] we represent all bunches by rigid distributions so that only their center-of-mass motions are allowed. The bunches, in addition to undergoing simple harmonic motion in the storage ring, now receive beam-beam kicks when they pass through each other at the collision points. All bunches (in both beams) are then coupled together through the beam-beam kicks to form a dynamic system in which all bunches oscillate in time.

Note that no such coherent motion is allowed in the strong-weak picture. This, in fact, is one serious drawback of the strong-weak picture, especially since coherent motions, as we will soon see, can potentially set tighter stability limits than incoherent motions.

Let us first consider a storage ring with two oppositely circulating bunches that collide alternately at two collision points, as shown in Fig. 1(b). The two bunches are specified by indices 1 and 2 respectively. Let the two bunches have small center-of-mass motions in the y-direction. The kicks given to the two rigid bunches are computed by averaging the kicks over the bunch distribution. In the linear approximation, the result is

$$\Delta y_1' = -\frac{1}{f}(y_1 - y_2)\frac{1}{\sqrt{2}} ,$$

$$\Delta y_2' = -\frac{1}{f}(y_2 - y_1)\frac{1}{\sqrt{2}} , \tag{29}$$

where y_1 and y_2 are the displacements of the two bunch centers at the moment of crossing, f is the focal length defined in Eq. (7), and $1/\sqrt{2}$ comes from a Gaussian form factor. See Problem 3.

Problem 3. Equation (29) gives the coherent kicks for Gaussian beams. More generally, show that the kicks are given by

$$\Delta y_1' = -\Delta y_2' = -\frac{2\sqrt{2\pi}\,Nr_0}{\sigma_x\gamma}(y_1 - y_2)\cdot\int_{-\infty}^{\infty} dy\psi^2(y) \tag{30}$$

where $\psi(y)$ is the normalized beam distribution and a flat beam is assumed. Show that (30) become (29) for a Gaussian beam. Show also that, for a uniform beam, the factor $1/\sqrt{2}$ in Eq. (29) is absent.

Equation (29) can be written in matrix form. Defining the vector (y_1, y_1', y_2, y_2'),

the matrix that describes the beam-beam transformation is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{\sqrt{2}f} & 1 & \frac{1}{\sqrt{2}f} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}f} & 0 & -\frac{1}{\sqrt{2}f} & 1 \end{bmatrix} . \tag{31}$$

After collision, the bunches execute free betatron motion for half a revolution. The transformation is

$$T_0 = \begin{bmatrix} \cos\mu & \beta_0^* \sin\mu & 0 & 0 \\ -\frac{1}{\beta_0^*}\sin\mu & \cos\mu & 0 & 0 \\ 0 & 0 & \cos\mu & \beta_0^* \sin\mu \\ 0 & 0 & -\frac{1}{\beta_0^*}\sin\mu & \cos\mu \end{bmatrix} \tag{32}$$

where μ is the betatron phase advance between two collision points.

Let $T_{\rm tot}$ be the product of the two matrices (31) and (32). The coherent motion is stable if all eigenvalues of $T_{\rm tot}$ have absolute value equal to 1. In the case of two colliding bunches, the motion consists of two modes: a 0-mode in which the two bunches move up and down together, and a π-mode in which the two bunches move out of phase. The 0-mode is always stable because, as the bunch centers move up and down together, there is no beam-beam force acting on them. The π-mode is stable if

$$\xi < \frac{1}{2\sqrt{2}\,\pi} \cot\frac{\mu}{2} . \tag{33}$$

Note that this is $\sqrt{2}$ times more stringent than the strong-weak stability limit described by Eq. (11). This is more clearly seen in the sawtooth diagram Fig. 16(a).

The picture becomes more complicated, although still straightforward, if there are more bunches in the storage ring. For instance, when there are 6 bunches (3 per beam), there will be 6 modes of coherent oscillations. The stability region is obtained by requiring that all 6 modes be stable. Fig. 16(b) shows the sawtooth diagram. The stability region is much smaller than in the incoherent case.

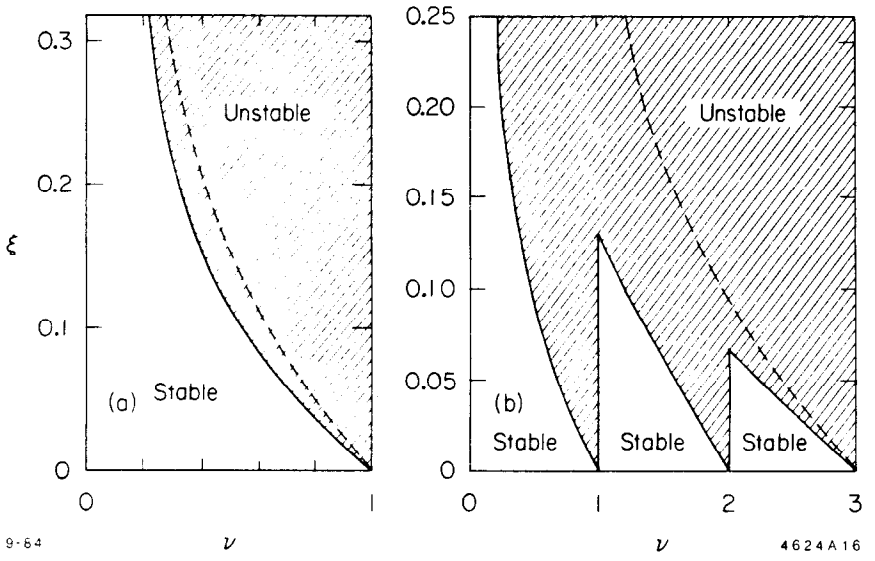


Fig. 16. Stability region for two strong rigid beams executing small center-of-mass oscillations for (a) two colliding bunches and (b) six colliding bunches. The figures are periodic in ν, the total tune of the storage ring; the periods are 1 in (a) and 3 in (b). The dashed lines show the strong-weak stability limit and are reproduced from Fig. 2.

Possibly the strong-strong coherent instability is more relevant than the strong-weak incoherent instability (i.e., the nonlinear mapping studies) in setting the observed beam-beam instability limit. This possibility is based on the following observation: in coherent motions, the separation between one piece of the beam and the corresponding piece of the on-coming beam is effectively twice the separation when one of the beams does not move, as would be the case in the strong-weak case. As a result, the beam-beam kicks are effectively stronger for coherent motions. This observation applies not only to the rigid dipole motions described in this section but also to motions of higher order modes to be described in the next section.

Center-of-mass beam motions induced by beam-beam collisions have been observed in storage rings for the case of two colliding bunches.[48,49] The measurements are less clean when there are more bunches. The rigid beam oscillations are not damped by the feedback systems because in the unstable region the growth rate is very fast.

13 HIGHER-ORDER BEAM-BEAM MODES

In case we relax the condition that all bunches are rigid, the calculation becomes more difficult. In addition to the inter-bunch modes associated with the coupling among bunches, the motion of an individual bunch can be described only as a superposition of modes in its transverse distribution. For instance, the lowest of such modes would be the dipole mode we just considered; and then there have to be quadrupole mode, sextupole mode, etc.

For small beam-beam parameters, coupling among different transverse modes is weak; it is possible to study the coherent beam-beam effect by considering each transverse mode separately. Our calculation of the dipole coherent instability then still applies. In particular, the stability region will be a sawtooth diagram with instability occuring near ν =integers, as shown in Fig. 16.

One can also perform a stability analysis on, say, the quadrupole mode and obtain its stability limit. Then, as we will see, there is instability if ν is close to half-integers. Similarly, sextupole modes are unstable if ν is close to 1/3, etc.

To study the coherent quadrupole modes,[6] let us consider the case of two bunches as sketched in Fig. 1(b). In this case, both bunches have Gaussian distributions, but the second moments of these Gaussian distributions oscillate in time around some equilibrium values. Each bunch will be described by a Σ matrix[51] whose elements are the second moments

$$\Sigma = \begin{bmatrix} < x^2 > & < xx' > & 0 & 0 \\ < xx' > & < x'^2 > & 0 & 0 \\ 0 & 0 & < y^2 > & < yy' > \\ 0 & 0 & < yy' > & < y^2 > \end{bmatrix} . \tag{34}$$

Let us designate the equilibrium sigma matrix as Σ_0. On top of Σ_0, each bunch has a small time-dependent perturbation, i.e.,

$$\Sigma^{(1)} = \Sigma_0 + \Delta\Sigma^{(1)} ,$$
$$\Sigma^{(2)} = \Sigma_0 + \Delta\Sigma^{(2)} . \tag{35}$$

We have assumed that the bunch distribution is not tilted in the x-y phase space. The stability of the coherent quadrupole modes will be determined by the stability of the two matrices $\Delta\Sigma^{(1)}$ and $\Delta\Sigma^{(2)}$.

---

[6]The work described in this section was done in collaboration with Y. Kamiya.[50]

The free betatron motion between collision points can be described by the transformation

$$\Delta\Sigma^{(1)}_{\text{out}} = T_0 \Delta\Sigma^{(1)}_{\text{in}} \tilde{T}_0 \ ,$$

$$\Delta\Sigma^{(2)}_{\text{out}} = T_0 \Delta\Sigma^{(2)}_{\text{in}} \tilde{T}_0 \ , \tag{36}$$

where T_0 is the matrix (32). Matrix elements of $\Delta\Sigma^{(1)}$ and $\Delta\Sigma^{(2)}$ each transform among themselves (without coupling to each other) according to Eq. (36).

The beam-beam transformation is more complicated. Beam 1, for example, will be transformed according to

$$\Sigma^{(1)}_{\text{out}} = T_{bb} \Sigma^{(1)}_{\text{in}} \tilde{T}_{bb} \tag{37}$$

where

$$T_{bb} = T_{bb0} + \Delta T_{bb}$$

is the beam-beam transformation matrix, and

$$T_{bb0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{f_x} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f_y} & 1 \end{bmatrix}$$

is the time-independent part of T_{bb}. The matrix ΔT_{bb} results from beam 2 executing quadrupole motion so that its distribution is not the equilibrium one. The matrix ΔT_{bb} that acts on beam 1 therefore depends on $\Delta\Sigma^{(2)}$. This provides the coupling between the two colliding bunches. Note that we have linearized the beam-beam force.

In terms of the perturbation quantities, Eq. (37) reads

$$\Delta\Sigma^{(1)}_{\text{out}} = \Delta T_{bb} \Sigma_{0\text{in}} \tilde{T}_{bb0} + T_{bb0} \Delta\Sigma^{(1)}_{\text{in}} \tilde{T}_{bb0} + T_{bb0} \Sigma_{0\text{in}} \Delta \tilde{T}_{bb} \ . \tag{38}$$

The coupling among elements of $\Delta\Sigma^{(1)}$ and $\Delta\Sigma^{(2)}$ is linear. That suggests we form a 12-dimensional vector

$$\left[\Delta\Sigma^{(1)}_{11}\ \Delta\Sigma^{(1)}_{12}\ \Delta\Sigma^{(1)}_{22}\ \Delta\Sigma^{(1)}_{33}\ \Delta\Sigma^{(1)}_{34}\ \Delta\Sigma^{(1)}_{44}\ \Delta\Sigma^{(2)}_{11}\ \Delta\Sigma^{(2)}_{12}\ \Delta\Sigma^{(2)}_{22}\ \Delta\Sigma^{(2)}_{33}\ \Delta\Sigma^{(2)}_{34}\ \Delta\Sigma^{(2)}_{44}\right] \tag{39}$$

and compute the two 12 × 12 matrices that transform this vector through the free betatron region and the collision point. The product of these two matrices, T_{tot}, then gives the stability of the quadrupole motions.

Among the 12 eigenvalues of T_{tot}, 4 are associated with constants of the motion and are always equal to unity. The remaining 8 correspond to 4 dynamic inter-bunch modes. Fig. 17(a) shows the region of stability in which all 4 quadrupole modes have stable eigenvalues.

Figure 17(b) shows the stability region for the case of 6 bunches in the two beams. In this case, there are 12 dynamic modes and they all have to be stable in order for the colliding beam system to be stable. We see from Fig. 17 that quadrupole coherent modes impose additional constraint on the parameters ν and ξ.

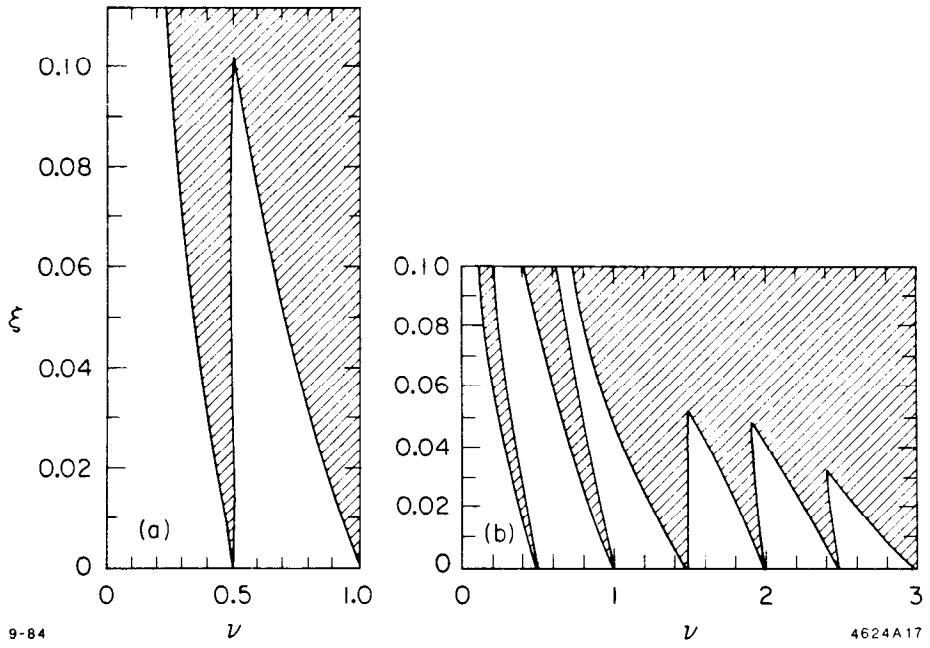
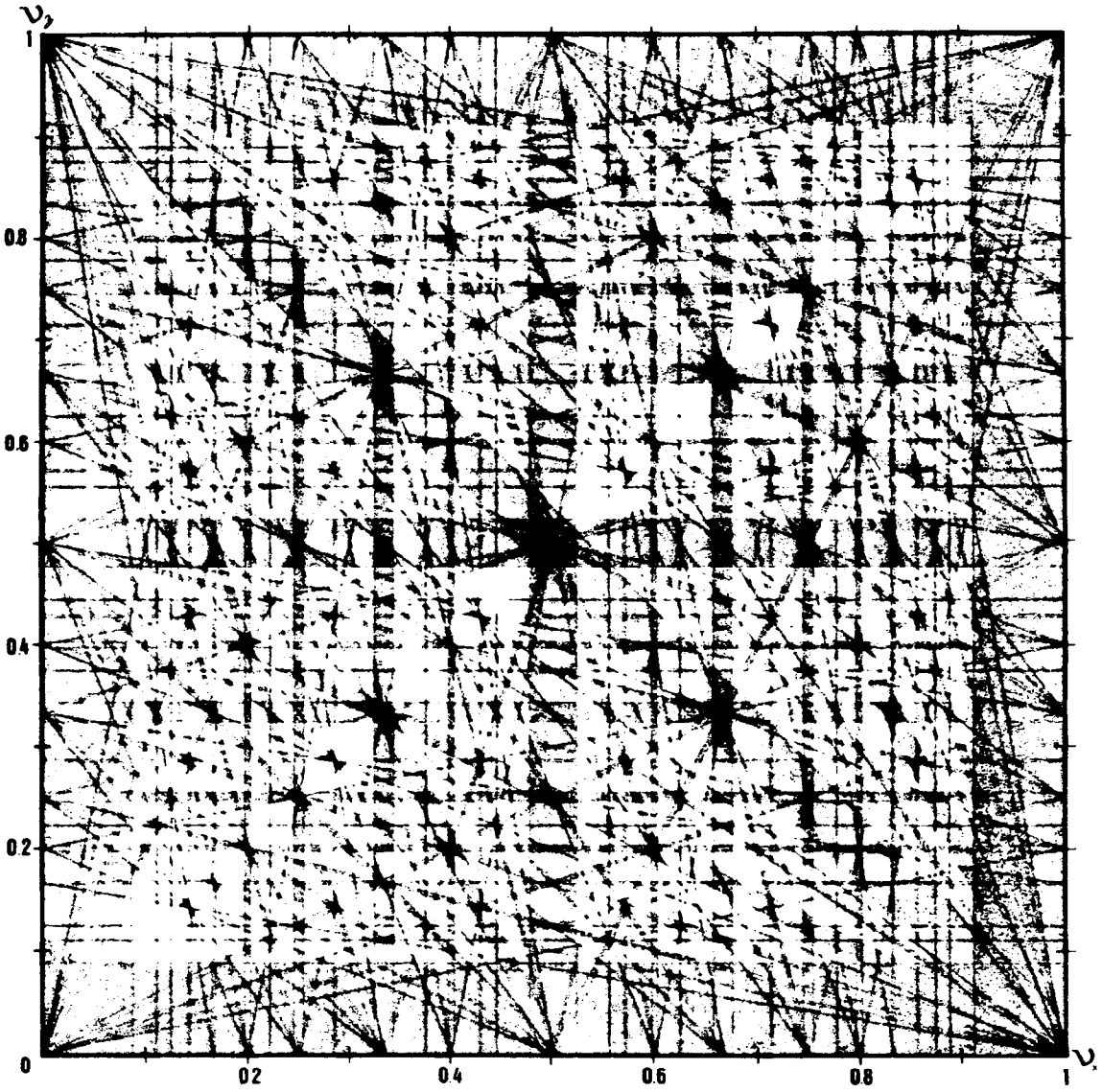


Fig. 17. Stability region for two strong beams executing small quadrupole oscillations. The figures are periodic in ν, the total tune of the storage ring; the periods are 1 in (a) and 3 in (b). A round beam at equilibrium has been assumed.

Other approaches can be taken to studying the higher-order coherent beam-beam modes.[52–54] The result of one such attempt[53] is reproduced in Fig. 18. For a given ξ, the shaded area represents the unstable region in the (ν_x, ν_y) space. The calculation is made for the very special case of the storage ring DCI,[55] in which each of the two colliding bunches is composed of electrons and positrons of equal intensity so that the net charge is neutral.



10-83 4056A16

Fig. 18. Coherent instability region in the ν_x, ν_y space for the storage ring DCI. There is an instability when a resonance condition, Eq. (14), is met. The value of ξ is taken to be 0.05.

The idea of DCI is of course to eliminate the incoherent beam-beam force. But this did not help the luminosity and, according to Fig. 18, one possible reason is that although the strong-weak instabilities are in principle eliminated, the coherent instabilities are actually enhanced. It should also be mentioned that more recently several advances have been made on the strong-strong case of beam-beam interaction in the form of numerical simulations.[44,56–58] These results generally agree quite well with the experimental observations that they

simulate. (It is not clear whether these simulations agree with one another since they attribute the beam-beam instability to different mechanisms. Nevertheless, they all seem to agree with observations.) In some cases, they were even used successfully to find working points that yield better luminosities.

14 IS THE BEAM-BEAM LIMIT GIVEN BY $\xi <$ UNIVERSAL CONSTANT?

We mentioned before that ξ is the scaling parameter of the beam-beam interaction effects. This idea led to the invention of low-β^* insertions. In this section, we will discuss the question of whether the beam-beam stability limit is indeed correctly given by the simple condition

$$\xi < \text{universal constant} \quad . \tag{40}$$

We should point out that the idea of having ξ as the scaling strength parameter is not challenged. Rather we are asking whether there are other parameters in addition to ξ that may also play a role.

The question can also be asked in a different way. We mentioned previously that, according to conventional wisdom, the beam-beam limit is given approximately by $\xi = 0.05$ for electrons and $\xi = 0.005$ for protons. We then attributed the difference to radiation damping. But a moment's reflection suggests that these ideas can not be a complete description of the beam-beam limit. If we consider an electron storage ring with beam-beam limit $\xi = 0.05$ and imagine that we slowly remove its radiation damping, then the difference between electron and proton diminishes, and the beam-beam limit will decrease toward $\xi = 0.005$. If so, the ξ limit is clearly not given by a universal constant; the radiation damping rate has to play a role — at least during the transition from $\xi = 0.05$ to 0.005.

One sensitive control on the radiation damping rate in an electron storage ring is the beam energy. The radiation damping time τ_{rad} is proportional to γ^{-3}. To see whether radiation damping plays a role in determining the beam-beam limit, therefore, one way is to examine the energy dependence of the luminosity.

First let us assume that condition (40) does correctly describe the beam-beam limit; then Eq. (10) would predict

$$N \propto \gamma^3 \tag{41}$$

since the beam dimensions σ_x and σ_y are proportional to γ in electron rings.[59] Inserting this into Eq. (20) then gives

$$\mathcal{L} \propto \gamma^4 \, . \tag{42}$$

The prediction (42) does not seem to agree with the measurements. Figure 19 is a compilation of luminosity versus beam energy for the storage ring SPEAR.[60] According to Fig. 19, luminosity depends on beam energy as

$$\mathcal{L} \propto \gamma^{6.7\pm0.1} \quad . \tag{43}$$

It has been suggested[61,62] that the discrepancy between (42) and (43) can be explained by a phenomenological "diffusion model" that incorporates the radiation damping into the beam-beam instability mechanism. This model is further discussed in the next section.

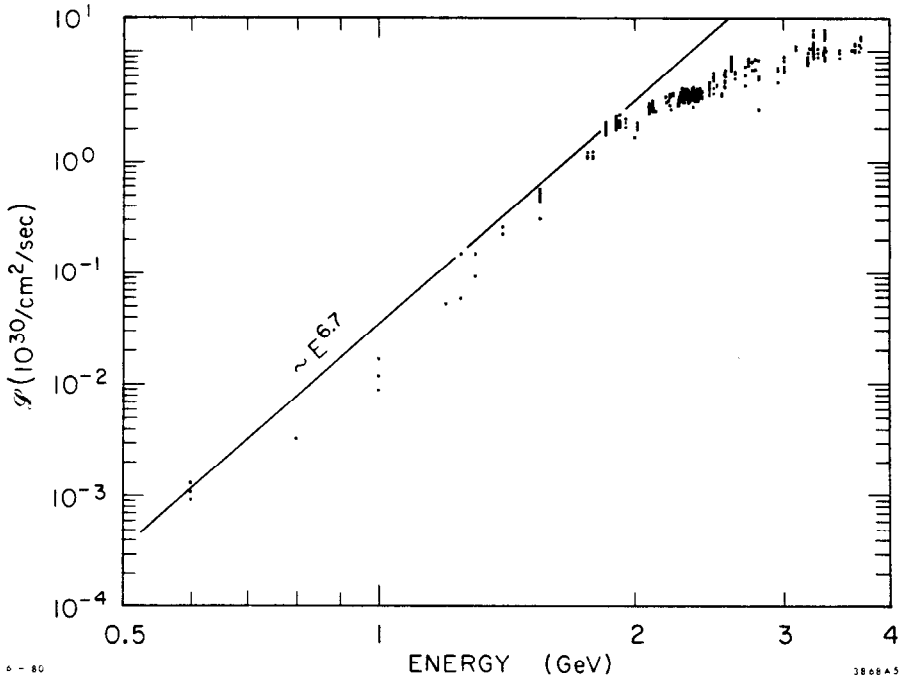


Fig. 19. Luminosity versus beam energy for the storage ring SPEAR.

15 THE DIFFUSION MODEL

The model begins with the assumption that the beam-beam kicks contain a "random" part in the sense that this part assumes a random value from one kick to the next. Clearly the linearized kicks described in Eq. (6) are totally correlated from one collision to the next and are not of interest here. The same applies to kicks considered in the single-resonance model.

It may be instructive to demonstrate that the random part of the beam-beam kick, if it exists, must be much less than the kick itself. To see this, note first that the magnitude of a typical beam-beam kick is of the order of $\Delta y' = \sigma_y / f$, where f is the beam-beam focal length defined in Eq. (7). Since the rms beam

divergence at the collision point, $\sigma_{y'}$, is equal to σ_y/β_y^*, we find

$$\frac{\Delta y'}{\sigma_{y'}} = \frac{\beta_y^*}{f} = 4\pi\xi \ . \tag{44}$$

If $\xi = 0.05$, this means $\Delta y'$ is about equal to 0.6 times $\sigma_{y'}$. In other words, one single beam-beam kick is comparable to the natural divergence of the beam! If these kicks were uncorrelated from one collision to the next, the beam-beam interaction would not allow the beams to be stored in the storage ring at all.

Let the random part of the beam-beam kicks be written as

$$\frac{\Delta y'_{\text{random}}}{\sigma_{y'}} = 4\pi\xi\eta \tag{45}$$

where η is a phenomenological constant yet to be determined from experimental data. We will consider a flat beam.

The beam-beam random kicks contribute to a diffusion in the beam size just like the random contribution from the synchrotron radiation noise. These random contributions are counteracted by the radiation damping, and on balance this gives an equilibrium rms beam size

$$\sigma_y^2 = \sigma_0^2\left[1 + \frac{\tau_{\text{rad}}}{T_0}(4\pi\xi_0\eta)^2\right] \tag{46}$$

where σ_0 is the beam size in the absence of beam-beam collisions, T_0 is the time between collisions, and ξ_0 is the beam-beam parameter calculated without beam-beam blow-up.

To get a rough idea, let us say that the beam size is doubled at the beam-beam limit of $\xi_0 = 0.05$. If we then take $\tau_{\text{rad}} = 10$ msec and $T_0 = 1$ μsec, we find $\eta = 0.03$, i.e., about 3% of the beam-beam kick strength is random at the beam-beam limit.

If the beam-beam instability is caused by an aperture limitation, then the beam-beam limit is reached when σ_y is equal to a certain defined fraction (say, 1/10) of the aperture. In case the beam-beam blow-up is significant, the beam-beam limit will be determined by

$$\frac{\text{aperture}}{10} \approx \sigma_0\left(\frac{\tau_{\text{rad}}}{T_0}\right)^{1/2} 4\pi\xi_0\eta \ . \tag{47}$$

Assuming that η is energy independent and keeping in mind that $\sigma_0 \propto \gamma$ and $\tau_{rad} \propto \gamma^{-3}$, we find, at the beam-beam limit,

$$\xi_0 \propto \gamma^{1/2} ,$$

$$\xi \propto \gamma^{3/2} ,$$

$$N \propto \gamma^{7/2} ,$$

$$\mathcal{L} \propto \gamma^{6} , \tag{48}$$

where ξ is the beam-beam parameter calculated with the blown-up beam. Note that the luminosity is proportional to γ^6, in reasonable agreement with Eq. (43).

The beam-beam diffusion model just described has other implications. For example, it predicts that the main beam-beam effect is a simple blow-up of the beam size rather than some dynamical instability. It does not have any sensitive dependence on the tunes of the storage ring. It also predicts that the beam-beam blow-up depends on the number of collision points according to $\sigma \propto \sqrt{N}$.

One problem with the diffusion model, partly due to its *ad hoc* nature, is that it is not clear how exactly the random part can be extracted from the beam-beam force. One possible source of randomness may be attributed to the stochastic motion when resonances overlap.[62]

Another possible source of randomness, which can be shown to be a very weak effect, is included here for amusement. It results from the fact that the beam is really not a smooth distribution of charges but, rather, a collection of a large number of discrete point charges. One can imagine that every time a particle collides with an on-coming beam, the average distribution of the on-coming beam is the same but the detailed distribution of the point charges within this average distribution has been randomly rearranged.

Consider then a Coulomb collision between the particle and a particle in the on-coming beam with an impact parameter b. The kicking angle is given by

$$\Delta y' = \frac{2r_0}{\gamma b} . \tag{49}$$

If the impact parameter is less than a certain minimum value $b_{\min}$, the collision will be so strong that both particles will be ejected from the storage ring acceptance because of this single Coulomb collision. The quantity $b_{\min}$ is given by

$$b_{\min} = \frac{2r_0}{\gamma \Delta \hat{y}'} \tag{50}$$

where $\Delta\hat{y}'$ is the maximum value of y' at the collision point that can be accepted by the storage ring acceptance.

These violent Coulomb collisions cause a continuous loss of particles. The beam lifetime τ, because of this effect,[7] is determined by the probability of a particle finding itself within a distance $b_{\min}$ from a particle in the on-coming beam. We then find

$$\frac{1}{\tau} = \frac{N\, b_{\min}^2}{T_0 \sigma_x \sigma_y} \ . \tag{51}$$

If we take $N = 10^{12}$, $\Delta\hat{y}' = 5$ mrad, $\sigma_x = 2$ mm, $\sigma_y = 0.05$ mm, $T_0 = 1$ μsec, and a beam energy of 10 GeV, we find that $b_{\min} = 6 \times 10^{-17}$ m and the beam lifetime is about 8000 hours.

For those particles that remain within the storage ring acceptance, there is a diffusion effect on the beam size. Very roughly,[8] each collision will contribute to the beam divergence by an rms amount of $\Delta y'/\sqrt{N}$, where $\Delta y'$ is the typical beam-beam kick given by Eq. (44). The growth rate for this diffusion process is therefore

$$\frac{1}{\tau} = \frac{(4\pi\xi)^2}{N\, T_0} \tag{52}$$

which is a much weaker growth than the radiation damping. Compared with the diffusion process described by Eq. (45), this effect is equivalent to replacing η by $1/\sqrt{N}$.

ACKNOWLEDGEMENTS

I have benefitted from discussions with K. Bane, L. Rivkin, D. Helm, A. Hofmann, S. Kheifets, P. Morton, and J. Seeman in preparing this material. George Bell has contributed substantially in writing several programs and then preparing the plots shown in Figures 3, 5, 8, 9, and 10. The work on the coherent quadrupole mode instability was done in collaboration with Yuri Kamiya, a visitor from the KEK laboratory.

---

[7] Another perhaps more serious lifetime restriction comes from the bremsstrahlung scattering of the two electrons. The present consideration, however, is sufficient to illustrate the point.

[8] A more careful calculation would probably introduce a logarithmic factor $\ln(\sigma/b_{\min})$, but this is ignored here.

REFERENCES

1. Nonlinear Dynamics and the Beam Beam Interaction, M. Month and J. C. Herrera, eds., AIP Conference Proceedings 57, Brookhaven National Laboratory, 1979.
2. Proceedings of the Beam Beam Interaction Seminar, SLAC-PUB-2624, 1980.
3. Proceedings of the 11th International Conference on High Energy Accelerators, pp. 701-70, CERN, 1980.
4. S. Kheifets, SLAC-PUB-2700, 1981.
5. J. F. Schonfeld, Summer School on High Energy Particle Accelerator, SLAC, 1982.
6. B. W. Montague, CERN/ISR-GS/75-36 (unpublished), 1975.
7. J. E. Augustin, SLAC note PEP-63 (unpublished), 1973.
8. E. D. Courant and H. S. Snyder, Ann. Phys. 3, 1 (1958).
9. F. Amman and D. Ritson, 1961 International Conference on High Energy Accelerator, p. 471 Brookhaven National Laboratory, 1961.
10. A. J. Dragt, Summer School on High Energy Accelerator, Fermi Lab, 1981.
11. J. M. Greene, Ref. 2, p. 235.
12. L. Jackson Laslett, Proceedings of the 9th International Conference on High Energy Accelerator, p. 394, Stanford, 1974. More references are quoted in this paper.
13. A. G. Ruggiero and L. Smith, SLAC Note PEP-52 (unpublished), 1973.
14. A. Lejcic and L. Le Duff, Proceedings of the 8th International Conference on High Energy Accelerator, p. 354, CERN, 1971.
15. E. Keil, CERN/ISR-TH/72-7 and CERN/ISR-TH/72-25 (unpublished), 1972.
16. A. Chao, Ref. 1, p. 42.
17. B. W. Montague, Ref. 3, p. 731.
18. M. Month, IEEE Trans. Nucl. Sci. NS-22, 1376 (1975).
19. H. G. Hereward, CERN/ISR-DI/72-26 (unpublished), 1972.
20. J. Le Duff, CERN/ISR-AS/74-53 (unpublished), 1974.
21. D. Neuffer and A. Ruggiero, Ref. 2, p. 332.

22. B. V. Chirikov, Phys. Rep. 52, 263 (1979).

23. J. Tennyson, Ref. 1, p. 158.

24. E. D. Courant, Ref. 3, p. 763.

25. V. I. Arnold, Dokl. Akad. Nauk. 156, 9 (1964).

26. B. V. Chirikov, E. Keil, and A. M. Sessler, CERN/ISR-TH/69-59 (unpublished), 1969.

27. T. C. Bountis, Ref. 2, p. 248.

28. L. C. Teng, IEEE Trans. Nucl. Sci. NS-20, 843 (1973).

29. F. Amman et al., International Conference on High Energy Accelerators, Dubna, USAEC Conference – 114, 1963, p. 309; B. Richter, Proceedings of the International Symposium on Electron and Positron Storage Rings, p. I-1-1, Saclay, 1966.

30. SPEAR Storage Ring Group, IEEE Trans. Nucl. Sci. NS-20, 838 (1973).

31. J. Le Duff, SLAC Note PEP-65, 1973.

32. S. Kheifets, Part. Accelerators, Vol. 15, No. 3, 153 (1984).

33. K. W. Robinson and G.-A. Voss, Cambridge Electron Accelerator Report CEAL-TM-149 (unpublished), 1965.

34. P. L. Morton and J. R. Rees, IEEE Trans. Nucl. Sci. NS-14, 630 (1967).

35. J. Rossbach et al., IEEE Trans. Nucl. Sci. NS-28, 2025 (1981).

36. J. Seeman, Cornell University Report CLNS 82/531 (unpublished), 1982.

37. R. Helm et al., Particle Accelerator Conference, Santa Fe, SLAC-PUB-3070, 1983.

38. J. Rees and L. Rivkin, Summer School on High Energy Particle Accelerator, SLAC, 1982.

39. J. M. Paterson, J. Rees, and H. Wiedemann, SLAC Note PEP-125 (unpublished), 1975.

40. R. H. Helm, M. J. Lee, and J. M. Paterson, Proceedings of the 9th International Conference on High Energy Accelerators, p. 100, Stanford, 1974.

41. A. Piwinski, DESY Report 77/18, 1977.

42. L. Smith, SLAC Note PEP-20 (unpublished), 1972.

43. G. Fischer, SLAC Note SPEAR-154 (unpublished), 1972.

44. S. Myers, Nucl. Instrum. Methods 211, 263 (1983).

45. A. Piwinski, Proceedings of the 8th International Conference on High Energy Accelerators, p. 357, CERN, 1971.

46. A. Chao and E. Keil, CERN/ISR-TH/79-31 (unpublished), 1979.

47. E. Keil, Ref. 3, p. 759.

48. SPEAR Group, Proceedings of the 9th International Conference on High Energy Accelerators, p. 66, Stanford, 1974.

49. A. Piwinski, Ref. 1, p. 115.

50. Y. Kamiya and A. Chao, SLAC Note, in preparation, 1983.

51. K. L. Brown, SLAC Report 75, 1967.

52. Ya. S. Derbenev, Proceedings of the 3rd All Union Particle Accelerator Conference, p. 382, Moscow, 1973.

53. N. N. Chau and D. Potaux, Orsay Tech. Report 5-74 and 2-75 (unpublished), 1974 and 1975.

54. R. E. Meller and R. H. Siemann, IEEE Trans. Nucl. Sci. NS-28, 2431 (1981).

55. J. Le Duff et al., Ref. 3, p. 707.

56. A. Piwinski, IEEE Trans. Nucl. Sci. NS-28, 2440 (1981).

57. S. Peggs and R. Talman, Ref. 3, p. 754.

58. E. Keil, Nucl. Instrum. Methods 188, 9 (1981).

59. M. Sands, SLAC-121, 1970.

60. H. Wiedemann, Ref. 2, p. 33.

61. S. Kheifets, Ref. 2, p. 40.

62. L. Teng, Ref. 2, p. 99.