

Theories of Statistical Equilibrium in Electron- Positron Colliding-Beam Storage Rings*

Jonathan F. Schonfeld
Fermi National Accelerator Laboratory
P.O. Box 500
Batavia, Illinois 60510

I. INTRODUCTION

In this lecture I want to introduce you to some recent theoretical work ([1]-[7]) that represents a significant and long overdue departure from the mainstream of ideas on the physics of colliding-beam storage rings. The goal of the work in question is to understand analytically--without recourse to computer simulation--the role that dissipation and noise play in the observed colliding-beam behavior of electron-positron storage rings. More traditional approaches have given such

*Invited Lecture given at Joint US/CERN School on Particle Accelerators, Topical Course on Nonlinear Dynamics, 31 January- 5 February 1985, Santa Margherita di Pula, Sardinia.

"thermal" processes relatively little attention, with results that have been qualitative at best. For a comprehensive overview of the conventional theoretical literature, see references [8]-[11].

Here are the colliding-beam phenomena that storage ring theories must explain:

-In proton-proton/antiproton colliders, enhanced beam loss when the machine tunes approach nonlinear resonances of anomalously high order.

-In electron-positron colliders, strong tune-dependence of loss rates and luminosity; anomalously low maximum storable currents, and beam heights that grow ("blow up") markedly with beam currents, with consequent degradation of luminosity. For a comprehensive picture of colliding beam storage ring phenomenology, see references [8],[9],[12], and [21].

Strictly speaking, theories that ignore damping and/or noise altogether can provide a complete ideal description only of proton-proton/ antiproton colliders. The dominant source of damping and noise in a ring with otherwise quiet components is synchrotron radiation. In rings that store protons and/or antiprotons, radiation damping times are between days (SSC) and months (Sp \bar{p} S, Tevatron). In electron-positron rings, radiation damping times are typically measured in milliseconds.

In the current state of the art, a theory of colliding effects in a storage ring can include--apart, possibly, from radiative damping and fluctuation--linear transverse restoring forces due to quadrupole magnets, longitudinal oscillations, and coulombic beam-beam encounters, idealized as instantaneous nonlinear kicks. Magnet nonlinearities, wake fields, beam-gas scattering, inelastic or non-electromagnetic inter-beam interactions, and intrabeam process are generally ignored. When computer simulation is not involved, it is common in addition to assume--unrealistically--that the current in one beam is much smaller than the current in the other. In such a "weak-strong" limit, the high current, strong beam is unaffected by collisions, while the weak beam reduces to an ensemble of uncoupled

particles, all independently perturbed by the opposing beam. For a discussion of the support that computer simulation provides for such simplifications, see reference [9], and the contribution of Myers to these Proceedings.

Traditionally--excepting the work to be described below--theorists have chosen to analyze such models in terms of Hamiltonian chaos, nonchaotic Hamiltonian nonlinear resonance, coherent (rigid body) bunch motion, and collective (plasma) oscillations. The balance between these effects and the damping and noise in electron-positron storage rings has largely been ignored. This style of analysis has produced a successful theory of ISR beam loss under certain conditions [13], and is also claimed to provide an account of Sp̄p̄S operating limits [14].

This style of analysis often begins by identifying regions of phase space in which unstable processes can act. Such identifications are meaningless, however, unless one can estimate the beam populations that occupy such regions. In the case of protons and antiprotons, these populations can in principle be obtained, over limited but non-negligible periods of time, directly from beam distributions at injection, and such distributions are understood well enough for some purposes. By contrast, in the case of the more strongly damped electrons and positrons, luminosity, loss rates, etc. can only be obtained from distributions that must be presumed to have relaxed rapidly to an equilibrium that is largely independent of initial conditions, and which therefore must be calculated in an entirely self-consistent manner, from first principles.

In this lecture, we shall consider three schemes that have recently been proposed for carrying out analytical calculations of equilibrium beam distributions, or related quantities, in weak-strong electron-positron colliding beam storage rings. In section II we shall discuss the formalism of Kheifets ([1]-[3]); in section III we shall discuss the formalism of F. Ruggiero [4]; and in section IV we shall discuss my own formalism [5]-[7].

Before proceeding to detailed descriptions, let us first place the physical problem at hand in the larger context of statistical mechanics in general, and let us also highlight in advance some of the more important differences between the three approaches under discussion.

The problem of electron and positron beams colliding in a storage ring differs from the familiar textbook paradigms of statistical mechanics (or, more narrowly, of stochastic differential equations) in two principal ways:

First the notion of temperature cannot be formulated in a useful way for an electron-positron storage ring. Indeed, under the most idealized circumstances (azimuth-independent fields, no coupling) the dynamics of synchrotron radiation [15] would have us assign no less than three very different phenomenological temperatures to the three dimensions in which an electron or positron oscillates about its bunch center. Under more realistic circumstances, even such a fragmented sort of temperature has no practical meaning.

Second, the environment of electrons or positrons in a storage ring varies periodically on a timescale much shorter than the timescale of (the analogue of) thermal relaxation: An electron or positron receives a delta-function kick from some bunch in the opposing beam several times per revolution, while the damping time is $O(10^3)$ - $O(10^4)$ revolutions. Thus, such a system is far from the adiabatic limit. In particular, even if the notion of temperature could be formulated for a storage ring, one could not bypass a complicated dynamical calculation in order automatically to conclude that the equilibrium beam distribution is obtained by exponentiating a time-dependent Hamiltonian, as in the familiar Boltzmann factor.

For these reasons, the problem of electron-positron storage rings requires the development of new theoretical techniques.

The formalisms of Kheifets, Ruggiero, and myself are all perturbative approaches to the calculation of equilibrium beam distributions. They differ in the choice of perturbative parameter(s). Kheifets uses the strong-beam (vertical)

tunesift ξ , which is typically a few percent in high energy e^+e^- colliders. Ruggiero uses both ξ and the transverse damping rate (normalized to revolution frequency T^{-1}) $T\gamma$, which, as already indicated, is $O(10^{-4})$ - $O(10^{-3})$. (In this review we shall not distinguish between vertical and horizontal damping.) I use only γ as a perturbative parameter. (Note that in all cases γ also sets the scale of anisotropic noise, as we explain below.)

Our experience with nonstatistical Hamiltonian mechanics teaches us to be wary of perturbative expansions in nonlinearity strengths such as ξ . We expect to encounter spurious unbounded time dependences that mask frequency shifts, and/or "small denominator" singularities that mask the appearance of topologically distinctive resonance structure in phase space. Resonant small denominators appear both in Kheifets' and in Ruggiero's calculations, and in either case it is the presence of damping and noise that keeps their expressions from becoming infinite.

My formalism was developed specifically to bypass small denominators, by building the true perturbed distortion of Hamiltonian orbits directly into the definitions of the variables used to parametrize phase space. Thus my scheme is uniquely suited to critical analysis of the resonant processes traditionally invoked to account qualitatively for colliding beam behavior. Ironically, the small- γ limit seems in principle to bring about its own variant of the small denominator problem, although computer simulation of some simple models [16] seems so far to indicate that this is not very important.

Some other comparisons worth underscoring: None of these formalisms has yet been extended to encompass longitudinal phase space. Indeed, only Kheifets has succeeded in carrying out his calculations without disregarding one of the transverse degrees of freedom. Accordingly, only Kheifets' calculations have been meaningfully compared with experimental data, with some interesting results.

These three approaches have in common their use of the Fokker-Planck equation to define the time evolution of weak-beam phase space densities. If the displacement of a weak-beam particle from its bunch center develops in time according to

$$\ddot{\vec{x}}_i + \gamma(t)\dot{\vec{x}}_i - F_i(\vec{x}, t) = \lambda_i(t)\zeta_i(t) \quad , \quad (1.1)$$

where the ζ_i are uncorrelated sources of normalized Gaussian white noise, then, for smooth γ, \vec{F} , and $\vec{\lambda}$ the Fokker-Planck equation for the weak-beam distribution $P(\vec{x}, \vec{v} \equiv \dot{\vec{x}}, t)$ is

$$\frac{\partial P}{\partial t} + \sum_i \left(v_i \frac{\partial P}{\partial x_i} + F_i \frac{\partial P}{\partial v_i} \right) = \sum_i \frac{\partial}{\partial v_i} \left(\gamma v_i + \frac{\lambda_i^2}{2} \frac{\partial}{\partial v_i} \right) P \quad . \quad (1.2)$$

(For a discussion of simplifications made in arriving at (1.1), see [7]. For pedagogical simplicity, we ignore the distinction between time and the more common azimuth parameter s .) The quantities γ, \vec{F} , and $\vec{\lambda}$ are all periodic as explicit functions of time. (Strictly speaking, we ought earlier to have referred to the time average, $\langle \gamma \rangle$, of γ , rather than to γ itself.) As a function of \vec{x} , \vec{F} is conventionally the sum of a linear part, due to quadrupole magnets, and the beam-beam contribution

$$\varepsilon \sum_{k=-\infty}^{+\infty} \delta(t-t_k) \vec{\nabla} f_k(\vec{x}) \quad , \quad (1.3)$$

where the t_k are collision times. The sequence of potentials $\{F_k\}$ repeats as a function of k . The sequence $\{t_{k+1} - t_k\}$ of collision-time spacing also repeats, with the same period. In what follows, the careful reader will recognize precautions that have been taken to avoid misusing (1.2) at collision times, where \vec{F} is

certainly not smooth.

In the small γ limit, λ is generally taken to scale as $O(\sqrt{\gamma})$. When this is done, then we can be sure that the equilibrium solution--if it exists--to (1.2) has a nonsingular $\gamma \rightarrow 0$ limit, at least in the absence of beam-beam interactions (and away from linear resonance [7]).

II. KHEIFETS' FORMALISM

In order to unify the discussions in this and the next two sections, my presentation of Kheifets' result will differ somewhat from Kheifets' own. I will indicate briefly how Kheifets actually proceeded when I reach the point at which our two presentations begin to coincide.

For simplicity (Kheifets himself does not make this simplification), we shall imagine only one interbunch collision per revolution, so that all f_k are identical ($\equiv f$) and all $t_k \equiv kT$.

We seek as an equilibrium a periodic solution P of (1.2), with period T . Because (1.2) (or (1.1) in $\vec{x}-\vec{v}$ space) is first order in the time derivative, it is sufficient to demand that

$$\lim_{t \rightarrow kT^-} P(\vec{x}, \vec{v}, t) \equiv P(\vec{x}, \vec{v}, kT^-) \quad (2.1)$$

be independent of the integer k . We guarantee this inductively by equating $P(\vec{x}, \vec{v}, 0^-)$ to $P(\vec{x}, \vec{v}, T^-)$, the result of evolving $P(\vec{x}, \vec{v}, 0^-)$ through time T . To evolve $P(\vec{x}, \vec{v}, 0^+)$ to $P(\vec{x}, \vec{v}, T^-)$, we write

$$P(\vec{x}, \vec{v}, T^-) = \int d^2x_0 d^2v_0 G_0(\vec{x}, \vec{v}, T; \vec{x}_0, \vec{v}_0, 0) P(\vec{x}_0, \vec{v}_0, 0^+) = G_0 P(0^+) \quad , \quad (2.2)$$

where G_0 is the Green's function for (1.2) with the beam-beam kicks omitted. With linear magnets, G_0 can be calculated exactly in closed form (see, e.g. [7]), and is a

Gaussian in $\vec{x}, \vec{v}, \vec{x}_0$, and \vec{v}_0 . To evolve P from 0^- to 0^+ , we use (1.1) and (1.3) to write

$$P(0^+) \equiv P(\vec{x}, \vec{v}, 0^+) = P(\vec{x}, \vec{v} - \xi \vec{\nabla} f(\vec{x}), 0^-) \quad (2.3)$$

So, combining (2.2) and (2.3), our equilibrium is defined (up to normalization) by

$$P(0^-) = GP(0^-) \equiv \int d^2x_0 d^2v_0 G_0(\vec{x}, \vec{v}, T; \vec{x}_0, \vec{v}_0 + \xi \vec{\nabla} F(\vec{x}_0), 0) P(\vec{x}_0, \vec{v}_0, 0^-) \quad (2.4)$$

It is a simple matter to expand (2.4) in powers of ξ . With $P(0^-) \equiv P_0 + \xi P_1 + \dots$, and with

$$G = \sum_n \frac{1}{n!} [\xi \vec{\nabla} f(\vec{x}_0) \cdot \vec{\nabla}_{v_0}]^n G_0 \equiv G_0 + \xi G_1 + \dots \quad (2.5)$$

we have

$$P_0 = G_0 P_0 \quad ,$$

$$P_1 = G_0 P_1 + G_1 P_0 \quad , \quad (2.6)$$

$$P_2 = G_0 P_2 + G_1 P_1 + G_2 P_0 \dots ,$$

with formal solutions

$$P_1 = (1-G_0)^{-1}G_1P_0$$

$$P_2 = (1-G_0)^{-1}G_1P_1 + (1-G_0)^{-1}G_2P_0 + \dots \quad (2.7)$$

$$= [(1-G_0)^{-1}G_1]^2P_0 + (1-G_0)^{-1}G_2P_0 + \dots$$

(Note that if the right hand sides of (2.6)--with G_0P_0, G_0P_1, G_0P_2 etc. omitted--had nontrivial overlap with that solution of the time-reversed adjoint of (1.2) that is dual to P_0 , then the inverses in the right hand sides of (2.7) could not be defined. However, the dual in question is in fact constant in $\vec{x}-\vec{v}$ space, so that the overlaps in question are zero because of the powers of \vec{v}_{v_0} in the definition of $G_{n \neq 0}$. This still leaves the operation $(1-G_0)^{-1}$ determined only up to addition of a constant multiple of P_0 , which serves to adjust the overall normalization. This adjustment is rendered unnecessary by the next step, and its physical interpretation.)

Kheifets' general perturbative expressions can be obtained from (2.7) by expanding in powers of G_0 :

$$P_1 = \sum_{m=0}^{\infty} G_0^m G_1 P_0$$

$$P_2 = \sum_{m,n=0}^{\infty} G_0^m G_1 G_0^n G_1 P_0 + \sum_{m=0}^{\infty} G_0^m G_2 P_0 + \dots \quad (2.8)$$

Kheifets himself obtains the series' in (2.8) directly, by perturbatively calculating the time-dependent solution of (1.2) that equals P_0 at $t=0^-$. Presuming that any solution of (1.2) relaxes to the desired equilibrium--but making no a priori assumption that "equilibrium" need be periodic--he arrives at what we call $P(0^-)$ by evaluating his solution at $t=kT^-$ (integral k) and taking the limit of

infinite k . (The powers of G_0 arise in this approach from unperturbed evolution through all possible integral multiples of T .) This procedure makes it clear that the normalization of P_0 carries over to the $P_0 + \xi P_1 + \dots$ that corresponds to (2.8), since time evolution according to (1.2) preserves normalizations, by construction. Note that if γ were regarded as a perturbation parameter, then this procedure would not have been possible, because we expect time evolution approximated by finitely many powers of γ to look like relaxation only until $t \sim 0(1/\langle \gamma \rangle)$, and thereafter to approach no definite limit.

Kheifets confines his explicit calculation to the mean square vertical beam amplitude (i.e. emittance) going into an interaction point in equilibrium,

$$\frac{1}{2} \langle x_2^2 + \beta_2^{*2} v_2^2 \rangle = \frac{1}{2} \int [x_2^2 + \beta_2^{*2} v_2^2] P(\vec{x}, \vec{v}, 0^-) d^2 x d^2 v, \quad (2.9)$$

where β_2^* is the vertical beta function at an interaction point, and P is assumed normalized to unity. In comparisons with experiment, Kheifets uses (2.9) as a measure of the mean square weak beam half-height, $\langle x_2^2 \rangle$ (even though (2.9) and $\langle x_2^2 \rangle$ are really exactly equal only when $\xi=0$, and far from linear resonance [7]). He finds that when (2.8) is substituted into (2.9), and summations and integrations interchanged, then, through second order in ξ , every term in every summation can be evaluated in closed form, assuming (as is appropriate for e^+e^- machines [15]) that $\langle x_2^2 \rangle \ll \langle x_1^2 \rangle$ when $\xi=0$. This discovery rests on the fact that both P_0 and G_0 are Gaussians, and that the collision potentials f_k can all be expressed in terms of integrals over Gaussians.

With this result in hand, Kheifets finds, further, that the full order $-\xi$ infinite summation can be carried out in closed form, and that the doubly infinite sum in order ξ^2 can be reduced in a simple way to a singly infinite sum, which must be evaluated numerically.

We need not reproduce Kheifets' exact expression here. However, one aspect of the final infinite sum in $O(\xi^2)$ is worth highlighting: The sum has the form

$$\sum_{q=0}^{\infty} h(e^{-\langle \gamma \rangle q}, e^{2\pi i v_1 q}, e^{2\pi i v_2 q}) \quad , \quad (2.10)$$

where the v_i are the unperturbed storage ring tunes. When the function h is expanded in powers of its arguments, (2.10) becomes an infinite series of subsums proportional to

$$\sum_{q=0}^{\infty} [e^{2\pi i \vec{n} \cdot \vec{v} - n \langle \gamma \rangle}]^q = [1 - e^{2\pi i \vec{n} \cdot \vec{v} - n \langle \gamma \rangle}]^{-1} \quad , \quad (2.11)$$

for integral \vec{n} and n . Thus we expect vertical emittance to show pole-like peaks in the tune plane, at $\vec{n} \cdot \vec{v} = \text{integer}$, with infinities regularized by nonzero damping.

In some calculations, Kheifets includes the linear part of (1.3) (i.e. with the f_k expanded in powers of \vec{x} and truncated at second order) in the unperturbed system, in order to treat at least part of the beam-beam interaction exactly. When this is done, (2.11) implies pole-like peaks in weak beam vertical emittance as a function of strong beam current, since in this case \vec{v} depends on ξ , and ξ is proportional to strong beam current, by definition. These peaks are the distinctive predictions of this formalism.

(Note that when the linear part is removed from $\vec{v}f$, then the remainder--i.e. the new perturbation--is left with spurious linear growth at large \vec{x} , because the full $\vec{v}f$ actually falls as $1/|\vec{x}|$ at large \vec{x} . Presumably, this large- \vec{x} distortion has little effect on the mean emittance, since the equilibrium distribution should in any event be sparse far from the beam center.)

Figure 1 (next page), taken from [3] (with permission), shows comparison between Kheifets' calculations and the results of a weak/strong PEP experiment. The vertical axis represents the ratio of weak-beam height with to without collisions.

The solid (dashed) curves correspond to calculations that do (not) include the linear part of the beam-beam interaction in the unperturbed system. Different curves of the same type correspond to different phenomenological choices of storage ring parameters. None of the curves seems to fit this data particularly well, although the dashed curves seem to do better than the the solid ones. The situation in Figure 2, corresponding to a SPEAR experiment (also from [3], with permission) is more encouraging. All curves in this figure were calculated with the linear part of the collisions included in the unperturbed system. The abrupt climb in the data at about 9mA is the kind of structure that one would expect from the pole terms

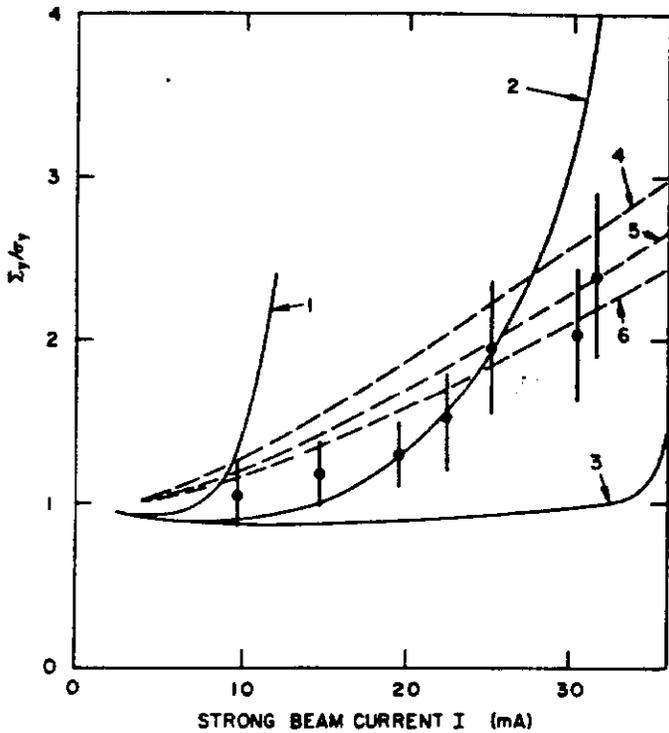


Figure 1

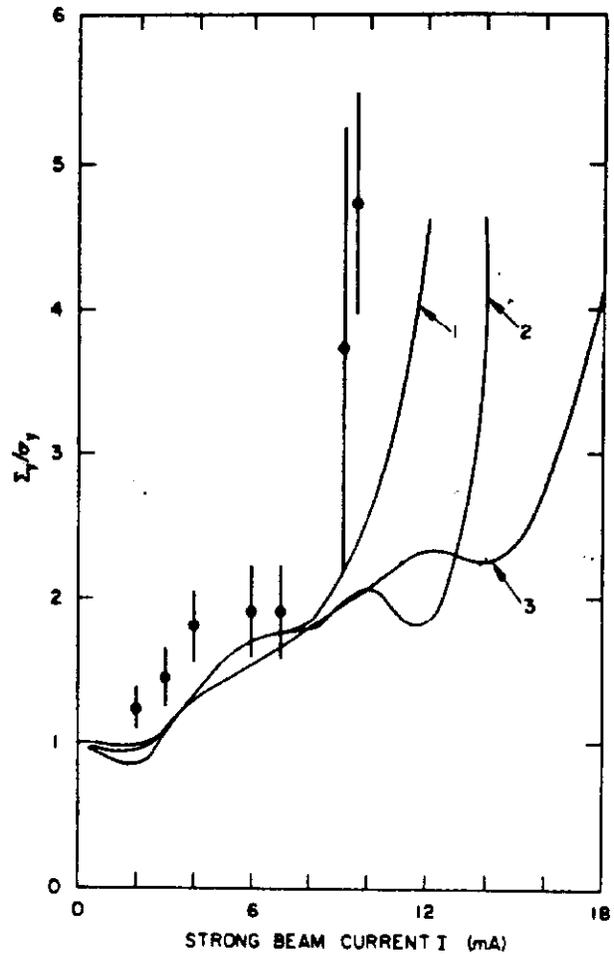


Figure 2

characteristic of this theory. Curves 1 and 2, which come the closest to fitting this rise, also give encouraging fits to the shoulder in the region 2-6 mA. Unfortunately, whereas curve 1 appears to give the best fit overall, the parameters assumed in obtaining curve 2 are in fact the closest to the nominal parameters of the real storage ring.

The reader who is not impressed with Figures 1 and 2 should realize that Kheifets' are the first analytical calculations even to obtain the right order of magnitude for e^+e^- beam blowup.

It is worth noting that Kheifets partially compensates for his omission of longitudinal oscillations by making provision in his calculations for nonidentical interaction regions and intercollision betatron phases, which computer simulations have strongly implicated in beam blowup.

III. RUGGIERO'S FORMALISM

As mentioned in the introduction, Ruggiero restricts himself to models with only one dynamical degree of freedom (vertical, conventionally, since beams do not blow up horizontally); and, as in the preceding section, with only one beam-crossing per machine period T .

He begins by noting that when $\gamma=\xi=0$, then the operator G , which in this limit we shall call g_0 , can easily be diagonalized in closed form. In particular, let J and ϕ be the canonical action and angle for the periodic linear system with vanishing γ and ξ ; then for integral n ,

$$g_0[p(J)e^{in\phi}] = [p(J)e^{in\phi}]e^{-2\pi i\nu n}, \quad (3.1)$$

for any $p(J)$, where ν is the unperturbed storage ring tune. For this reason, Ruggiero tailors his formalism to the harmonic expansion

$$P(0^-) = \sum_{n=-\infty}^{+\infty} p^n(J) e^{in\phi} \quad , \quad (3.2)$$

where the coefficient functions $p^n(J)$ are to be computed perturbatively.

The distribution P_0 introduced in the preceding section turns out to correspond to $n=0$, at least for small γ and far from linear resonance [7]. Thus Ruggiero conducts his perturbative analysis under the assumption that p^0 is much larger than all but at most one ($n=N$, for definiteness) $p^{n \neq 0}$.

Because g_0 is diagonal in the basis (3.2), Ruggiero separates out the diagonal from the nondiagonal terms in (2.4), writing, in an obvious notation,

$$(1-G^{nn})p^n = \sum_{m \neq n} G^{nm} p^m \quad . \quad (3.3)$$

Note that the objects G^{nm} in (3.3) are now operators on functions of J only. (For continuity's sake, we have rearranged Ruggiero's expressions in inessential ways, in presenting (3.3) and the manipulations that follow.)

For small γ and ξ , the G^{nm} with $n \neq m$ are all naively of roughly the same (small) size, while presumably $p^{0,N} \gg p^{n \neq 0, N}$, so that Ruggiero replaces (3.3) by

$$(1-G^{nn})p^n \approx G^{n0} p^0 + G^{nN} p^N \quad (3.4)$$

for $n \neq 0$. The second term on the right is absent for $n=N$. For $n=0$, he substitutes (3.4) into (3.3) in order to obtain

$$(1-G^{00})p^0 \approx \sum_{m \neq 0} G^{0m} (1-G^{mm})^{-1} G^{m0} p^0 \quad . \quad (3.5)$$

(Terms involving p^N that one might have expected to see on the right turn out to be subdominant [22].) Ruggiero calls this the "renormalized Fokker-Planck equation," because, after the approximations that follow, it is formally identical to the

equation that one would write for $\xi=0$, but with extra ξ - dependent contributions to what would have been the noise term.

For future reference, note that the existence of at most one large $p^{n \neq 0}$ implies that at most one term ($m=N$) in (3.5)'s sum can be appreciable.

In order to find simple but meaningful approximations to G^{nm} and G^{m0} , Ruggiero makes the explicit assumption that

$$\xi \ll \left(\frac{\gamma T}{\xi}\right)^{1/3} \ll 1 \quad . \quad (3.6)$$

Physically, these inequalities ensure that damping and noise act much more slowly than any other dynamical process. In practice, Ruggiero seems to rely on a hierarchy of orders of magnitude that is sharper than (3.6), because in several expressions he expands through second order in ξ and through first order in γ . Thus, the true state of affairs seems to correspond to $\gamma T \sim O(\xi^2)$, which is consistent with (3.6). We shall comment later on the physical appropriateness of this identification.

In order to understand some of the manipulations that follow, it is useful to know how to represent the full operator G --including perturbations--as a product of exponentials. For simplicity, let us assume time-independent damping, and time-independent unperturbed linear force $-x(2\pi\nu/T)^2 \equiv -x\omega^2$. Then

$$G = \exp T \left[\gamma \frac{\partial}{\partial v} \left(v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right) + x \omega^2 \frac{\partial}{\partial v} - v \frac{\partial}{\partial x} \right] \cdot \exp \left[-\xi \frac{df}{dx} \frac{\partial}{\partial v} \right], \quad (3.7)$$

where σ^2 is the unperturbed mean-square half-height of the beam. With

$$J \equiv \frac{1}{2}(\omega x^2 + v^2/\omega) \quad , \quad \phi = -\tan^{-1} \frac{v}{x \omega} \quad , \quad (3.8)$$

the γ -independent terms in the first exponent add to $-2\pi v \partial/\partial \phi$, from which one directly recovers (3.1).

It follows easily from the foregoing that the leading perturbative contributions to G^{m0} and G^{0m} , for $m \neq 0$, are

$$G^{m0} \equiv e^{-2\pi i m v} \left[-\xi \frac{df}{dx} \frac{\partial}{\partial v} \right]^{m0}$$

$$G^{0m} \equiv \left[-\xi \frac{df}{dx} \frac{\partial}{\partial v} \right]^{0m} \quad , \quad (3.9)$$

and that the leading perturbative contribution to $1-G^{00}$ is

$$(1-G^{00}) \equiv -T\gamma \left[\frac{\partial}{\partial v} \left(v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right) \right]^{00} - \frac{1}{2} \left[\left(\xi \frac{df}{dx} \frac{\partial}{\partial v} \right)^2 \right]^{00}$$

$$\equiv -T\gamma \left[\frac{\partial}{\partial v} \left(v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right) \right]^{00} - \frac{1}{2} \sum_{m \neq 0} e^{2\pi i m v} G^{0m} G^{m0} \quad . \quad (3.10)$$

(We can exclude $m=0$ from the summation in (3.10) because the 00 component of $\frac{df}{dx} \frac{\partial}{\partial v}$ vanishes identically.) By virtue of (3.10), the renormalized Fokker-Planck

equation becomes

$$T\gamma \left[\frac{\partial}{\partial v} \left(v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right) \right]^{00} p^0$$

$$+ \sum_{m \neq 0} G^{0m} \left[\frac{1}{2} e^{2\pi i m v} + (1-G^{mm})^{-1} \right] G^{m0} p^0 \equiv 0 \quad . \quad (3.11)$$

At this point the natural temptation is to supplement (3.9) by replacing $(1-G^{mm})^{-1}$ with its leading behavior for small ξ and γ , namely, $(1-e^{-2\pi imv})^{-1}$. However, Ruggiero does not do this for at least two reasons. First, if G^{mm} were replaced by $e^{-2\pi imv}$, then the square brackets under the summation in (3.11) could effectively be replaced by

$$e^{2\pi imv} \left[\frac{1}{2} + \frac{1}{2} (e^{2\pi imv} - 1)^{-1} + \frac{1}{2} (e^{-2\pi imv} - 1)^{-1} \right] = 0 \quad , \quad (3.12)$$

since approximation (3.9) makes $e^{2\pi imv} G_{m0} G_{m0}^0$ an even function of m . In this way, the collision parameter ξ would disappear from (3.11) altogether, which is unacceptable. Second (although this is not stated explicitly in [4]) if G^{mm} were replaced by $e^{-2\pi imv}$, then some $(1-G^{mm})^{-1}$ would be singular whenever v was rational. Our experience with Hamiltonian dynamics leads us to expect that such singularities are artifacts of the perturbation method, and accordingly disappear when viewed in the proper light.

For the first of these reasons Ruggiero extends his approximation for $(1-G^{mm})^{-1}$ beyond the naively dominant term. For the second reason, he carries out the necessary extension in the exponents in G^{mm} , so as not to compound the presumably spurious singularities at rational v . This exponential approximation is carried out as follows: First, one naively expands the decomposition in (3.7) to orders ξ^2 and γ to obtain

$$G^{mm} \approx e^{-2\pi imv} \left\{ 1 - \xi \frac{df}{dx} \frac{\partial}{\partial v} + \frac{1}{2} \left(\xi \frac{df}{dx} \frac{\partial}{\partial v} \right)^2 + \gamma T \frac{\partial}{\partial v} \left[v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right] \right\}^{mm} \quad . \quad (3.13)$$

One then observes that this is identical to the expansion to $O(\xi^2)$ and $O(\gamma)$ of

$$\exp \left\{ -2\pi i m v + \gamma T \left[\frac{\partial}{\partial v} \left(v + \sigma^2 \omega^2 \frac{\partial}{\partial v} \right) \right]^{mm} \right\} \exp \left\{ -\xi \left[\frac{df}{dx} \frac{\partial}{\partial v} \right]^{mm} + \frac{1}{2} \xi^2 \left[\left(\frac{df}{dx} \frac{\partial}{\partial v} \right)^2 \right]^{mm} - \frac{1}{2} \xi^2 \left(\left[\frac{df}{dx} \frac{\partial}{\partial v} \right]^{mm} \right)^2 \right\} \quad (3.14)$$

This is the desired approximation to G^{mm} .

Note that this enables us to make some contact with conventional Hamiltonian theory, because, to leading order in ξ , (3.14) is equivalent to

$$\exp - 2\pi i m \left[v - \frac{\xi}{2\pi} \left(\frac{d}{dJ} \int_0^{2\pi} F \left(\frac{J}{\omega} \cos \phi \right) \frac{d\phi}{2\pi} \right) \right] \quad (3.15)$$

whose exponent is $(-2\pi i m)$ times the familiar first-order amplitude-dependent beam-beam-shifted tune [9]. First order Hamiltonian nonlinear resonance occurs at amplitudes J for which (3.15) is equal to unity for some nonzero m .

The final result of all these manipulations--namely (3.11), supplemented by (3.9) and (3.14)--has these significant virtues:

-If one neglects the $O(\xi^2)$ term in (3.14) (one can show²² that this is not inconsistent with assuming $\gamma T = O(\xi^2)$ in (3.10) and 3.11)), then an argument based on (3.6) shows (modulo concerns about making approximations before infinite sums are evaluated) that the operator $(1-G^{mm})^{-1}$ reduces simply to multiplication by an m -dependent function of J .

-With this simplification, one can show that (3.11) reduces to

$$0 = \frac{d}{dJ} \left\{ \gamma J + \left[\gamma \sigma^2 \omega J + \xi^2 \sum_m^2 |f^m|^2 \left\{ \frac{1}{2} + \operatorname{Re} \left[e^{-2\pi i m v} (1-G^{mm})^{-1} \right] \right\} \right] \frac{d}{dJ} \right\} p^0 \quad (3.16)$$

where

$$f^m(J) \equiv \int_0^{2\pi} e^{-im\phi} f\left(\frac{2J}{\omega} \cos\phi\right) \frac{d\phi}{2\pi}, \quad (3.17)$$

and where we have used (3.9), and the fact that G^{mm} transforms into its complex conjugate when m changes sign. From (3.16), we see that, within this scheme, beam-beam collisions can be thought of as renormalizing noise in an amplitude-dependent way. More importantly, (3.16) enables us quickly to find a nonsingular solution for p^0 , by ignoring the outside d/dJ , and then trivially integrating the first order ordinary differential equation that remains.

Finally, we note that a formalism that focuses on p^0 is in principle ideally suited to calculations that use--as did Kheifets--emittance as a stand-in for the square of the vertical coordinate. In particular, (2.9) in the present context becomes

$$\begin{aligned} \frac{1}{2} \int (x^2 + v^2/\omega^2) P(x, v) dx dv &= \frac{1}{\omega} \int_0^\infty J dJ \int_0^{2\pi} P(J, \phi) d\phi \\ &= \frac{2}{\omega} \pi \int_0^\infty p^0(J) J dJ, \end{aligned} \quad (3.18)$$

where no approximations have been made.

Together with these virtues, Ruggiero's formalism has the following shortcomings:

-It is difficult to see how systematically to compute corrections to Ruggiero's approximate calculation of p^0 , because the perturbations are treated in very different ways in different parts of equations (3.5)/(3.16): Explicit expressions for G^{m0} , G^{0m} , and $(1-G^{00})$ are obtained by conventional expansion in powers of ξ and γ , while an explicit expression for $(1-G^{mm})^{-1}$ is obtained by power expansion of the exponents in a particular factorization of G^{mm} .

-It is not clear that the relative orders of magnitude of ξ and γT are treated realistically in this formalism. In approximating $(1-G^{00})$, the perturbation expansion is truncated at $O(\xi^2)$ and $O(\gamma)$; in demonstrating that $(1-G^{mm})^{-1}$ is a multiplication operator, $O(\gamma)$ is retained while $O(\xi^2)$ is discarded. Experimentally, γT generally lies between .001 and .0001, while ξ can be as large as .06, so that ξ^2 and ξ^3 can be as large as about .004 and .0002 respectively. Thus, one might want to regard $O(\xi^3)$ and $O(\gamma T)$ as roughly equivalent. This would still be consistent with (3.6).

-Finally, the infinite sum that defines the multiplicative approximation to $(1-G^{mm})^{-1}$ is somewhat cumbersome. For this reason Ruggiero attempts, within his idealized one-dimensional model, to draw some quantitative conclusions about the beam-beam instability threshold directly from the form of (3.16), without actually generating an explicit graph of p^0 vs. J . We conclude this section by sketching his reasoning, and offering a few comments.

Ruggiero begins by supposing that the onset of beam blowup corresponds to the appearance of a range of J in which p^0 becomes substantially flatter than the otherwise rapidly decreasing unperturbed distribution $\exp-J/\sigma^2\omega$. According to (3.16), p^0 can deviate from $\exp-J/\sigma^2\omega$ only when the quantity in curved brackets in the sum is substantially different from zero for some nonzero m , and this happens only in some small region, of computable width, about a value of J for which (3.15) is equal to unity. Since the effect of one such "resonance" acting in isolation is presumably weak, Ruggiero asserts that a necessary condition for this flattening is substantial resonance overlap. Accordingly, Ruggiero (1) chooses J_R , a reference value of J ; (2) computes the value of m that corresponds roughly to the resonances overlapping (according to his definition) at J_R ; (3) approximates the sum in (3.16) by the term corresponding to that value of m ; and, finally, (4) requires that the nonsingular solution of (3.16), so approximated, be nearly flat at J_R . This yields a threshold condition for blowup in the form $\xi > \xi_{th}(J_R, T, \sigma, \omega)$.

I have the following misgivings about this type of construction:

-The notions of the reference action J_R , and of flattening, appear to be ad hoc, and without phenomenological basis.

-Flattening per se does not ensure that a beam distribution will empirically look "blown up," unless the flattened region contains a significant fraction of the total beam population. Ruggiero's threshold calculation does not include an estimate of the beam population in the flattened region.

-The notion of resonance overlap requires several terms simultaneously to contribute significantly to the sum in (3.5), and this is not obviously consistent with the one-dominant- $p^{n \neq 0}$ assumption on which the renormalized Fokker-Planck equation is based.

-Finally, Ruggiero's rush to implement resonance overlap seems premature in the light of computer simulations ([17],[18]) that suggest that isolated nonlinear resonance can play a significant role in beam blowup in circular electron-positron colliders. (Note, however, that Ruggiero's resonances need not be completely equivalent to nonlinear resonance. See below.)

In any event, his tuneshift threshold turns out, numerically, to be much higher than anything measured experimentally, as is characteristic of one dimensional models, even in nonstatistical frameworks [8].

IV. MY FORMALISM

The most significant difference between my approach and Ruggiero's is illustrated in Figures 3 and 4. Figure 3 shows unperturbed (harmonic oscillator) orbits in the phase space of the one-dimensional model of the preceding section. These orbits follow curves of constant J . Figure 4 shows, schematically, Hamiltonian orbits for the same model, but with nonzero ξ included. A particle moves along one of the curves shown, while the entire topographical pattern of curves rotates, with shear, about the origin, repeating itself every period T .

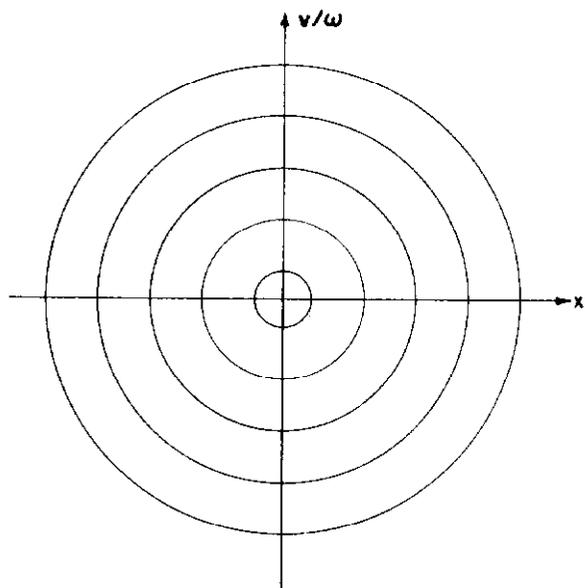


Figure 3

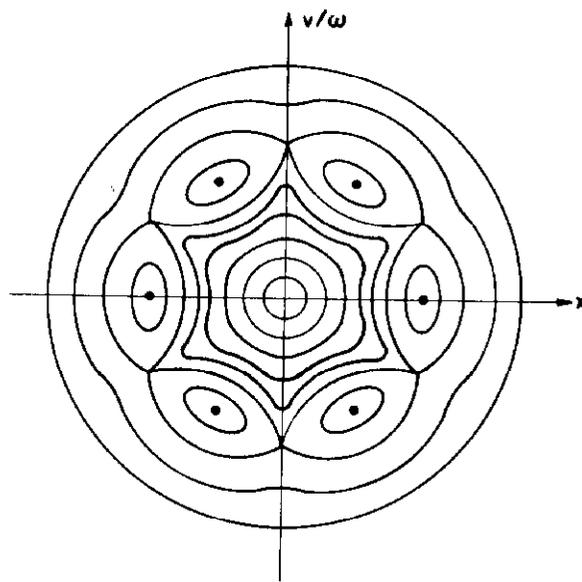


Figure 4

The curves that do not encircle the origin, forming an "island chain," correspond to nonlinear resonance. In Ruggiero's formalism, the presumed dominant contribution to the equilibrium beam distribution is constant on the curves shown in Figure 3; in my formalism, the leading approximation to the equilibrium distribution is constant on the curves shown in Figure 4.

These two points of view are nearly equivalent for the parts of phase space that lie near the origin or beyond the island chain. Near the island chain they are evidently quite different. In the first-order theory of nonlinear resonance [9],

the radius of such a chain corresponds roughly to a J such that (3.15) is equal to one, for some nonzero integer m.

Physically, the beam distribution in my scheme is approximately constant on the orbital curves of Figure 4 (or their higher-dimensional analogues) because, in the small- γ limit, collections of particles are smeared along by the Hamiltonian flow more rapidly than they are transported across Hamiltonian orbits by damping or noise. (This can also be argued in a more formal way.)

This kind of approximation has several significant virtues:

-It provides a systematic way of reducing the problem of finding an equilibrium beam distribution function of 2d phase space variables (plus time) to a corresponding problem involving d variables (plus time).

-It is not difficult to implement explicitly, because the equations that describe orbital curves (or their higher-dimensional analogues, KAM tori) often can be approximated in simple ways [9]. For example, in the case of Figure 4, the curves close to the origin or well beyond the island chain are given approximately by $J=\text{constant}$, as when $\xi=0$; while close to the resonant island chain they are given by

$$a (J-J_r)^2 + b \cos[\delta + 6(\phi-wt)] \cong \text{constant}, \tag{4.1}$$

where the parameters a,b, J_r , δ , and w are readily computable.

-It satisfies a relatively simple partial differential equation, which is derived by first directly substituting the ansatz $\{P(\vec{x},\vec{v},t)=\text{constant on KAM tori}\}$ into the Fokker-Planck equation, and then, in the result, replacing certain coefficients by suitably defined averages over oscillatory dependence on time and on (perturbed) canonical angles.

For example, in the case of the one-dimensional model discussed in the preceding section, this "thermally averaged" Fokker-Planck equation takes the form

$$0 = \left\{ -\frac{\partial}{\partial t} + \gamma \frac{\partial}{\partial I} \left[\int_0^{2\pi} v \frac{\partial I}{\partial v} \frac{d\theta}{2\pi} + \omega_{\sigma}^2 \int_0^{2\pi} \left(\frac{\partial I}{\partial v} \right)^2 \frac{d\theta}{2\pi} \frac{\partial}{\partial I} \right] \right\} P(I, t) \quad , \quad (4.2)$$

where θ and I are perturbed canonical angle and action, with $\xi \neq 0$ structure fully included. The integrals in (4.2) are to be regarded as functions of θ, I , and t . Modulo "thermal resonance" problems (see below) this is the first term in a systematic expansion in powers of γ . Notice that (4.2) is a continuous-time equation, not a single-time eigenvalue condition as in (2.4). Therefore, in determining equilibria, equation (4.2) must be supplemented with some sort of time-periodicity constraint. The higher-dimensional version of (4.2) is no more complicated.

By contrast with the expressions derived by Kheifets and by Ruggiero, neither equation (4.2) nor its solutions exhibit small-denominator singularities in the limit of vanishing damping and noise. These have been replaced by gentler topological discontinuities in the $d\theta$ integrals at the island-chain boundaries, because a $d\theta$ integral is equivalent to a line integral along an orbit curve, and at a chain boundary an orbit curve that encloses the origin abruptly breaks into several segments, each forming only half of an orbit curve that encloses an island center.

At present, this scheme faces the following difficulties:

-We cannot formulate the thermally averaged Fokker-Planck equation where the Hamiltonian dynamics is chaotic, i.e. where KAM surfaces do not exist. This is not necessarily fatal, because previously cited computer results do suggest that at least some observed colliding beam behavior is associated with nonchaotic single resonances. Moreover, even if chaotic effects did in fact turn out to dominate colliding beam physics, we might still hope to use this formalism to understand

their onset, since, according to Chirikov [19], chaos can develop out of the overlap of otherwise nonchaotic isolated resonances.

-Even though it involves fewer phase space variables than its exact counterpart, the thermally averaged Fokker-Planck equation is at present still too difficult to solve for more than one degree of freedom. Efforts to identify limits in which the equation is separable, or possesses some helpful symmetry, have so far not succeeded. The reader is invited to try his or her own hand at this game. A useful place to start might be the particularly interesting limit in which resonance streaming ([17]; see also [9], sections III. 2.d, IV. 3.b, and V.3) is conspicuous.

-The averaging procedure that leads to (4.2) and its higher-dimensional counterparts involves evaluating limits of the form

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau dt e^{i[n\Omega + \vec{n} \cdot \vec{\omega}]t} \quad , \quad (4.3)$$

where $\Omega = 2\pi/T$, n and \vec{n} are integral, and the components of $\vec{\omega}$ are the canonical angular frequencies of the fully perturbed canonical angles $\vec{\theta}$ that parametrize some KAM surface. In deriving the thermally averaged Fokker-Planck equation, it is assumed that averages of the form (4.3) vanish identically, unless $n = \vec{n} = 0$. However, strictly speaking, this is true only if $n\Omega + \vec{n} \cdot \vec{\omega}$ is nonzero for all n and \vec{n} ; whereas in fact there is in general a countable but dense set of KAM surfaces on which this is false.

The vanishing of $n\Omega + \vec{n} \cdot \vec{\omega}$ for some n and \vec{n} , in this statistical context, is called "thermal resonance." It is distinguished from nonlinear resonance in that $\vec{\omega}$ is a multiplet of fully perturbed frequencies. In nonlinear resonance, on the other hand, an approximate $\vec{\omega}$ is substituted into $n\Omega + \vec{n} \cdot \vec{\omega} = 0$ in order to locate the center of what is to become a resonant island. In principle, thermal resonance can give rise to structure that is superimposed on the familiar island structure of nonlinear

resonance. Are these two types of resonance mixed together in the small denominators in Kheifets' and/or Ruggiero's formalism(s)?

In arguing for thermal averaging, I am tacitly hoping that thermal resonance can be ignored because it only takes place on a set of measure zero in phase space. Is this fair? I have explored three different ways of answering this question:

(1) I have adapted methods from the theory of the Feynman path integral, reasoning that it may be easier to study a path, that can encounter thermal resonances sequentially, one at a time, than to study an entire distribution, which could suffer from many resonances in many different locations simultaneously [6]. One can go quite far in developing such path techniques, but as yet not far enough to yield a clear statement about thermal resonance. (2) I have studied the statistical behavior of linear systems near linear resonance [7], which have some formal features in common with nonlinear systems in the presence of thermal resonance. My results suggest that thermal resonance for small γ can give rise to pronounced but very narrow structures in phase space.

(3) I have numerically studied the system

$$\begin{aligned} \dot{x} &= + pQ(x^2+p^2) \\ \dot{p} + \gamma(t)p - \lambda(t)\epsilon(t) &= - x Q(x^2+p^2) \end{aligned} \quad , \quad (4.4)$$

where the function Q is not constant [16]. This system faces thermal resonance problems in equilibrium whenever $\gamma(t)/\lambda^2(t)$ is not independent of time. (4.4) is especially suited to numerical work because it can be solved explicitly when γ and λ vanish. The solution is simply

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = \begin{pmatrix} \cos\Delta t Q(x_0^2 + p_0^2) & +\sin\Delta t Q(x_0^2 + p_0^2) \\ -\sin\Delta t Q(x_0^2 + p_0^2) & \cos\Delta t Q(x_0^2 + p_0^2) \end{pmatrix} \begin{pmatrix} x_0 \\ p_0 \end{pmatrix} \quad (4.5)$$

In practice, I represent damping and noise by lumped kicks separated by time Δt ; and between kicks I have x and p evolve according to the map (4.5). Thermal averaging leads one to expect that the equilibrium distribution for (4.4), for small γ and λ , is given approximately (up to normalization) by

$$\exp - \frac{\langle \gamma \rangle}{\langle \lambda^2 \rangle} (x^2 + p^2) \quad , \quad (4.6)$$

where $\langle \rangle$ indicates time average. My data has not yet shown significant departures from this Gaussian.

V. CONCLUDING REMARKS: ANALYTICAL THEORY vs. COMPUTER SIMULATION

We must conclude that computer simulation at present compares very favorably with other available theoretical approaches to electron-positron storage ring analysis: It is not restricted to one or two degrees of freedom; it is not restricted to weak/strong systems; and it is not restricted only to perturbative limits. However, at present computer simulation has two significant weaknesses:

-We do not yet know, in an algorithmic way, how best to choose the operating conditions that a numerical study should preferentially simulate. I.e., computer simulation is not yet systematic in an optimized way.

-Computer simulation has so far been ill-suited to the calculation of loss rates or beam lifetimes. Numerical models typically simulate several hundred "macroparticles" for a few damping times; thus they can simulate loss rates as low as about one part in 10^2 per damping time. However, at CESR, for example, lifetimes of about a hundred minutes, equivalent to loss rates of about one part in 10^7 per

damping time, can be fatal (because of background in the CLEO detector)[20].

Personally, I feel that it is too much to hope that intuitive, back-of-the-envelope or hand-calculator theory will ever completely supplant large scale computer simulation of electron-positron circulating colliders. At the same time, it is perfectly reasonable to hope that further theoretical research will teach us how to make simulations better and faster, and how to learn much more from them.

High energy physics has so much to gain from an increased understanding of collider behavior. Let us hope that more scientists give the subject serious and creative attention.

ACKNOWLEDGEMENTS

I am grateful to Sam Kheifets and Francesco Ruggiero for helpful conversations.

REFERENCES

1. S. Kheifets, Particle Accelerators 15 (1984), 153.
2. S. Kheifets, in Proceedings, International Conference on High Energy Accelerators, Fermilab, 1983, pp. 303-305.
3. S. Kheifets, R. Helm, and H. Shoaee, in Proceedings, International Conference on High Energy Accelerators, Fermilab, 1983, pp. 306-308.
4. F. Ruggiero, Ann. Phys. (N.Y.) 153 (1984), 122.
5. J.F. Schonfeld, Ann. Phys. (N.Y.) 160 (1985), 149.
6. J.F. Schonfeld, Ann. Phys. (N.Y.) 160 (1985), 194.
7. J.F. Schonfeld, Ann. Phys. (N.Y.) 160 (1985), 241.
8. J.F. Schonfeld, in "Physics of High Energy Particle Accelerators (Fermilab Summer School, 1981)"(R.A. Carrigan, Jr., F.R. Huson, and M. Month, Eds.), AIP Conference Proceedings No. 87, 1982, pp. 314-344.
9. J.F. Schonfeld, in "Physics of High Energy Particle Accelerators (SLAC Summer School, 1982)"(M. Month, Ed.), AIP Conference Proceedings No. 105, 1983, pp. 524-650.

10. A.W. Chao, in "Physics of High Energy Particle Accelerators (BNL/SUNY Summer School, 1983)" (M. Month, P.F. Dahl, and M. Dienes, Eds.), AIP Conference Proceedings No. 127, pp. 201-242; and in these Proceedings.
11. F. Vivaldi, Rev. Mod. Phys. 56 (1984), 737.
12. J.T. Seeman, in Proceedings, International Conference on High Energy Accelerators, Fermilab, 1983, pp.212-216; and in these Proceedings.
13. A.W. Chao and M. Month, Nucl. Instr. and Meth. 121 (1974), 129; M. Month, in Proceedings, International Conference on High Energy Accelerators, SLAC, 1974, pp. 402-404; M. Month, IEEE Trans. Nucl. Sci. NS-22 (1975), 1376.
14. S. Peggs, CESR Report CBN 84-12.
15. M. Sands, in "Physics with Intersecting Storage Rings" (B. Touschek, Ed.), pp. 257-411, Academic Press, New York, 1971.
16. J.F. Schonfeld, unpublished.
17. J.L. Tennyson, in "Physics of High-Energy Particle Accelerators (Fermilab Summer School, 1981)" (R.A. Carrigan, Jr., F.R. Huson, and M. Month, Eds.), AIP Conference Proceedings No. 87, 1982, pp. 345-394.
18. A. Piwinski, IEEE Trans. Nucl. Sci. NS-30 (1983), 2378.
19. B.V. Chirikov, Phys. Rep. 52 (1970), 263.
20. CESR Operations Group, private communication, November, 1983.
21. J. Gareyte, in these Proceedings.
22. F. Ruggiero, private communications, February-March, 1985.