Dynamic aperture calculations for circular accelerators and storage rings

David R. Douglas

Citation: AIP Conference Proceedings **153**, 390 (1987); doi: 10.1063/1.36369
View online: http://dx.doi.org/10.1063/1.36369
View Table of Contents: http://aip.scitation.org/toc/apc/153/1
Published by the American Institute of Physics

Dynamic Aperture Calculations for Circular Accelerators and Storage Rings

David R. Douglas
Continuous Electron Beam Accelerator Facility
12070 Jefferson Avenue
Newport News, Virginia 23606

Table of Contents

0094-243X/87/1530390-84$3.00

Dynamic Aperture Calculations for Circular Accelerators and Storage Rings*

David R. Douglas
Continuous Electron Beam Accelerator Facility
12070 Jefferson Avenue
Newport News, Virginia 23606

I. Objectives of This Document

It is the author's hope that this document will help provide the novice with both an understanding of what dynamic aperture is and some insight as to why it is limited. In addition, we hope that this paper will provide information as to the analytical and numerical methods available for the study of dynamic aperture. Finally, we wish to present some interpretations and applications of these concepts and tools by means of examples. Those factors having an impact on the dynamic aperture of circular machines and storage rings will be discussed; applications to two types of machine of current popular interest (namely, large hadron colliders and synchrotron radiation electron storage rings) will be made in order to illustrate the ideas that are presented.

II. Single Particle Motion in an Accelerator

A. Coordinates, Phase Space, and Hamilton's Equations

The motion of a charged particle in an accelerator is most easily described by reference to an "ideal" or "design" trajectory. The design trajectory is the closed orbit comprised of the straight lines and circular arcs, along which a particle would move at precisely the design energy in a perfect machine. Such an idealised particle would arrive at the rf stations at exactly the proper phase of acceleration or bunching and is therefore referred to as the "synchronous particle". A physical particle ("real" particle in a "real" machine, with errors) will not move precisely on the design trajectory, nor will it have precisely the design energy. Instead, it will oscillate about the design trajectory. The oscillations will make the length of the real trajectory different than the length of the ideal trajectory and the energy differences will make the real velocity different than the ideal velocity; as a consequence the physical particle will arrive at the rf stations somewhat out of phase with the accelerating fields, leading to oscillations in the energy and phase of the particle.

---

*Work done at the Lawrence Berkeley Laboratory, 1 Cyclotron Rd., Berkeley, CA. 94720

Even under circumstances in which the oscillating motion of a physical particle is unstable, the differences between the behavior of the physical particle and the idealised synchronous particle are (on some scale) relatively small. We therefore describe the motion of the physical particle in terms of deviations from the motion of the synchronous particle. A suitable set of coordinates is depicted in Figure 1. The x coordinate measures tranverse deviations of the physical particle location from the design trajectory in the nominal plane of the accelerator, the y coordinate measures transverse deviations perpendicularly out of the nominal plane of the accelerator. Together, the x and y axes define a plane which is always perpendicular to the design trajectory. The z coordinate measures the distance of this plane (along the reference trajectory) from a chosen reference point at the particular instant of interest.

The evolution of the motion of a physical particle as a function of time is completely specified by the six functions

$$x(t), p_x(t), y(t), p_y(t), z(t), p_z(t).$$

The functions p_x, p_y, and p_z are the "canonical momenta conjugate, respectively, to x, y, and z". The vector $[x(t), p_x(t), y(t), p_y(t), z(t), p_z(t)]$ may be viewed as describing a path in a six-dimensional space called "phase space". Given the coordinates and momentum components of a particle at a single instant of time, this path in phase space is completely and uniquely determined by the equations of motion. If we neglect radiation (a fairly good approximation for protons, even at 20 TeV, and a reasonable approximation for electrons, for the time scales that will be of interest) these equations of motion are simply Hamilton's equations:

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial H}{\partial p_x} & \frac{dp_x}{dt} &= -\frac{\partial H}{\partial x} \\ \frac{dy}{dt} &= \frac{\partial H}{\partial p_y} & \frac{dp_y}{dt} &= -\frac{\partial H}{\partial y} \\ \frac{dz}{dt} &= \frac{\partial H}{\partial p_z} & \frac{dp_z}{dt} &= -\frac{\partial H}{\partial z}. \end{aligned} \tag{1}$$

For our purposes, the Hamiltonian may be written in the following manner:

$$H = H(x, p_x, y, p_y, z, p_z) = \sqrt{m_0^2 c^4 + c^2(\mathbf{p} - q\mathbf{A})^2} + q\psi\,. \tag{2}$$

In this expression, $\mathbf{A} = \mathbf{A}(x, y, z, t)$ is the magnetic vector potential and $\psi = \psi(x, y, z, t)$ is the electric scalar potential.

So far, we have chosen to parameterise the calculation of the trajectory of a physical particle in terms of the time. This parameterisation is somewhat inconvenient in that it does not allow a simple comparison between the physical and the synchronous particles. For example, our present choice of coordinates

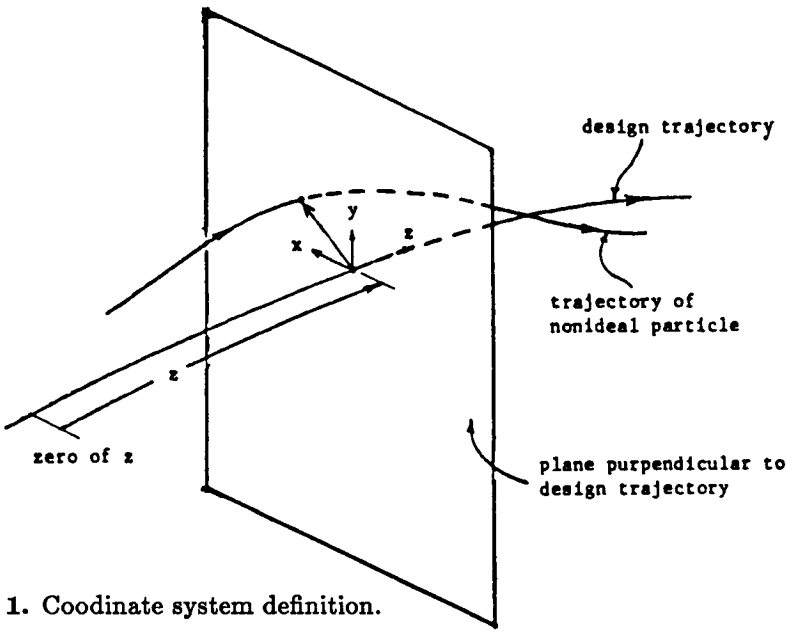

Figure 1. Coodinate system definition.

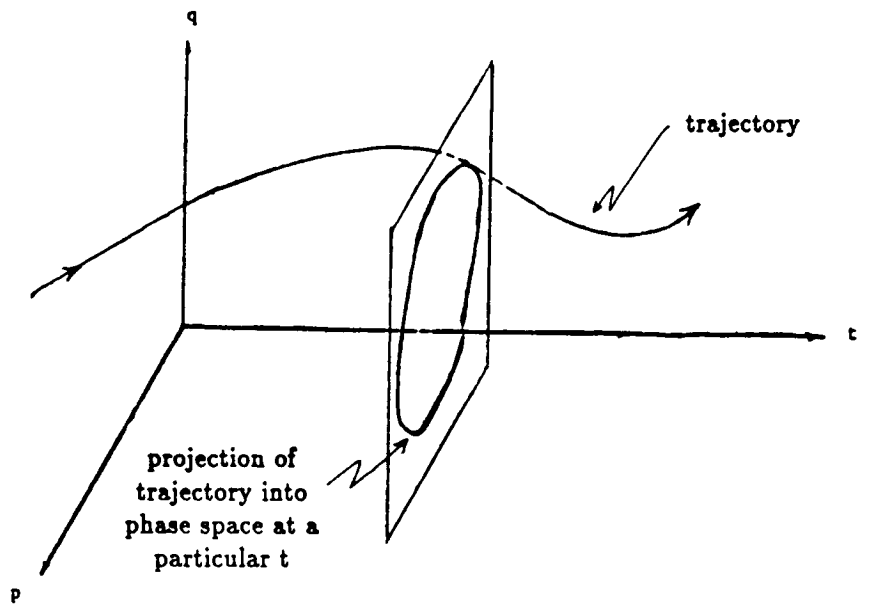

Figure 2. Schematic representation of state space.

will allow us to say that at 2:00 p.m. on 7 July 1985 a proton is 135 metres upstream of rf cavity # 7, but it does not allow a ready determination of whether or not this proton will be in phase with the acclerating field in cavity # 7. We are more interested in the time at which the particle arrives at a particular location in the ring (such as an rf station) than we are in the location of the particle at a specified time. We therefore will make a change in the parameterisation of the problem. Time t will be taken to be an independent variable (rather than the parameter) and the distance z along the design trajectory will be taken to be the parameter for the equations of motion. That we can do this, and that the procedure results in a set of equations which are Hamiltonian, is insured by the following theorem.

Theorem 1[1] Let $H(\mathbf{w};t)$ be the Hamiltonian for a system with n degrees of freedom $[\mathbf{w} = (q_1,...,q_n,p_1,...,p_n)]$. Suppose $q_1 = \partial H/\partial p_1 \neq 0$ in some region of phase space for some time interval of interest. Then within this region, for the time interval under consideration, q_1 can be introduced as the independent variable replacing time t. Moreover, the equations of motion with q_1 as independent variable may be obtained by using $K = -p_1$ as an effective Hamiltonian.

We remark that p_z must always be nonzero if the accelerator is to work. Identifying the q_1 of Theorem 1 with z, we then have that the motion of a particle in an accelerator is completely defined by the phase space vector $[x(z),p_x(z),y(z),p_y(z),t(z),p_t(z)]$ the components of this vector are solutions of the following set of canonical equations:

$$\begin{aligned} \frac{dx}{dz} &= \frac{\partial K}{\partial p_x} & \frac{dp_x}{dz} &= -\frac{\partial K}{\partial x} \\ \frac{dy}{dz} &= \frac{\partial K}{\partial p_y} & \frac{dp_y}{dz} &= -\frac{\partial K}{\partial y} \\ \frac{dt}{dz} &= \frac{\partial K}{\partial p_t} & \frac{dp_t}{dz} &= -\frac{\partial K}{\partial t} \end{aligned} \quad . \tag{3}$$

In these equations, the momentum p_t conjugate to t is simply the negative of the "old" Hamiltonian $H(x,p_x,y,p_y,z,p_z;t)$ given by equation (2):

$$p_t = -H = -\sqrt{m_0^2c^4 + c^2(\mathbf{p} - q\mathbf{A})^2} + q\psi \ .$$

It is therefore useful to think of p_t as the negative of the energy of the physical particle. The system (3) is generated by the following "effective Hamiltonian".

$$\begin{aligned} K &= K(x,p_x,y,p_y,t,p_t;z) \\ &= -p_z \\ &= -\frac{1}{c}\sqrt{(p_t + q\psi)^2 - m_0^2c^4 - c^2[(p_x - qA_x)^2 - (p_y - qA_y)^2]} - qA_z \ . \end{aligned} \tag{4}$$

The path-length parameterised form (3) of Hamilton's equations is most convenient for comparing the motion of a physical particle to that of a synchronous particle. Because of the particularly simple nature of the design trajectory, it is possible to compute the time $t(z)$ at which the synchronous particle arrives at any point along the design trajectory. A canonical transformation may then be performed to "remove" this time from the problem; the resulting set of equations then provide not only the transverse position of a physical particle relative to the design trajectory, but also give the deviation in energy and arrival time of a physical particle from the energy and arrival time of the sychronous particle. This canonical transformation can be thought of as a move to a coordinate frame coincident with the synchronous particle; all the coordinates of a physical particle will then be measured relative to the synchronous particle and will therefore be, on some scale, "small". It will then be possible to treat all terms ("longitudinal" - *i.e.*, energy - terms as well as transverse positions) in Hamilton's equations on an equal footing. The problem then reduces to studying a six-dimensional phase space in which the phase space trajectories are in the vicinity of the origin, a problem which is amenable to treatment using any of the tremendous selection of tools available from a number of Hamiltonian perturbation theories.

Let us continue by actually carrying out the transformation to coordinates moving with the synchronous particle. We wish to make another canonical transformation, from x, p_x, y, p_y, t, p_t to variables x, p_x, y, p_y, T, P_T, with T and P_T defined by the following equations:

$$\begin{aligned} T &= t - t^0(z) \\ P_T &= p_t - p_t^0(z) \ . \end{aligned} \tag{5}$$

Here, $t^0(z)$ is the time at which a synchronous particle arrives at displacement z along the design trajectory; $p_t^0(z)$ is the energy variable of a synchronous particle at location z (and is just $-H$ evaluated on the design trajectory at location z - think of it as $-1 \times$ (design energy). That equations (5) define a canonical transformation is insured by the fact that they may be derived from a generating function $F(t, P_T)$ via the following equations:

$$\begin{aligned} T &= \frac{\partial F}{\partial P_T} \\ p_t &= \frac{\partial F}{\partial t} \end{aligned} \tag{6}$$

and by using as F the function

$$F(t, P_T) = [\, t - t^0(z) \,]\,[\, P_T + p_t^0(z) \,] \ . \tag{6'}$$

When a Hamiltonian system is subjected to a transformation of the form (5), the result is also a Hamiltonian system. The Hamiltonian generating this

new system is directly related to both the generating function and the Hamiltonian generating the system in terms of the old variables. The relationship is given by the following equation[2]:

$$K^{NEW}(x, p_x, y, p_y, T, P_T; z) = K(x, p_x, y, p_y, t, p_t; z) + \frac{\partial F}{\partial z} \,. \qquad (7)$$

As noted above, this canonical transformation is simply a translation into a coordinate frame moving with the design particle. The $\partial F/\partial z$ term will in idealised, steady-state systems (no errors or acceleration) render K^{NEW} independent of any terms which are linear in the dynamical variables x, p_x, y, p_y, T, P_T. That is, K^{NEW} will be at least quadratic in deviations from the synchronous particle values in time and energy, as well as in transverse positions and momenta. It is worth noting that both the new effective Hamiltonian K^{NEW} and the old effective Hamiltonian K are explicit functions of the parameter z. This is, of course, because the magnetic vector potential $\mathbf{A}$ and the electric scalar potential ψ change in form as a particle traverses the different types of elements which comprise the accelerator. Thus, in this full six-dimensional treatment of particle motion, the "longitudinal" momentum component p_z $(-K)$ changes as a particle circulates in the machine. Similarly, the deviation of the p_z of a physical particle from that of the synchronous particle (this deviation being exactly $-K^{NEW}$) varies as the particle circulates. Hence, neither K nor K^{NEW} provide us with an integral of the motion, which we might use to understand the particle behavior.

B. Transfer Maps (Canonical Transformations)

The previous section concerned itself with the selection of the coordinates with which we shall describe the motion of a particle in an accelerator. In the preceding discussion, a number of canonical transformations were employed to bring the equations of motion to a fairly simple form and to help us select a particularly useful and revealing set of coordinates from which to view our problem. All of these transformations were "passive"; that is, they had only to do with the selection of coordinates and were not directly concerned with the dynamical behavior of the particle motion. In this section, we will be concerned with another type of canonical transformation. This transformation is "active" in that it describes the actual evolution of a Hamiltonian system and is generated by the Hamiltonian function itself. This Hamiltonian-generated canonical transformation describing the evolution of the particle motion in the accelerator may be thought of (in a manner described below) as a mathematical mapping, and will be refered to as a "transfer map".

In the variables of interest, Hamilton's equations read as follows:

$$\begin{aligned} \frac{dx}{dz} &= \frac{\partial K^{NEW}}{\partial p_x} & \frac{dp_x}{dz} &= -\frac{\partial K^{NEW}}{\partial x} \\ \frac{dy}{dz} &= \frac{\partial K^{NEW}}{\partial p_y} & \frac{dp_y}{dz} &= -\frac{\partial K^{NEW}}{\partial y} \\ \frac{dt}{dz} &= \frac{\partial K^{NEW}}{\partial p_t} & \frac{dp_t}{dz} &= -\frac{\partial K^{NEW}}{\partial t} \end{aligned} \quad . \tag{8}$$

The effective Hamiltonian K^{NEW} is given by equation (8). These "canonical" equations of motion may be considered a differential-equation based description for the phase space trajectory of a charged particle. However, constraining the motion of a particle to be Hamiltonian imposes an extremely rich mathematical structure on the equations governing that motion. Because of this, one is not limited to descriptions of Hamiltonian systems which are based soley on differential equations. In particular, there are pictures in which it is appropriate to use the methods of differential geometry or group theory. We turn our attention to one such alternative viewpoint which is especially fruitful when dealing with dynamic aperture problems.

Consider a charged particle moving in an accelerator as a dynamical system with three degrees of freedom. This dynamical system is governed by the canonical equations of motion (8). It is useful to consider the 7-dimensional space consisting of all points $(x, p_x, y, p_y, T, P_T; z)$; in the following we shall refer to this as the "state space" of the system. For brevity, we denote coordinates collectively as $\mathbf{q} = (x, y, T)$ and the momentum as $\mathbf{p} = (p_x, p_y, P_T)$, and make the definition $\mathbf{w} = (x, p_x, y, p_y, T, P_T)$. Points in the state space may then be denoted as $(\mathbf{w}; z)$.

We may schematically represent state space as in Figure 2. Any point $(\mathbf{w}; z)$ in this space represents a particular configuration of the dynamical system described by the Hamiltonian K^{NEW}. Consider the evolution in "time" z of the system starting at the inital point $(\mathbf{w}^1; z^1)$ [where $\mathbf{w}^1 \equiv \mathbf{w}(z^1)$]. As z increases from z^1 to z^2, the "system point" $[\mathbf{w}(z); z]$ moves from $(\mathbf{w}^1; z^1)$ to the point $(\mathbf{w}^2; z^2)$ along a trajectory which is governed by the canonical equations. The point $(\mathbf{w}^2; z^2)$ at which the system resides at "time" z^2 is uniquely determined by the intial point $(\mathbf{w}^1; z^1)$ and the equations (8). As the system evolves from z^1 to z^2, for a given initial point, the system point therefore traces out a unique trajectory in state space (Figure 2). The set of all such trajectories in state space generated by the equations of motion starting at all possible initial points is referred to as a "Hamiltonian flow".

Hamiltonian flows, as they are generated by the canonical equations, are subject to various constraints. One may therefore make a variety of statements about the behavior that such a flow must exhibit. A familiar example of such a statement is Liouville's theorem; other examples include statements about Poincaré integral invariants and straightening-out theorems[3].

A second property of Hamiltonian flows is a consequence of the uniqueness of the solutions to the equations of motion and the fact that one may integrate Hamilton's equations either "forward" or "backward" in time. Because of this, it is possible to consider any point $(\mathbf{w}^2; z^2)$ in state space to be the image of a unique point $(\mathbf{w}^1; z^1)$ (at which the system resided at some earlier "time" z^1) under a nonlinear mapping $\mathcal{M}$. The uniqueness of solutions requires that any point in state space have only one trajectory passing through it. We may therefore take $(\mathbf{w}^2; z^2)$ to be the point of interest, and integrate "backwards" along the trajectory through it until we arrive at "time" z^1 and the point $(\mathbf{w}^1; z^1)$. Uniqueness then insures that this is the only point in the phase space at "time" $z = z^1$ which gives rise to a trajectory passing through $(\mathbf{w}^2; z^2)$. It is therefore always possible to treat the evolution of $(\mathbf{w}^1; z^1)$ into $(\mathbf{w}^2; z^2)$ as a mapping from z^1 to z^2 generated by the equations of motion. The image of any point under the mapping is defined in the natural fashion: allow the point of interest to serve as the initial condition for a solution of the equations of motion. Integrate the equations forward in "time" until z^2 is reached; the image of $\mathbf{w}^1$ under $\mathcal{M}$ is then taken to be $\mathbf{w}^2$. We write

$$\mathbf{w}(z^2) = \mathcal{M}\mathbf{w}(z^1) \tag{9}$$

to denote that $\mathbf{w}^2$ is the image of $\mathbf{w}^1$ under the nonlinear map $\mathcal{M}$.

Because the map $\mathcal{M}$ is generated by following a Hamiltonian flow, it must have special properties. Specifically, it must be a "symplectic map".

Definition 1 Let M be a 6×6 matrix, and let J be the 6×6 matrix with the following block form.

$$J = \begin{pmatrix} \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & & \begin{matrix} & \\ & \end{matrix} \\ & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & \begin{matrix} & \\ & \end{matrix} \\ \begin{matrix} & \\ 0 & \end{matrix} & & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \end{pmatrix}. \tag{10}$$

If M obeys the "symplectic condition"

$$M J \widetilde{M} = J \tag{11}$$

("$\sim$" denotes "transpose") then M is said to be a "symplectic matrix".

Definition 2 Let $\mathcal{M}$ be a (nonlinear) transformation acting on dynamical variables $\mathbf{w} = (w_1, w_2, w_3, w_4, w_5, w_6) = (x, p_x, y, p_y, T, P_T)$, mapping them to $\bar{\mathbf{w}} = (\bar{w}_1, \bar{w}_2, \bar{w}_3, \bar{w}_4, \bar{w}_5, \bar{w}_6) = (\bar{x}, \bar{p}_x, \bar{y}, \bar{p}_y, \bar{T}, \bar{P}_T)$:

$$\bar{\mathbf{w}} = \mathcal{M}\mathbf{w} .$$

Note that $\bar{\mathbf{w}} = \bar{\mathbf{w}}(\mathbf{w}; z)$. If the Jacobian matrix M of $\mathcal{M}$, defined by

$$M_{ij} \equiv \frac{\partial \bar{w}_i}{\partial w_j}$$

is symplectic for all $\mathbf{w}$, then $\mathcal{M}$ is said to be a "symplectic mapping".

The fact that any two points along a trajectory are related by a symplectic map can provide useful knowledge about the Hamiltonian flow. In fact, complete knowledge of $\mathcal{M}$ constitutes a solution to the canonical equations. For, given any initial point $(\mathbf{w}^1; z^1)$, we may use (9) to generate the state of the system $(\mathbf{w}^2; z^2)$ at any later "time" z^2. We thus arrive at an alternative formulation of Hamiltonian dynamics, one in which we seek the symplectic mapping generated by a Hamiltonian flow.

That such a formulation is always possible is insured by the following theorem, which tells us that every Hamiltonian flow generates a symplectic map.

Theorem 2[1] Every Hamiltonian flow generates a symplectic map, which is defined in the "natural fashion" of equation (9).

Given that a system is described by Hamilton's equations, one is therefore assured that the evolution of the system is governed by a symplectic map defined as in (9). If this map can be determined, a solution for the system (all dynamical variables specified as functions of "time" z and initial conditions) is obtained via equation (9). One is therefore motivated to inquire as to what, if any, information may be obtained about this mapping. This requires an understanding of what mathematical tools are appropriate to the study and representation of symplectic maps.

Consider a Hamiltonian system at three times $z^1 < z^2 < z^3$. Let $\mathcal{M}^{2\leftarrow 1}$ be the map governing the evolution of the system from time z^1 to time z^2, $\mathcal{M}^{3\leftarrow 2}$ be the map from time z^2 to time z^3 and $\mathcal{M}^{3\leftarrow 1}$ be the map over the entire interval, from time z^1 to time z^3 (see Figure 2). At z^3 the state of the system $\mathbf{w}^3$, in terms of the initial conditions $\mathbf{w}^1$, is given by the following expression:

$$\mathbf{w}^3 = \mathcal{M}^{3\leftarrow 1}\mathbf{w}^1 .$$

This transformation, however, is equally well described by the following pair of transformations, which take the initial condition first from z^1 to the system configuration $\mathbf{w}^2$ at z^2, and then onward to the final condition at z^3.

$$\mathbf{w}^2 = \mathcal{M}^{2\leftarrow 1}\mathbf{w}^1$$
$$\mathbf{w}^3 = \mathcal{M}^{3\leftarrow 2}\mathbf{w}^2 .$$

We may thus represent $\mathcal{M}^{3\leftarrow 1}$ as a composition of the two maps $\mathcal{M}^{2\leftarrow 1}$ and $\mathcal{M}^{3\leftarrow 2}$. As the equations of motion continuously generate the trajectory, and as all maps involved are symplectic, it is reasonable to treat $\mathcal{M}^{3\leftarrow 1}$ as the "product"

of $\mathcal{M}^{3\leftarrow 2}$ and $\mathcal{M}^{2\leftarrow 1}$. This product is defined by concatenation: act first with $\mathcal{M}^{2\leftarrow 1}$ and then with $\mathcal{M}^{3\leftarrow 2}$. All this discussion of products of symplectic maps hints at the topic of our next theorem.

Theorem 3[4] The set of symplectic maps generated by a Hamiltonian flow is a group.

Heuristic Proof We must prove that the set is closed under the product defined by concatenation, that an identity exists, that every element has an inverse, and that the product is associative.

Exercise 1 Prove that the product of two symplectic matrices is a symplectic matrix.

The product of two symplectic maps, defined by concatenation, is itself symplectic. Writing

$$\mathcal{M}^{3\leftarrow 1} = \mathcal{M}^{3\leftarrow 2}\mathcal{M}^{2\leftarrow 1}$$

and letting $M^{3\leftarrow 2}$ and $M^{2\leftarrow 1}$ be the Jacobians of $\mathcal{M}^{3\leftarrow 2}$ and $\mathcal{M}^{2\leftarrow 1}$, we see that the Jacobian $M^{3\leftarrow 1}$ of $\mathcal{M}^{3\leftarrow 1}$ is, by the chain rule, given by the following set of matrix elements (i and j run from 1 to 6):

$$\begin{aligned}(M^{3\leftarrow 1})_{ij} &= \frac{\partial w_i^3}{\partial w_j^1} \\ &= \sum_{k=1}^{6} \frac{\partial w_i^3}{\partial w_k^2}\frac{\partial w_k^2}{\partial w_j^1} \\ &= \sum_{k=1}^{6} (M^{3\leftarrow 2})_{ik}(M^{2\leftarrow 1})_{kj} \\ &= (M^{3\leftarrow 2}M^{2\leftarrow 1})_{ij} \ .\end{aligned}$$

As $\mathcal{M}^{2\leftarrow 1}$ and $\mathcal{M}^{3\leftarrow 2}$ are symplectic maps, the Jacobians $M^{2\leftarrow 1}$ and $M^{3\leftarrow 2}$ are symplectic matrices. However, by exercise 1 the product of these two matrices is also a symplectic matrix, whereby $M^{3\leftarrow 1}$ is a symplectic matrix. By definition 2, $\mathcal{M}^{3\leftarrow 1}$ is a symplectic mapping.

Other group properties follow as readily. For example, the fact that the equations of motion may be integrated backwards in time insures that the symplectic maps generated by Hamiltonian flows are invertible. the identity transformation is trivially symplectic ($IJ\tilde{I} = J$). Finally, the concatenation process clearly defines an associative product.

We therefore find that the set of symplectic maps generated by a Hamiltonian flow forms a group. By viewing a dynamical system in terms of a Hamiltonian flow, we can therefore shift our attention from a differential-equation based

description to one in which the fundamental mathematical tools are those of group theory. In this description, the problem of characterising the time (or "path-length parameterised") evolution of a Hamiltonian system is equivalent to determining the properties of the group of symplectic mappings generated by the Hamiltonian flow.

This rather mathematical vewpoint is readily connected to the properties of Hamiltonian systems with which physicists are well aquainted. Specifically, Hamiltonian systems have to do with "active" canonical transformations. The time evolution of a Hamiltonian system can be described in terms of a "canonical transformation generated by the Hamiltonian"[2]. The symplectic maps generated by Hamiltonian flows, described above, are in fact simply these very canonical transformations. This is the content of the next theorem.

Theorem 4 Canonical transformations are equivalent to symplectic mappings applied to canonical variables.

Before proving this assertion, we make a few definitions. They serve to refine the concept of canonical transformation.

Definition 3 Let $\mathbf{w} = (x, p_x, y, p_y, T, P_T)$ be the phase space of a Hamiltonian system with three degrees of freedom. Let $f = f(\mathbf{w})$ and $g = g(\mathbf{w})$ be any two functions on this phase space. The "Poisson bracket" (PB) of f with g is a function on phase space denoted by $[f, g]$ and defined by the following collection of derivatives:

$$[f,g] = \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial p_x} - \frac{\partial f}{\partial p_x}\frac{\partial g}{\partial x}\right) + \left(\frac{\partial f}{\partial y}\frac{\partial g}{\partial p_y} - \frac{\partial f}{\partial p_y}\frac{\partial g}{\partial y}\right) + \left(\frac{\partial f}{\partial T}\frac{\partial g}{\partial P_T} - \frac{\partial f}{\partial P_T}\frac{\partial g}{\partial T}\right).$$

The PB of the canonical phase space variables with themselves are particularly simple and important. These are the "fundamental" Poisson brackets:

$$[x, p_x] = [y, p_y] = [T, P_T] = 1; \quad \text{bracket of all other pairs } = 0.$$

Exercise 2 Show that $[f, g]$ may be written in the following fashion by using the antisymmetric matrix J:

$$[f,g] = \sum_{i,j=1}^{6} \left(\frac{\partial f}{\partial w_i} J_{ij} \frac{\partial g}{\partial w_j}\right)$$

where $\mathbf{w} = (w_1, w_2, w_3, w_4, w_5, w_6) = (x, p_x, y, p_y, T, P_T)$. Show that the fundamental PB are then compactly written as follows:

$$[w_i, w_j] = J_{ij} \ .$$

The concept of a "canonical" transformation was origionally invented to allow changes in coordinate systems without changing the form of the equations

of motion. Graduate mechanics texts illustrate how to cast these equations of motion in a compact form using PB's[2]. Thus, canonical transformations may be viewed as transformations that leave the "fundamental" Poisson brackets unchanged or invariant.

Definition 4 Let $\mathcal{M}$ be a transformation of canonical variables $\mathbf{w}$ to a new set $\bar{\mathbf{w}}$. This transformation is defined to be canonical provided that it preserves the fundamental Poisson brackets – that is, provided

$$[\bar{w}_i, \bar{w}_j] = [w_i, w_j] = J_{ij} \quad \text{where } \bar{\mathbf{w}} = \mathcal{M}\mathbf{w}.$$

We now may prove Theorem 4.

Proof of Theorem 4: Let $\mathcal{M}$ be a canonical transformation from variables $\mathbf{w}$ to variables $\bar{\mathbf{w}}$:

$$\bar{\mathbf{w}} = \mathcal{M}\mathbf{w}.$$

We claim $\mathcal{M}$ is a symplectic map. The Jacobian M of $\mathcal{M}$ is the following matrix of partial derivatives:

$$M_{ij} = \frac{\partial \bar{w}_i}{\partial w_j}.$$

Hence, using Exercise 2 and the assumption that the transformation is canonical:

$$\begin{aligned}(MJ\widetilde{M})_{ij} &= \sum_{k,l=1}^{6} \frac{\partial \bar{w}_i}{\partial w_k} J_{kl} \frac{\partial \bar{w}_j}{\partial w_l} \\ &= [\bar{w}_i, \bar{w}_j] \\ &= [w_i, w_j] \\ &= J_{ij}.\end{aligned}$$

The last equalities follow from the fact that $\mathcal{M}$ is a canonical transformation and therefore preserves the fundamental PB. Thus we find $MJ\widetilde{M}$=J, whence $\mathcal{M}$ is symplectic.

Now let $\mathcal{M}$ be a symplectic transformation

$$\bar{\mathbf{w}} = \mathcal{M}\mathbf{w}\ .$$

Denote by M the Jacobian of $\mathcal{M}$. We show that $\mathcal{M}$ preserves the fundamental PB and that it therefore is a canonical transformation:

$$\begin{aligned}[\bar{w}_i, \bar{w}_j] &= \sum_{k,l=1}^{6} \frac{\partial \bar{w}_i}{\partial w_k} J_{kl} \frac{\partial \bar{w}_j}{\partial w_l} \\ &= \sum_{k,l=1}^{6} M_{ik} J_{kl} (\widetilde{M})_{lj} \\ &= (MJ\widetilde{M})_{ij} \\ &= J_{ij}\end{aligned}$$

We therefore conclude that a symplectic mapping of canonical variables and a canonical transformation are one and the same thing. We remark that it is mathematically allowable to a apply a symplectic transformation to non-canonical variables; however, the result need not be physically meaningful. For example, noncanonical variables need not preserve the canonical structure of the canonical equations of motion. In physical applications, one should insure that the variables and equations being subjected to symplectic transformations are in fact canonical variables and equations.

Recalling our discussion of symplectic maps, we may now identify the "symplectic mapping generated by a Hamiltonian flow" with the classic "canonical transformation generated by the Hamiltonian which carries the system forward in time". If this symplectic map can be determined, the time evolution of the Hamiltonian system is then known. This viewpoint is particularly useful to accelerator physicists, because a circular accelerator or storage ring provides an extremely efficient analog computer for the study of such canonical transformations and symplectic maps. A circular machine may be viewed as an analog "black box", which evaluates the action of a map $\mathcal{M}$ on an initial condition. This is depicted in Figure 3. A particle is "injected" in the machine with the initial condition (x, p_x, y, p_y, T, P_T); after a single turn it comes back out of the machine with the "image under $\mathcal{M}$" of $(\bar{x}, \bar{p}_x, \bar{y}, \bar{p}_y, \bar{T}, \bar{P}_T)$. If we want to see where the particle goes on subsequent turns, we simply re-inject it, using the image from a previous turn as the initial condition for a subsequent turn. The black box (the accelerator) contains a canonical transformation (a symplectic map of canonical variables) which is controlled by the Hamiltonian describing the particle motion in the accelerator. The value of the phase space variables after each turn are analytic functions of the initial condition for that turn:

$$\begin{aligned}
\bar{x} &= \bar{x}(x, p_x, y, p_y, T, P_T) \\
\bar{p}_x &= \bar{p}_x(x, p_x, y, p_y, T, P_T) \\
&\vdots \\
\bar{P}_T &= \bar{P}_T(x, p_x, y, p_y, T, P_T).
\end{aligned}$$

These equations represent the active canonical transformation (they define the symplectic map) carrying the particle once around the ring.

A somewhat more physical point of view is depicted by Figure 4. This illustrates the actual situation that will occur when we discuss tracking calculations. A cut is made through the accelerator, normal to the design trajectory. We stand beside the machine and launch a particle into the machine with the initial condition (x, p_x, y, p_y, T, P_T). We then watch as it circulates around the machine (that is, as we repeatedly re-insert the particle into the black box) and record the phase space location of the particle every time it returns to the "surface of section" (our plane of observation). If we iteratively apply the transformation $\mathcal{M}$ many times, and observe the behavior of the point at which the particle orbit intersects the surface of section, it becomes possible to make

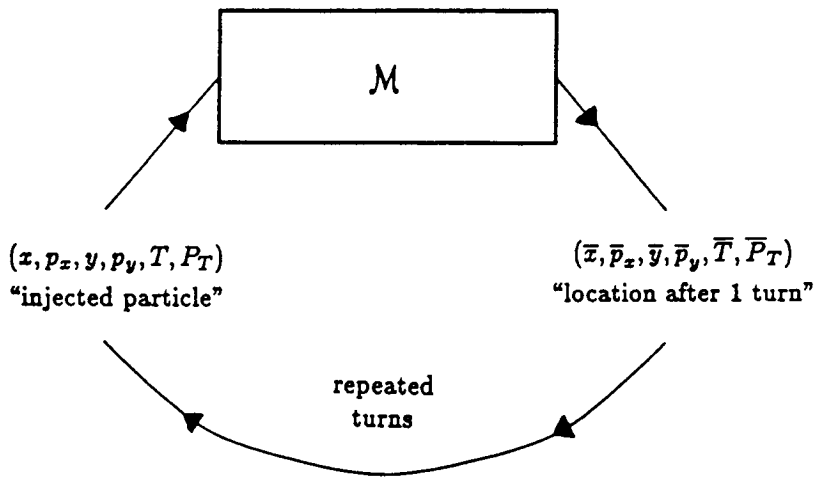

Figure 3. Analog "black box" view of accelerator.

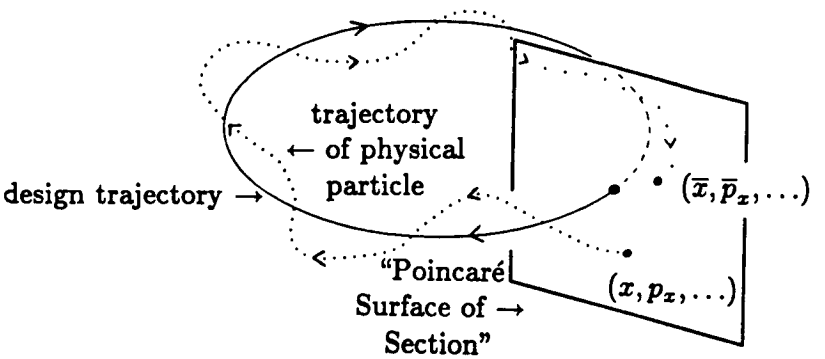

Figure 4. "Surface of Section" viewpoint of accelerator.

statements about the stability of that particular orbit. If, for example, the x coordinate of the point of intersection grows monotonically with every turn, we can pronounce that the orbit is unstable and that the particle will hit the wall of the beam pipe and be lost.

An advantage to thinking of the accelerator as an analog computer that evaluates a symplectic map is that we know many ways to mathematically model this dynamical system. It is therefore possible to predict how (or if) a particular accelerator will behave, in terms of the stability of particle orbits, without actually building the machine! In particular, the symplectic mapping that carries physical particles around the machine has a variety of representations, which may be numerically implemented. (The "group of symplectic mappings admits a variety of representations".) Each particular representation emphasises one or another aspect of the machine behavior. It is thus possible to study the question of the stability of particle orbits in great detail using a number of different numerical methods. In section IV.B we will compare a number of these algorithms to see under what circumstances each is most applicable. Before doing so, we will spend some effort on defining the problem of orbit stability more precisely.

III. The Concept of Dynamic Aperture

The concept of dynamic aperture arises because the transfer map $\mathcal{M}$ for a circular machine or storage ring is a nonlinear map. This necessarily implies that certain regions of the surface of section will not be "stable" (in the sense described in the previous section) under iterated transformations by the transfer map. This will be illustrated by the following example.

A. Example: "Linear" Ring With "Point" Sextupole and Octupole

In this example, we study a superficially simple 1-dimensional nonlinear system, which models certain types of motion in circular accelerators quite well. The phase space in this system is specified by the vector $\xi = (x, p_x)$, and the accelerator acting on this phase space will be taken to be a "perfect linear ring" containing one nonlinear element. We take the surface of section to lie adjacent to the nonlinear element (as shown in Figure 5) so that the transfer map of the ring, which carries points in the surface of section back into the surface of section, consists of a pair of canonical transformations $\mathcal{M}_1$ and $\mathcal{M}_2$. We denote action of this pair as follows.

$$\xi \xrightarrow{\mathcal{M}_1} \tilde{\xi} \xrightarrow{\mathcal{M}_2} \bar{\xi} .$$

The first transformation, $\mathcal{M}_1$, describes purely linear transport around the ring. We use a matrix representation to compute the effect of $\mathcal{M}_1$ on the initial phase space vector:

$$\begin{pmatrix} \tilde{x} \\ \tilde{p}_x \end{pmatrix} = \begin{pmatrix} \cos 2\pi\nu_x & \beta_x \sin 2\pi\nu_x \\ -(1/\beta_x)\sin 2\pi\nu_x & \cos 2\pi\nu_x \end{pmatrix} \begin{pmatrix} x \\ p_x \end{pmatrix} . \tag{10}$$

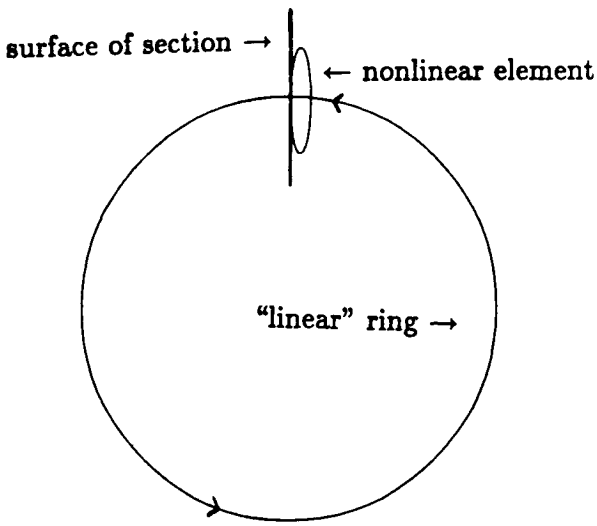

Figure 5. Schematic representation of "ideal" ring with "point nonlinear element".

The second transformation, $\mathcal{M}_2$, is a nonlinear transformation, which we shall represent using the following Taylor's series expansion of the image in terms of the initial condition:

$$\begin{cases} \bar{x} = \tilde{x} \\ \bar{p}_x = \tilde{p}_x + a_2\tilde{x}^2 + a_3\tilde{x}^3 \end{cases} . \tag{11}$$

Such representations of this type of transformation are often used to model the effect of a short, strong sextupole or octupole (or higher "multipole") on a particle traversing a beam line.

Exercise 3. A "multipole expansion" of the magnetic field in a beam-line element is often written in terms of the field measured in the plane $y = 0$. Specifically, one writes:

$$B_y = B_y(x) = B_2x^2 + B_3x^3 + \ldots .$$

In this expansion, B_2 is the magnetic field due to the sextupole component of the field in the beam-line element, B_3 is that due to the octupole component, and so forth. Using the Lorentz force law, show that if the above expansion is truncated at the "third-order" term B_3 and the equations of motion solved in the "impulsive" or "kick" approximation (element length $L \to 0$, element strength $B_n \to \infty$, but $LB_n \to$ finite constant), equation (11) will result. How are the the a_n in equation (11) related to the "integrated strength" parameters LB_n, which are finite, of the impulsive approximation?

From discussions at various summer schools[5] and/or the universally useful paper by Courant and Snyder[6] we realise that the map $\mathcal{M}_1$, which is represented by the matrix equation (10), will give stable betatron oscillations. That is, any initial condition ξ to which we apply $\mathcal{M}_1$ will remain bounded, for any number of applications of the map. Of interest to us is what happens when the "thin lens" nonlinear element is turned on: Will the motion remain stable or not??! To find out, we do an experiment! We launch particles into the ring and "track" them for many turns. We empirically determine if they are stable (*i.e.* if they "survive") for various settings of a_2 (sextupole strength) and a_3 (octupole strength).

In this experiment (for reasons that will become apparent later in this chapter) we have chosen $\beta_x = 10$ m and $\nu_x = 0.34$ to parameterise the linear transformation $\mathcal{M}_1$ given by equation (10). To do the experiment, first we choose a setting for a_2 and a_3. We then launch 10 particles, with initial displacements of 1 to 10 mm (with $p_x = 0$ in each case) from the surface of section, and iterate the transfer map [first act with $\mathcal{M}_1$ *via* equation (10) and then act with $\mathcal{M}_2$ via equation (11)] 500 times. After each application of the transfer map, we record the position of the 10 particles in the surface of section (the $x - p_x$ plane), and then use these positions as the initial conditions for the next iterate (see Figure 3 to be reminded of how this is done ...). After 500 iterates ("500 turns around the machine") we plot all the (x, p_x) pairs to see if (or which) particles are stable. If all particles are stable, we pick new values for the a_n and repeat the experiment. By going through this process several times, we can empirically determine a relationship between the values of the a_n and the regions of the surface of section which are stable under the transfer map. The results of our experiment are summarised in Figures 6a to 6i.

The inescapable conclusion of this exercise is that even a "simple" numerical example shows that the behavior of even a highly idealised model can get quite complicated. Not all of the single particle phase space is stable when nonlinear elements are present in the ring, and, moreover, even the stable regions can exhibit complex behavior such as "stochasticity". It is therefore necessary to understand how much of the single particle phase space is stable. If "not enough" of the phase space of an accelerator is stable, it might not be possible to inject beam into the machine or to perform the beam manipulations (*e.g.* orbit correction, bringing beams into collision, measuring β's or ν's,...) that are required for normal operation. The importance of this fact motivates us to give this stable region a special name:

Definition 5. The stable region of the single particle 6-dimensional phase space is called the "dynamic aperture" of the machine.

The importance of this definition is due to the fact that if a particle is put into the machine with an initial condition outside of the dynamic aperture (or if, during operation, the dynamic aperture changes so that it moves inside of the particle position), the trajectory of that particle will be unstable and the particle will be lost!

Exercise 4. A sufficiently strong sextupole can make the entire single particle phase space unstable under certain circumstances. Write a program to reproduce the calculations done in this section, and use it to conduct a different set of experiments. For given values of the tune ν_x near 1/3, find the maximum stable displacement as a function of sextupole strength a_2. What happens to this maximum stable displacement as ν_x gets very close to 1/3?

This is related to a common method employed to extract beams from circular accelerators. The method is referred to as "third-integer resonant extraction".[7]

Exercise 5. Repeat the "tracking" experiment of this section with $\beta_x = 10$ and $\nu_x = 0.41$ In what details do the plots you obtain differ from those shown here? You should find 5 "islands" at some amplitude, surrounded by a "stochastic layer". This effect is related to a "fifth-integer resonance".[8] (Hint: For $a_2 = 1$ and $a_3 = 0$, islands can be seen with the initial condition $x = 0.0075, p_x = 0.0$).

Figure 6. Results of tracking experiment; for all cases $\beta_0 = 10$ m, $\nu = 0.34$. The horizontal axis is in meters while the vertical is in radians.

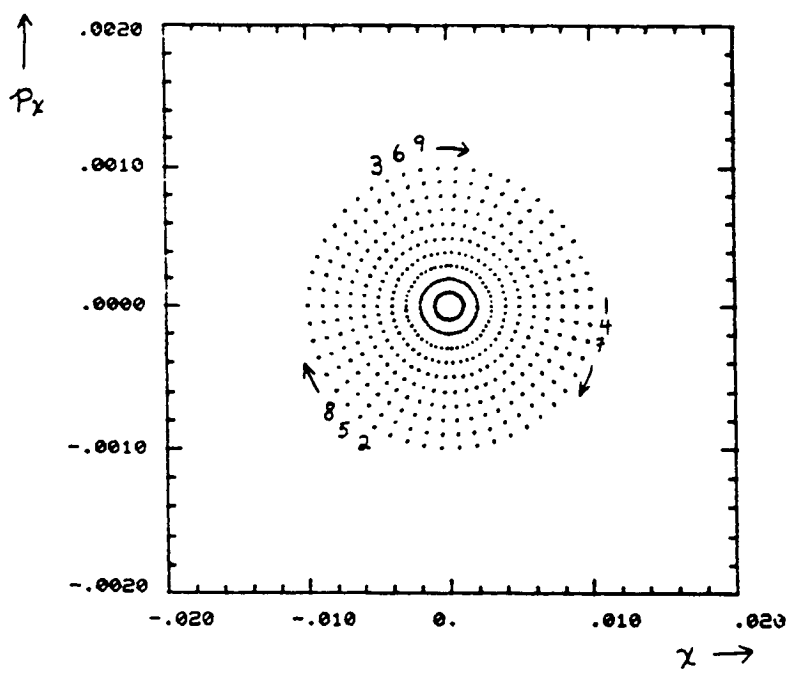

Figure 6a. Initial conditions x_0 =1 to 10 mm, p_x^0 =0; $a_2 = a_3 = 0$.

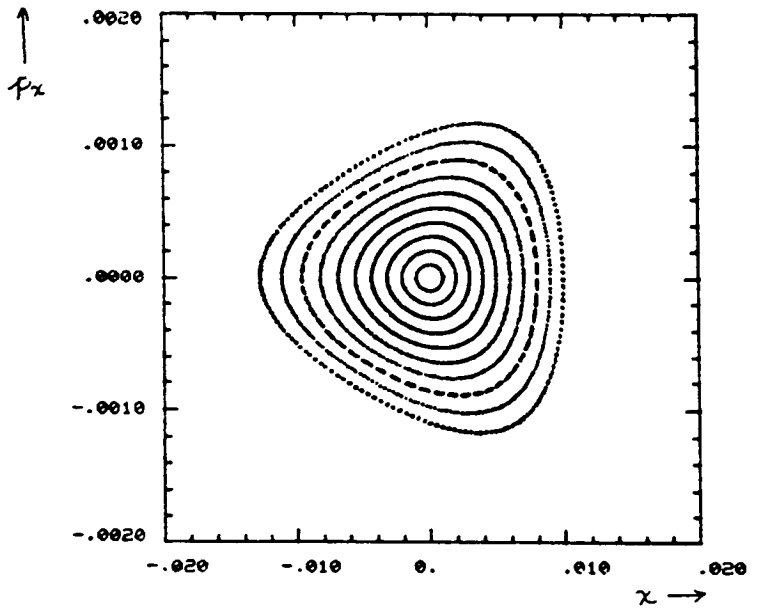

Figure 6b. Initial conditions x_0 =1 to 10 mm, p_x^0 =0; $a_2 = 0.5, a_3 = 0$.

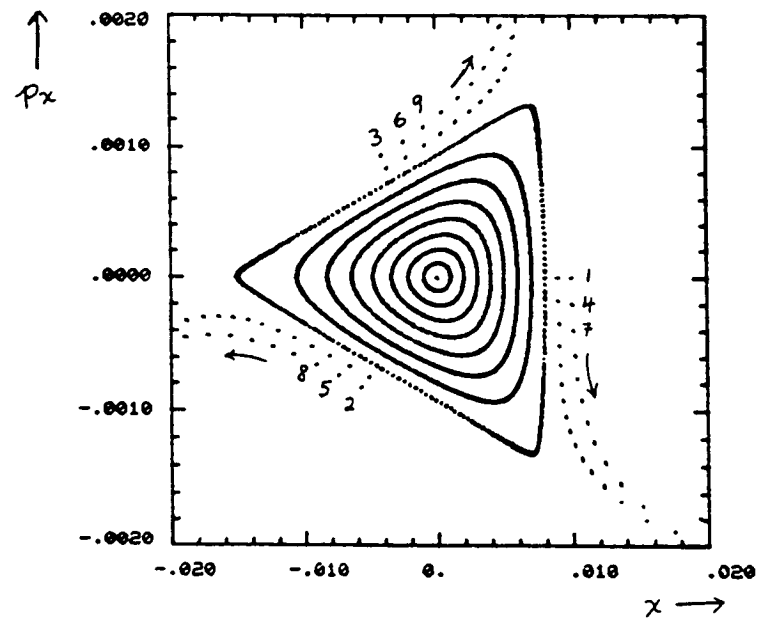

Figure 6c. Initial conditions x_0 =1 to 10 mm, p_x^0 =0; $a_2 = 1, a_3 = 0$. The maximum stable amplitude is about 8 mm.

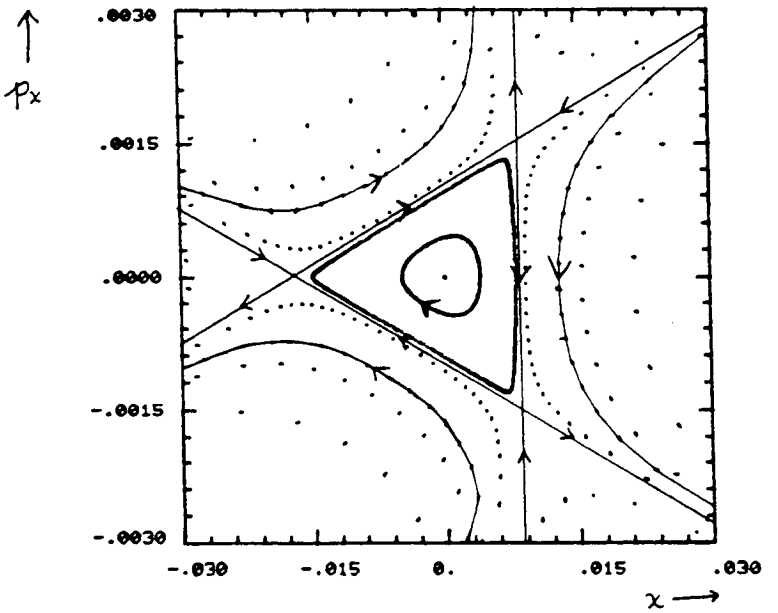

Figure 6d. Phase portrait illustrating separatrices and fixed points.

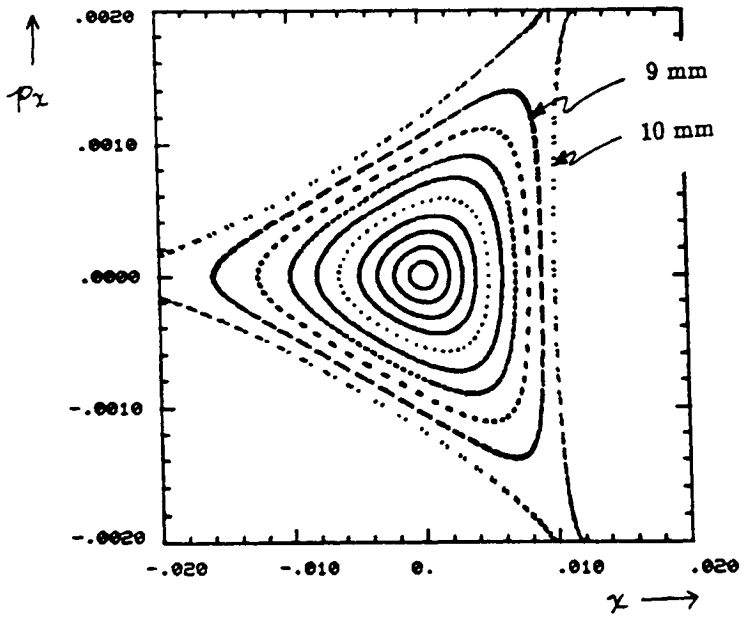

Figure 6e. Initial conditions x_0 =1 to 10 mm, p_x^0 =0; $a_2 = 1, a_3 = 10$. Octupole has been activated; orbits at 9 and 10 mm, which were unstable without octupole, are now stable.

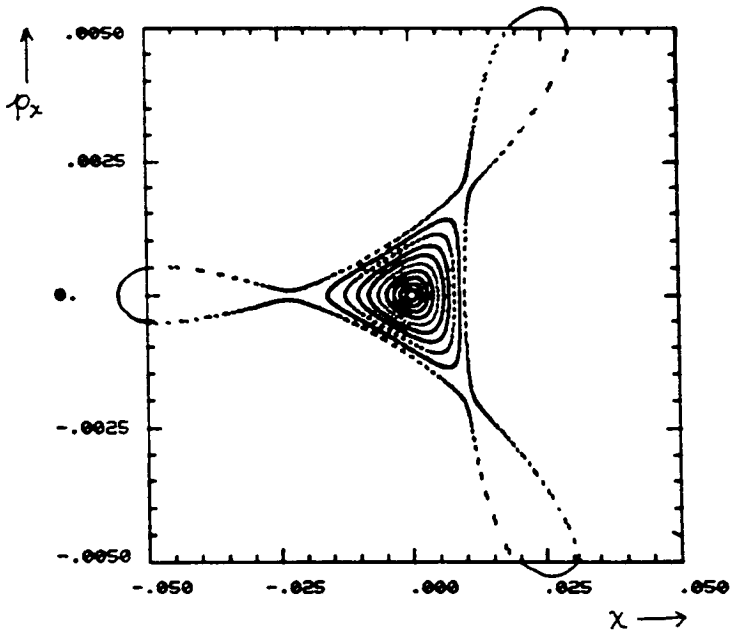

Figure 6f. Enlargement of Figure 6e illustrating "Mickey Mouse ears" due to octupole. The octupole driving term wraps the separatrices back around each other, to form the three-lobed figure.

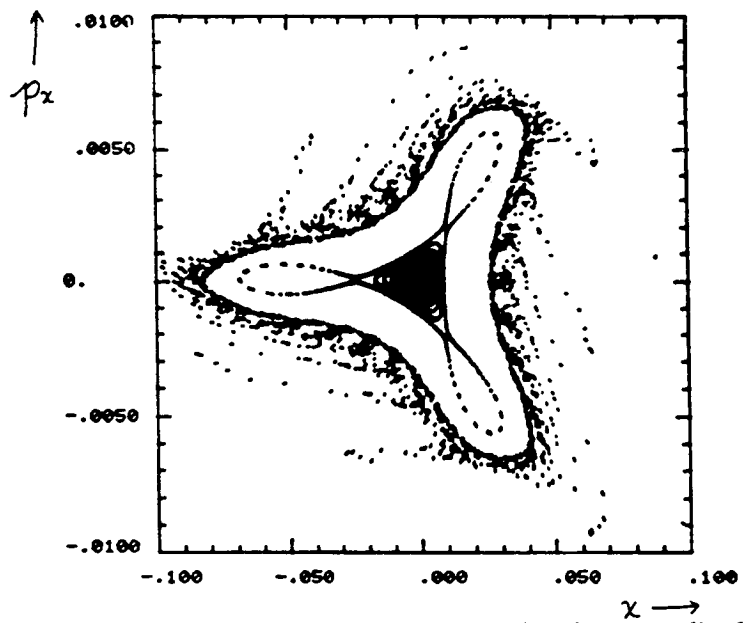

Figure 6g. Case of Figures 6e and 6f, including large amplitude particles. The maximum stable amplitude is about 28 mm.

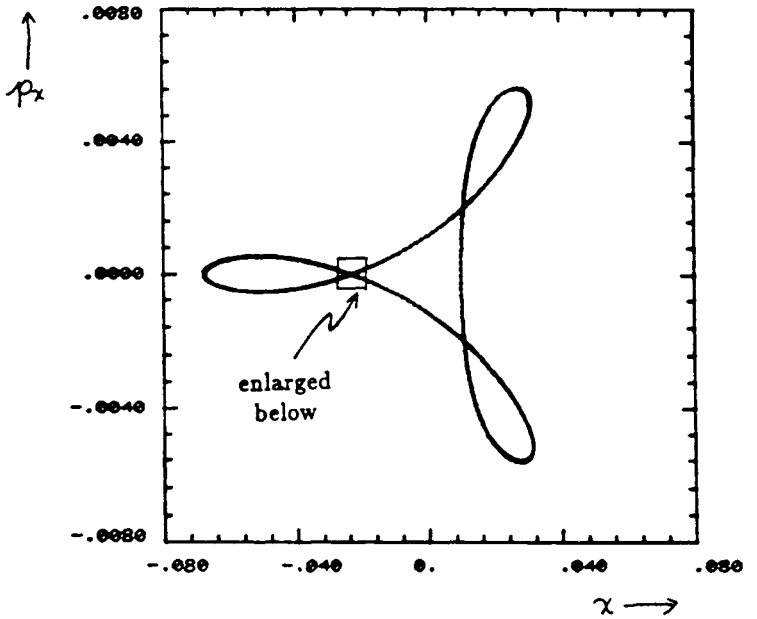

Figure 6h. Result of tracking $x_0 = -0.023276, p_x^0 = 0.0$, which is near a period three unstable fixed point, when $a_2 = 1, a_3 = 10$.

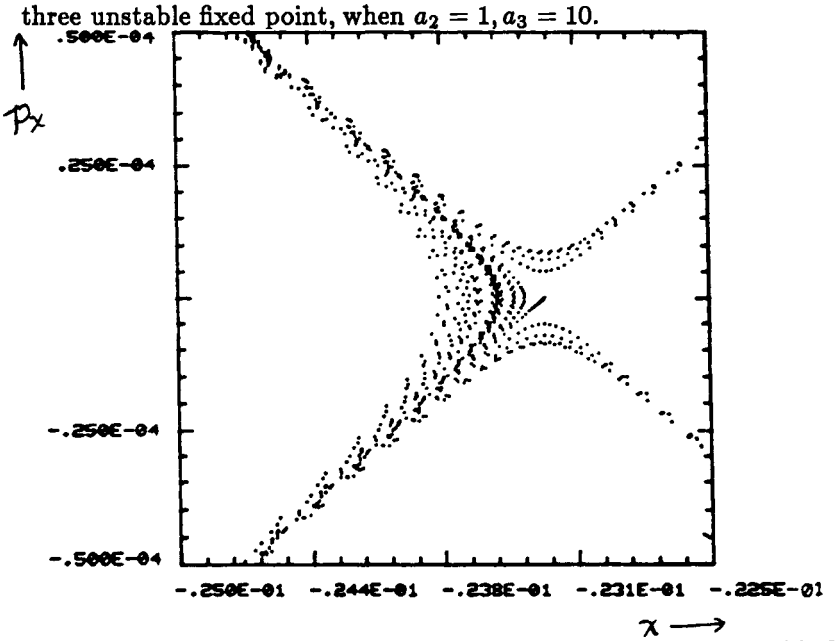

Figure 6i. Enlargement of Figure 6h in the vicinity of the unstable fixed point. This "stochastic layer" is due to the overlap of the separatrices generated by the octupole. The crossing of the separatrices forms "homoclinic" points, which are a signature of this type of chaotic motion.

B. Applications to Circular Accelerators

Before discussing specific applications of the concept of dynamic aperture, we must review one fundamental assumption we have made, namely, that the motion is Hamiltonian. The primary approximation implied by this assumption is the neglect of synchrotron radiation effects. This is quite reasonable for proton accelerators and storage rings, in which synchrotron radiation is, insofar as the single particle dynamics is concerned, completely negligiable. This is true even at 20 TeV energies, where the natural time scale for radiation effects, the damping time, is on the order of hours. We should inquire, however, as to the accuracy of this assumption when studying electron accelerators.

It is certainly true that radiation effects are non-negliable for electron machines. For the computation of dynamic aperture, however, the primary influence of synchrotron radiation is to provide damping to the single particle phase space, with a natural time scale (the damping time)[9]. Thus, when computing dynamic aperture for electron accelerators, radiation is viewed as a stablizing effect that introduces a time scale over which the motion is to be studied. The basic assumption we make is that if the motion is treated as Hamiltonian and is found to be stable for several damping times, then the inclusion of synchrotron radiation damping (as occurs in the physical accelerator) will only improve the situation. In particular, the primary effect of the radiation damping is assumed to be a stabilization of nonlinear effects that could, in a truely Hamiltonian system, lead to particle loss on time scales much longer than the damping time. If the motion, when modeled with Hamiltonian dynamics, is found to be stable for several damping times, then it is viewed as stable for all times. The concept of dynamic aperture presented above thus applies to the case of electron machines, with the additional feature that a particular time scale is implied.

As noted, determination of dynamic aperture is critical during the design phase of a machine. If the dynamic aperture proves to be too small, work must be done to make it larger. This is usually done by changing the design in order to improve the nonlinear "condition" that is causing the reduction of dynamic aperture. There are three types of machines for which this has been a critical process in recent years. The first type is that of dedicated synchrotron radiation sources[10]. Electron rings designed for this purpose are, in general, extremely strongly focussing (to make a "small" beam with correspondingly high intensity synchrotron radiation). Consequently, the sextupoles used in these rings are very strong, and lead to a rather small dynamic aperture[11]. Various schemes have therefore been devised to improve the dynamic aperture of the machines. These usually are based on either a reduction of focussing strength (by careful control of the electron beam parameters, to yield the desired small beam without strong focussing), or on a cancellation or correction of the nonlinear effects generated by strongly focussing sextupoles[12].

The second type of machine for which dynamic aperture is a major issue is e^+-e^- colliders. In such machines, the single-particle dynamic aperture is

almost always limited by chromaticity sextupoles. (As an aside, we note that multi-particle effects, such as the beam-beam interaction or collective effects, are often the actual limiting factor in the performance of such machines.) Often simple chromaticity correction schemes (using two families of sextupoles - one to correct horizontal chromaticity, the other the vertical chromaticity) provide "inadequate" dynamic aperture. More "sophisticated" correction schemes can often enlarge the available aperture[13].

A final type of accelerator for which dynamic aperture is an important issue is that of superconducting hadron colliders[14]. This is because superconducting magnet field quality is "poor" - that is, the fields have very strong nonlinear dependences (relative to conventional low-field magnets). Moreover, in most designs for these machines, the magnet aperture is small (so that the beam is "close" to the superconducting coils), and (because the machines are big) the β functions are large, so that the beam may, in some circumstances, execute large amplitude oscillations. All these effects generate strong systematic and random nonlinear terms, which occur throughout the machine. As a result, complicated nonlinear forces act on the particles.

The issue of adequate dynamic aperture may be the critical single issue for the Supeconducting Super Collider or other large hadron colliders. It is therefore vital to understand what the dynamic aperture of such machines will be with all "multipoles" (*i.e.* nonlinear errors) and how this dynamic aperture can be improved.

IV. Determination of Dynamic Aperture

In the example of section III, we numerically modeled a simple accelerator and computed its dynamic aperture by direct experimentation. We launched a particle and tracked its motion (by iterating the transfer map on the initial condition) to determine if the particle was stable. This "tracking" procedure is an important tool for determining the dynamic aperture of accelerators. However, as the simple example demonstrated, the results of such experiments can become quite complicated even in the most straightforward cases.

Before numerical modeling of a machine is attempted, it is therefore reasonable to try to make use of analytic tools, in order to elucidate the relevant features of the dynamical behavior of particles in the machine. Specifically, the extensive array of analytic tools that are available should be applied in an effort to ascertain what effects have the greatest impact on the dynamic aperture. After the relevant features and effects have been determined, it is possible to proceed with numerical studies, including models for important effects, to obtain more exact estimates of the dynamic aperture.

In this section, we will follow this philosophy, and first sketch some of the analytic methods that are available for estimation of dynamic aperture. We will then provide a somewhat more detailed discussion of the numerical tools

currently employed by the "nonlinear experimentalists" that study machine behavior.

A. Analytic Approaches to Dynamic Aperture

1. Scaling Arguments A very simple method of studying dynamic aperture is to employ scaling arguments to determine the dependence of the aperture on some parameter. To use this method, one must choose a "figure of merit" that must be controlled in order to operate the accelerator. An estimate of this figure of merit is made, and its dependence on the strength of various nonlinear terms in the machine design is determined. This will place limits on the strength of the nonlinear terms.

Example: Tune shift with momentum.[15] Superconducting magnets of the type used in large hadron colliders have a rich "harmonic content". That is, the magnetic field in the midplane has a strong positional dependence:

$$B_y(x) = B_0(1 + b_2x^2 + b_4x^4 + \ldots)$$

where B_0 is the dipole field, and the b_n are relatively (relative to conventional magnets) large. For example, for the Superconducting Super Collider (SSC) Reference Designs Study (RDS) magnet design C[16] the multipoles are as follows:

$$\begin{aligned} b_2 &= 0.540/\mathrm{m}^2 \\ b_4 &= 4.00\times10^4/\mathrm{m}^4 \\ b_6 &= 4.15\times10^8/\mathrm{m}^8 \\ b_8 &= 3.01\times10^{12}/\mathrm{m}^{12}\ . \end{aligned}$$

A particle moving through a dipole magnet off-axis will therefore experience some focusing from these multipole moments (beyond what it will obtain from the accelerator quadrupoles alone) and will, therefore, experience a tune shift. This tune shift is given by the following expression[9]:

$$\Delta\nu = \frac{1}{4\pi}\int_{ring} \beta\, k\, dz\ .$$

The position of an off axis particle is readily determined, if the particle is not at the design momentum. In this case, the displacement at a position z is simply $x(z,\delta) = \eta(z)\delta$, where η is the dispersion at z and δ the fractional momentum offset $(\Delta p/p_0)$. In the integral, $k = B'/B\rho$ is the "gradient error". Thus, for off momentum particles,

$$\begin{aligned} k = k(\delta) &= \frac{B'}{B\rho} \\ &= \frac{B_0}{B\rho}\left(2b_2x(\delta) + 4b_4[x(\delta)]^3 + \ldots\right)\ . \end{aligned}$$

If we view the integral as a discrete sum over bend magnets (which make up "most" of an accelerator such as the SSC), the tune shift may be written as follows:

$$\Delta\nu = \frac{1}{4\pi}\sum_{bends}\beta(\delta)k(\delta)L_{bend}\ .$$

For the SSC RDS lattices, the terms in this sum may be approximated as follows.

- $L_{bend} = 140$ m for design C magnets.
- Assume $\beta(\delta) = \beta(0)$ to get the estimate for $\Delta\nu$. (This estimate is true to within better than a factor of two for a well behaved lattice and the momentum range of interest.)
- Use $k(\delta)$ from above and view the sum as an average:

$$\sum_{bends}\beta k(\delta) = N_{bends}\overline{\beta\ k(\delta)}\ .$$

- $\overline{\beta} = 250$ m in design C.
- Using the above expressions

$$\begin{aligned}\overline{k(\delta)} &= \overline{\frac{B_0}{B\rho}(2b_2x(\delta) + 4b_4(x(\delta))^3 + \ldots)} \\ &= \frac{B_0}{B\rho}(2b_2\overline{x(\delta)} + 4b_4\overline{(x(\delta))^3} + \ldots) \\ &= \frac{B_0}{B\rho}(2b_2\overline{\eta}\delta + 4b_4\overline{\eta}^3\delta^3 + \ldots)\end{aligned}$$

where $\overline{x} = \overline{\eta}\delta$ has been used. For SSC RDS design C, $\overline{\eta} \sim 2.5$ m.

The tune shift therefore scales as follows:

$$\Delta\nu = \frac{N_{bends}L_{bends}B_0}{4\pi B\rho}\,\overline{\beta}\,(2b_2\overline{\eta}\delta + 4b_4\overline{\eta}^3\delta^3 + \ldots)\ .$$

With the parameters given above and using $B_0 = 3$ T, $B\rho = 66712.8$ T-m, and $N_{bend} = 1000$, we get the following result:

$$\Delta\nu = 125(5b_2\delta + 62.5b_4\delta^3 + \ldots)\ .$$

We can use this to set a limit, for example, on the allowable sextupole component b_2. Suppose that we must have $|\Delta\nu| < 0.01$ (to avoid resonances; see the following section for a discussion this effect) and need to allow for $|\delta| \leq 0.001$ (for beam handling, rf manipulations, *etc.*). We then need

$$\begin{aligned}0.01 > \Delta\nu &= 125 \times 5 \times b_2 \times 0.001 \\ &= 0.625\ b_2\end{aligned}$$

or, for the required "dynamic aperture" in momentum, we need to have

$$b_2 < \frac{0.01}{0.625} = 0.016 \ .$$

This is a factor of 30 times smaller than the RDS value given above. Based on this simple estimate, we may therefore state that the magnet design must be modified to reduce the sextupole component b_2

A more careful evaluation of the integral gives the tune shift as a detailed function of momentum. This is plotted in Figure 7. The resulting tune shift is quite large, indicating that either the multipoles must be reduced in magnitude, the lattice must be changed to lower $\overline{\beta}$ and $\overline{\eta}$ (this can be done by shortening the FODO cells comprising the lattice), or correction elements must be put into the ring to cancel the tune shift generated by the error multipoles in the dipoles.

The above example illustrates the type of qualitative information about dynamic aperture that can be provided by simple scaling arguements. More rigorous methods can provide more detailed information. Subsequent sub-sections will introduce some of these methods.

2. Resonance Analysis If the tune of a particle in a circular accelerator is rational (that is, the ratio of two integers), the motion of the particle can be in resonance with the (periodic) magnetic focussing forces (either linear or nonlinear) of the machine, and the particle may be lost. This is the same effect as that studied in elementary dynamics courses; the particle is executing harmonic motion, and a driving term acts on the particle with frequency near the natural frequency of the particle motion (which is the tune in this case). This driving term increases the amplitude of the particle motion - eventually leading to particle loss. (Recall that "eventually" can mean 10^6 periods of the motion, which is only a fraction of a second to a few seconds of real time.) Powerful tools for describing resonant motion have been developed, by both celestial mechanicians and accelerator physicists[17]. Many of the key features of a resonance analysis can be understood on the basis of Hamilton's equations alone.

In the case of a single isolated resonance (one in which the particle tune is near a rational number at which a single driving term in the equations of motion occurs), we can attack the problem of estimating the limits of particle stability as follows:

- Write down Hamilton's equations, with all nonlinear driving terms.
- Fourier-transform the driving terms.
- Identify the "resonant" driving term (the one with frequency near the natural frequency ("tune") and eliminate all the others.
- Make a series of canonical transformations on the equations of motion (with the resonant driving term only) to obtain an integral of the motion.

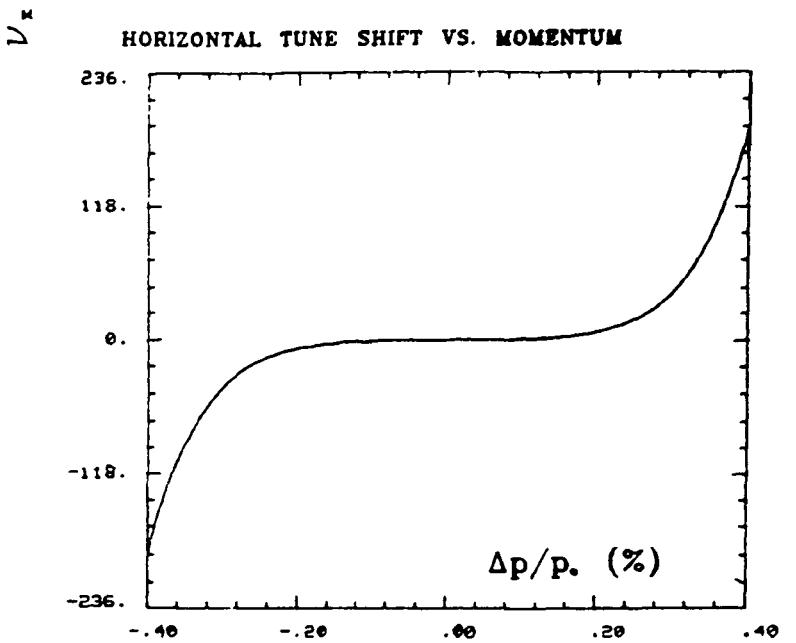

Figure 7a. Tune shift with momentum offset due to systematic errors in SSC RDS design C dipoles.

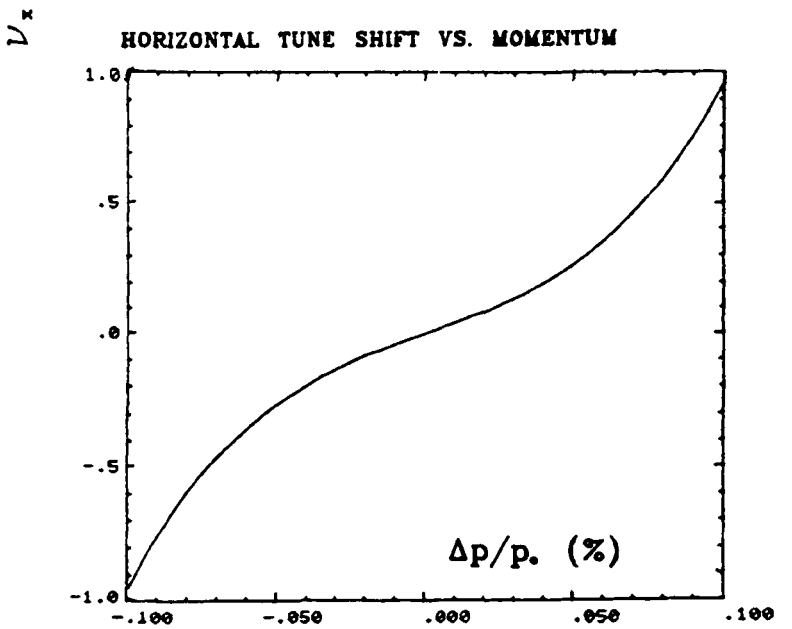

Figure 7b. Enlargement of Figure 7a. illustrating the momentum range of interest. $\Delta\nu_x$ is $\sim$ 100 times larger than acceptable.

- Use the integral of the motion to characterise the stability of various regions of the phase plane. That is, identify the dynamic aperture by studying the geometry in phase space of the surfaces on which the particle will move (the "integral surfaces").

We illustrate these ideas with an analysis of our numerical example of section III (a one-dimensional, isolated, resonance analysis).

Example Resonance Analysis of Linear Ring with Point Sextupole: In this system, the equation of motion is as follows:

$$\frac{d^2x}{dz^2} + K(z)x = -\sigma\delta(z)x^2 \ .$$

The $K(z)$ term details the linear focussing of the ring, while the delta-function term describes the effect of the impulse (which is proportional to the displacement squared and the strength σ of the sextupole) at $z = 0$.

We begin by making a "Floquet transformation" to new variables u and θ defined as follows:

$$u = \frac{x}{\sqrt{\beta}}$$
$$\theta = \int \frac{dz}{\nu\beta} \ .$$

Then, for example,

$$p = \frac{du}{d\theta} = \nu\left(\frac{\alpha x + \beta p_x}{\sqrt{\beta}}\right)$$

is the momentum conjugate to u, where $p_x = dx/dz$ is the momentum conjugate to x and α, β, and ν are the linear lattice parameters.

This transformation serves two purposes. First, it "normalises" the motion so that the elliptical linear phase space becomes circular; that is, the particle motion lies on a circle in the (u, p) phase plane, and, as time progresses, the particle's coordinates circulate on the circle. Secondly, this transformation makes θ the independent variable; thus, θ becomes the "time", and varies from 0 to 2π on each turn. Hence, we may think of θ as a "generalised azimuthal angle", giving the position of the particle around the ring.

With this transformation completed, the equation of motion becomes the following:

$$\frac{d^2u}{d\theta^2} + \nu^2 u = -\nu\beta^{3/2}\sigma\delta(\theta) \ .$$

We remark that the delta function is now a function of θ rather than of z. The Jacobian weight $dz/d\theta$ has thus been included; when making the transformation, the identity $\delta(z) = \delta(\theta)/\nu\beta$ is employed. The above equation of motion is canonical, in that it can be derived from the following Hamiltonian:

$$H(u, p, \theta) = \frac{1}{2}(p^2 + \nu^2 u^2) + \frac{1}{3}\nu\beta^{3/2}\sigma u^3\delta(\theta) \ .$$

We next make another canonical transformation; this one is to "polar coordinates" J and ϕ in phase space, or, as they are more commonly known, to action-angle variables. The transformation is defined by the following system:

$$u = \sqrt{\frac{2J}{\nu}} \cos\phi$$
$$p = -\sqrt{2J\nu} \sin\phi \quad .$$

Exercise 6. Find a generating function for the above canonical transformation.

The equations of motion remain Hamiltonian, and are now generated by

$$\mathcal{H}(\phi, J, \theta) = \nu J + \frac{\sigma}{3}\nu\beta^{3/2}\delta(\theta)\left(\frac{2J}{\nu}\right)^{3/2}\cos^3\phi \quad .$$

The action J corresponds to the radius of the circular phase space trajectory of the particle, while the angle ϕ corresponds to the angle between the radius vector of the particle position on this circle and the u axis in phase space. We note that (Hamilton's equations)

$$\frac{d\phi}{d\theta} = \frac{\partial \mathcal{H}}{\partial J}$$

so that in the linear ($\sigma = 0$) case,

$$\frac{d\phi}{d\theta} = \nu \ ,$$

implying

$$\phi = \nu\theta \quad .$$

That is, ϕ is "locked" to θ. If, for example, $\nu = 1/3$, ϕ goes from 0 to 2π in 3 turns. In the linear case, the phase space would then appear as in Figure 8a. If, however, $\nu > 1/3$, but still near $1/3$, the phase space would appear as shown in Figure 8b. We therefore interpret ϕ to be the number of oscillations that the particle undergoes between $\theta = 0$ and the θ at the current position around the accelerator. That is, ϕ may be regarded to be the "instantanious tune" or the "phase advance".

A dangerous situation arises if $\phi(2\pi)$ is rational. A particle can then return to the delta function impulsive driving force "in phase" with every kick the delta function provides, a situation that could make the particle motion unstable. In general, we must worry if $\nu = m/3$ for some integer m. To understand why, Fourier-transform the delta function

$$\delta(\theta) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik\theta} \ .$$

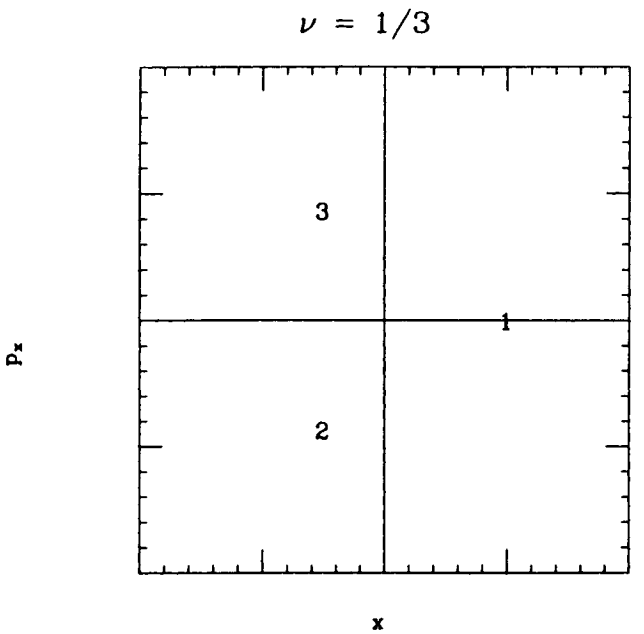

Figure 8a. Single particle phase space when $\nu = 1/3$. The particle returns to the starting point after three turns.

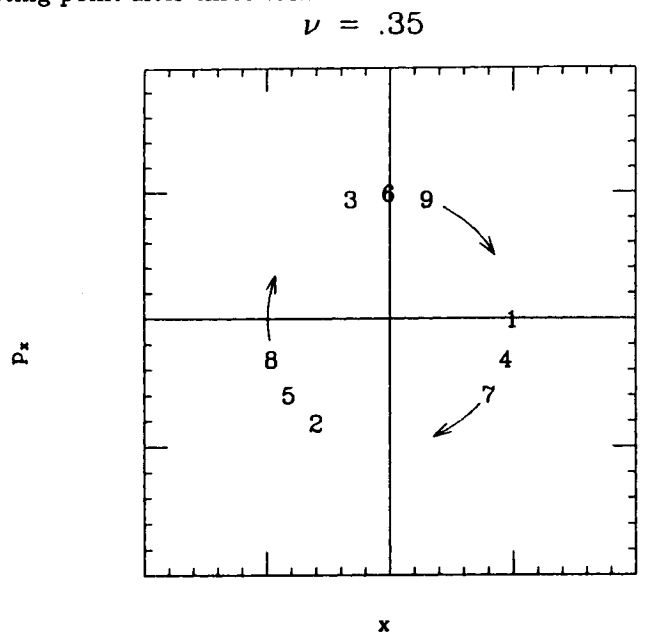

Figure 8b. Single particle phase space over nine turns when the tune is greater than, but near to, 1/3. The points of intersection trace out a circle.

Also, observe that

$$\cos^3\phi = \frac{1}{8}(e^{3i\phi} + 3e^{i\phi} + 3e^{-i\phi} + e^{-3i\phi}) \ .$$

If $\nu \approx m/3$, we then have that

$$\phi = \frac{m}{3}\theta + \epsilon \ ,$$

where ϵ is a numerically small, but complicated, function of the particle amplitude, azimuthal position, and sextupole strength. Therefore,

$$\delta(\theta)\cos^3\phi = \frac{1}{16\pi}(e^{i(3\phi - m\theta)} + e^{-i(3\phi - m\theta)}) + R$$

where R is a remainder term which is a function of sinusoids of $3\phi - l\theta$ and $\phi - l\theta$ for all other integers l.

The key feature of this calculation is that $3\phi - m\theta = \epsilon$ is small, so that $e^{\pm i(3\phi - m\theta)}$ is slowly varying as θ changes. This Fourier component therefore acts as a nearly-constant driving term. All other Fourier components vary rapidly: $3\phi - l\theta \neq 0$ for nearly all θ. These terms therefore fluctuate and do not drive the particle in a consistant manner; the effect of these terms thus averages to zero over many terms. We can therefore neglect them and keep only the "resonant term", yielding the following simplified Hamiltonian, in which β_o is the value of β at the site of the delta function force:

$$\mathcal{H}(\phi, J, \theta) = \nu J + \frac{\sigma}{3}\beta_0^{3/2}(\frac{2J}{\nu})^{3/2}\frac{1}{8\pi}\cos m(\frac{3}{m}\phi - \theta) \ .$$

To summarize the calculation so far, we note that we have taken $\nu = m/3$, so that $\phi = (m/3)\theta + \epsilon$, with ϵ "small". This circumstance has led to a slow variation in the phase $\psi = (3/m)\phi - \theta$ (we refer to this phase ψ as the "slow phase"), which causes the m^{th} harmonic of the nonlinear driving term to act in a "resonant" fashion on the particle motion. We refer to this as an "m^{th} harmonic third-order (or, third-integer) resonance". In the numerical example of section III, we had $\nu = 0.34 \approx 1/3$. So, we consider the "1^{st} harmonic third-integer resonance". The Hamiltonian when near this resonance is as follows:

$$\mathcal{H}(\phi, J, \theta) = \nu J + \frac{\sigma}{24\pi}\nu\beta_0^{3/2}(\frac{2J}{\nu})^{3/2}\cos(3\phi - \theta) \ .$$

This Hamiltonian, like every other one encountered so far, is a function of θ and is therefore not an integral of the motion. However, we are now in a position to obtain an integral of the motion! We make a canonical transformation to the "slow phase" ψ in the following manner:

$$\phi \rightarrow \psi = 3\phi - \theta$$
$$J \rightarrow K = J \ .$$

This is, in fact, canonical and is generated by the generating function

$$F(\phi, K) = (\phi - \frac{1}{3}\theta)K \quad .$$

The Hamiltonian, in these variables, is

$$\begin{aligned} \mathcal{K}(\psi, K) &= \mathcal{H}(\phi, K) + \frac{\partial F}{\partial \theta} \\ &= (\nu - \frac{1}{3})K + \frac{\sigma}{24\pi}\nu\beta_0^{3/2}(\frac{2J}{\nu})^{3/2}\cos 3\psi \ . \end{aligned}$$

This IS explicitly θ independent. $\mathcal{K}(\psi, K)$ is therefore an integral of the motion. Particle trajectories in phase space therefore always lie on level curves of the function $\mathcal{K}(\psi, K)$. To characterise particle motion, we need only plot these level curves in (x, p_x) space (at some particular θ) and see if any particles can go off to infinite displacement.

We will investigate the geometry of the phase space at $\theta = 0$ (the location of the surface of section) using this invariant. Returning (by means of a chain of inverse transformations) to the (x, p_x) plane (with $\theta = 0$), we find that the invariant may be written as follows:

$$\begin{aligned} \mathcal{K} = \frac{\sigma\nu}{24\pi}\{(x - \frac{4\pi(\nu - \frac{1}{3})}{\sigma\beta_0})(x - \sqrt{3}\beta_0 p_x + \frac{8\pi(\nu - \frac{1}{3})}{\sigma\beta_0}) \\ \times (x + \sqrt{3}\beta_0 p_x + \frac{8\pi(\nu - \frac{1}{3})}{\sigma\beta_0})\} \quad + \quad \frac{\sigma\nu}{6\pi}(\frac{4\pi(\nu - \frac{1}{3})}{\sigma\beta_0})^3 \ . \end{aligned}$$

Level curves of this function are plotted in Figure 9. They represent particle paths in the surface of section. There are stability limits identical to those in the numerical example, and phase space distortion identical to that observed numerically. We may therefore conclude that the behavior in our numerical example was due to the first-harmonic third-integer resonance we have been discussing.

The method discussed in this example can be generalised to treat two- or three-dimensional cases[18]. Further calculations can give other information of interest, such as:

- tune shifts with amplitude,
- "stop band widths" (these are figures of merit which quantify the fact that particle tunes need not be exactly rational for the particle to get "locked on" to a resonance and to be lost. A stop band width is the minumum unsafe distance from the resonant tune; if a particle tune is less than the stop band width from the resonant tune, it will be lost),
- the effect of "crossing" a resonance. Often, machine tunes are modulated, for example by synchrotron (energy) oscillations during acceleration, so

that a particle crosses and recrosses a resonance. It is possible to estimate the effect (most importantly, the amplitude increase) from such repeated resonance crossings[19].

3. Non-Resonant Calculations Considerations such as those just discussed dictate that the tune of a machine be far from any rational number. This is, of course, not possible, as the rationals are dense on the reals, but it is possible to make the tune "irrational enough" to insure that the accelerator will operate. This fortunate circumstance is due to the fact that resonances have ever weaker effects as the order of their nonlinearity increases. That is, if the tune is near a rational number m/n, the effect of the associated resonance becomes smaller as the denominator n becomes larger. This is due simply to the fact that the nonlinear driving terms must become smaller as their order n increases if the accelerator is to operate properly.

In conventional machines resonances through fifth order ($n = 5$) should be avoided (though in some machines, such as the $Sp\bar{p}S$, resonances of very high order are driven by the beam-beam force and can be clearly observed[20]). In superconducting machines, by virtue of their stronger high-order multipole fields, resonances through tenth or higher order may have to be avoided[21]. In either case, if the machine tune is properly chosen, it will be possible to avoid strong resonant effects.

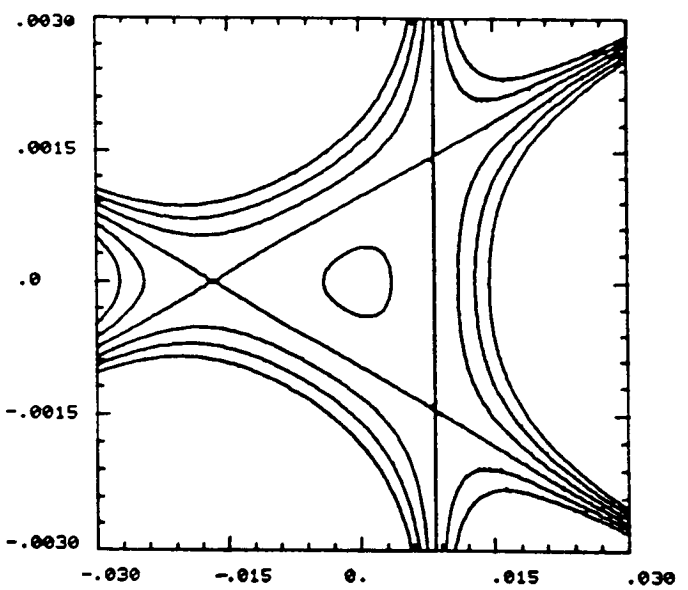

Figure 9. Level curves of invariant for third-integer first-harmonic resonance in linear ring with point sextupole; in this example $\beta_0 = 10$ m, $\nu = 0.34$, and $a_2 = 1, a_3 = 0$. The horizontal axis is in meters, the vertical in radians.

Provided the tunes are chosen so as to avoid resonant effects, all nonlinear driving terms in the equations of motion contribute equally to the particle motion. The slow phase approximation will break down, and the "resonance" calculation of the above example collapses. If no single term is pricipally responsible for driving the particle motion, however, we can study the nonlinear behavior using various non-resonant perturbation theory tools. We will now sketch a few of these approaches.

a) Phase Space Distortion Calculations. These methods examine the "distortion" of the phase space due to a specific nonlinear effect (such as sextupole magnets). They are based on pertubation techniques insofar as they retain only those terms in the equations of motion which are first or second order in the strengths of the nonlinear elements. Several authors have presented procedures for performing these calculations; the principle features of a few are given here.

(1) Calculation of Phase Space Distortion from the Equations of Motion[22]. At the 1984 Snowmass Workshop, T. Collins presented a treatment that accounts for phase space distortion due to sextupoles. In this method, it is noted that a perfectly "linear" machine will transport a phase space ellipse through a complete turn with no distortion. That is, if a transfer map is applied to any point on this ellipse, the image will lie on the ellipse. (This is simply a statement about the Courant-Snyder invariant.) A machine with nonlinearities will distort the ellipse, producing a "squashed-egg" effect or other undesirable result. However (for sufficiently weak sextupoles) the distortions will be linear or quadratic functions of the sextupole strengths.

The method therefore proceeds by propagating the linear invariant through the lattice, "kicking" each point on the invariant at each sextupole. At the end of one turn, the distorted invariant (with terms linear in sextupole strengths only) is compared to the origional "linear" invariant. The distortion is attributed to a set of "distortion functions", which are to be viewed as "nonlinear" analogues to the linear lattice functions. Comparison of the linear invariant and its distorted image allows us to "read off" the distortion functions (or at least their linear dependence on sextupole strengths).

Given the distortion functions, we can construct an invariant phase space "shape", which will be unchanged when the transfer map is applied. That is, this shape is mapped back onto itself by the transfer map, if only terms of first order in the sextupole strengths are used in the mapping. If second-order terms are used, the shape will be distorted still further; the additional distortion can now be attributed to a set of "second order distortion functions", which are second order in the sextupole strengths. These can be read off by comparing the "first-order invariant shape" to its "second-order distorted image" under the transfer map containing the terms quadratic in the sextupole strengths.

In this way, a self-consistant result for the "shape" of a single particle phase space is developed, order by order in the sextupole strengths. Although this does not provide an absolute limit to particle stability, it is possible to ascertain if the phase space is horribly distorted, and then to conclude that poor machine performance would be the consequence, with the sextupoles being the culprits. Moreover, this method provides analytic expressions for the first- and second-order distortion functions. It is therefore possible to determine specifically what is causing the distortion (*e.g.* unfortunate choice of tune, excessively large sextupole strengths, oversized β at a particular sextupole, *etc.*). This provides information that may be used to obtain a solution for the machine lattice, which can reduce the distortion.

The distortion functions themselves serve as a nonlinear generalisation of the linear lattice parameters α, β, γ, and ν. (This is discussed below, in the section on Lie algebraic analytic methods).

(2) Phase Space Distortion Through the Hamiltonian: Birkhoff- Moser Transformations. Willeke[23] and Ando[24] employ a method based on a standard canonical perturbation technique. They begin with the full Hamiltonian for the accelerator (in which all phase space distortions are buried), rather than the equations of motion. They then employ a Birkhoff-Moser transformation to develop an expression for a phase space invariant.

The Birkhoff-Moser normalisation technique proceeds as follows[25]:

- Begin with the Hamiltonian for single particle motion, which includes specific nonlinear driving terms:

$$H(u,p,\theta) = \frac{1}{2}(p^2 + \nu^2 u^2) + F(\theta,u,p) \quad .$$

  Here, θ is the azimuthal position around the machine, and $F(\theta,u,p)$ describes the nonlinear driving terms. For example, for point sextupoles of strengths g_i distributed at azimuthal angles θ_i about the machine,

$$F(\theta,u,p) = \sum_i g_i u^3 \delta(\theta - \theta_i) \quad .$$

  In general, the nonlinear term will be of the form

$$F(\theta,u,p) = \sum_{k=3}^{\infty} A_k(\theta) u^k \quad .$$

- This Hamiltonian is not an invariant because it is time (i.e., θ) dependent.
- If H could be made θ-independent, it would provide a constant of the motion, which would define invariant curves in phase space. If the curves

were closed, the motion would be stable; if the curves were open, the motion would be unstable (see the third integer resonance example).

- We therefore proceed to try to eliminate the θ dependence in H. We make a series of canonical transformations; the first eliminates the θ dependence in the u^3 term, the second eliminates the θ dependence in the u^4 term, and so forth. Hence, we take an order-by-order approach in which we "push the θ dependence to sufficiently high order that it may be neglected". In this approach, after the first transformation, θ influences only $u^4, u^5, \ldots$; after the second transformation, θ affects only $u^5, u^6, \ldots$; and so on.
- After the desired number of transformations, the Hamiltonian will have the following form:

$$\overline{H} \sim (p^2 + u^2) + B_3 u^3 + B_4 u^4 + \ldots + B_n u^n + A_{n+1}(\theta) u^{n+1} + \ldots \quad .$$

In this Hamiltonian, all terms of order n and lower are θ-independent.

- We then proclaim terms of order $n+1$ and higher to be negligiable and thereby obtain a θ-independent invariant:

$$\overline{H} \sim (p^2 + u^2) + B_3 u^3 + B_4 u^4 + \ldots + B_n u^n \quad .$$

- This invariant may be used to draw invariant curves in phase space (i.e., to examine the geometry of the phase trajectories in a surface of section), and thus to study the stability of single particle motion.

This treatment assumes that we are working on eliminating θ-dependences order-by-order in powers of u or p. In need not be done in this way; all that is required is that there be a small parameter to treat in a perturbative fashion. The strengths of magnetic elements (or the strength of perturbations of magnetic fields), if small, could provide a perturbative parameter that can be employed in the generation of a Birkhoff series.

Example Consider a storage ring containing a number of nonlinear elements with fields of strength g (this strength could be considered, for example, to be the pole-tip field strength) that are small. The position dependence of the field is taken to be $f(u)$ and the distribution of these elements about the azimuth of the ring is described by $h(\theta)$ (which will often be taken to be a sum of delta functions). Then,

$$H(u, p, \theta) = p^2 + u^2 + g f(u) h(\theta) \quad .$$

The object of the Birkhoff-Moser procedure is to make a transformation to new variables $\overline{u}$, $\overline{p}$ with Hamiltonian $\overline{H}$ that is to have no "g" term dependent on θ. Specifically, we pick a canonical transformation that eliminates any terms linear in g depending on θ. When this is done, the Hamiltonian is of the following form:

$$\overline{H}(\overline{u},\overline{p},\theta) = \overline{p}^2 + \overline{u}^2 + g\overline{f}(\overline{u}) + g^2\widetilde{f}(\overline{u})\widetilde{h}(\theta) \ .$$

In general, the new Hamiltonian $\overline{H}$ will contain terms of order g^2, which are θ dependent.

A second such transformation can send the θ dependence to order g^3, generating a newer Hamiltonian $\hat{H}$ that has the following general structure:

$$\hat{H}(\hat{u},\hat{p},\theta) \sim \hat{p}^2 + \hat{u}^2 + g\hat{f}(\hat{u}) + g^2\hat{\widetilde{f}}(\hat{u}) + g^3\hat{\overline{f}}(\hat{u})\hat{h}(\theta) \ .$$

This process can be repeated indefinately, making the θ dependence go to continually higher order in g. At some point (g is declared to be small), the procedure is stopped and the statement $g^n h(\theta) = 0$ made. The result is an invariant of order $n-1$ in g, the strength of the magnetic elements.

This procedure has been employed by various authors; we now discuss the work of two of these.

Ando[24] includes chromaticity sextupoles in the Hamiltonian. A canonical transformation is made, which eliminates all terms that are linear in the sextupole strengths. This yields an invariant containing only terms of second order in the sextupole strengths; using this invariant, the configuration of the distorted phase space can be determined

This method will reduce to the usual single resonance treatment if there is a particularly strong driving term due to a tune near a third integer.

Willeke[23] explicitly includes in the Hamiltonian terms describing all relevant linear and nonlinear fields in a machine. This Hamiltonian is subjected to a Birkhoff-Moser transformation, which eliminates all θ-dependent terms that are linear in the strengths of the nonlinear elements (*i.e.* linear in the "pole-tip field strengths" of the nonlinear elements). The result is subjected to a second such transformation, which eliminates θ dependences that are quadratic in the element strengths. This pair of transformations produces a Hamiltonian that has θ-independent terms that are linear and quadratic in field strengths; the only θ dependences are of cubic or higher-order in the field strengths.

The strengths of the nonlinear elements are then declared to be small, so that the series is "descending". The cubic and higher order terms are then neglected, so that a θ independent quantity — an invariant — is obtained. This invariant then provides phase trajectories of particles.

Although this method does not give an absolute limit to stability (which is the definition of dynamic aperture that we have adopted) it does motivate an alternative definition of dynamic aperture, in terms of phase space distortion. Willeke notes that small amplitude particles will lie on the linear phase ellipses but at larger amplitudes, where nonlinear elements "take hold", the phase space trajectory with nonlinearities oscillates about the nomininal linear trajectory of a particle launched with the same initial conditions. This is illustrated by Figure 10.

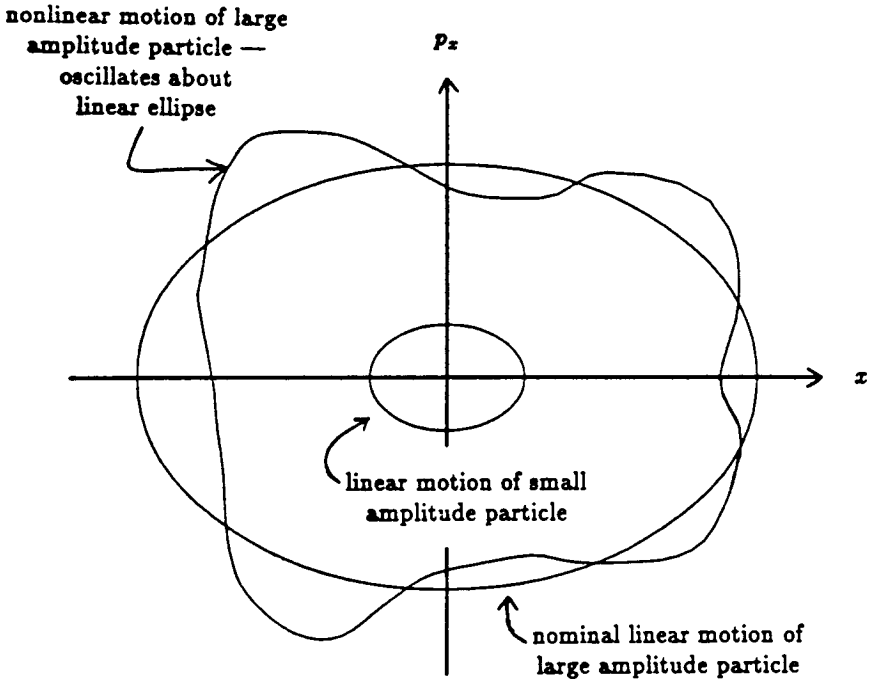

Figure 10. Oscillation of particle trajectory (including nonlinear fields) about nominal linear trajectory.

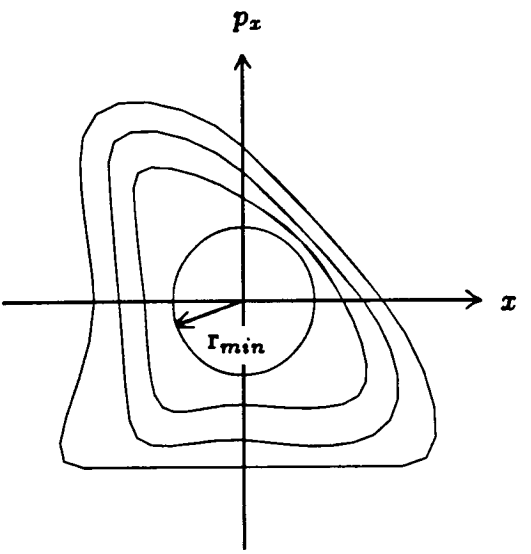

Figure 11. Definition of minimum radius of distorted phase trajectories and maximal inscribed ellipse.

As the initial amplitude increases, the oscillations about the linear phase ellipse increase in magnitude. At some point, the minimum radius of the distorted ellipse ceases to increase; particles launched at increasing amplitudes return to the same minimum radius r_{min} because of the increased amplitude of oscillation. This is illustrated in Figure 11. The area of the largest phase ellipse that fits inside r_{min} is then taken to be the dynamic aperture.

The methods of Ando and Willecke both provide analytical information about the causes of phase space distortion (and hence, about the sources of limitations on dynamic aperture). Both methods require computation of the Fourier transformations (in θ) of the nonlinear driving terms in the accelerator. It is thus possible, within these procedures, to study these driving terms and attempt to minimise their most dangerous effects, thereby producing a reduction in the level of phase space distortion.

Methods such as the ones just discussed, as well as resonance analysis procedures, have been implemented on computers and are regularly employed as design and operation tools for accelerators. The computer program HARMON[26] performs computations of a similar nature (analytically evaluating parameters such as resonance strengths, as well as various chromatic and geometric aberrations) and uses the information thus obtained to optimise the distribution of sextupoles in accelerators.

b) Transfer Map Methods Using Lie Transformations. We will mention two intimately related procedures employing Lie transformation representations of the transfer map. The first, due to Dragt[27], deals with the computation of "nonlinear lattice functions" and provides a connection between the more conventional methods described above and the group theoretical tools used in the Lie transformation description. The second procedure, due to Forest[28], provides a powerful method for analysis of phase space distortion and quantification of the effects of nonlinearities. Both these methods are perturbative in powers of displacements from the origin of phase space; they work with expansions about the design orbit. A third important and powerful method has been developed by Michelotti[29], based on an algorithm invented by the celestial mechanician Deprit[30]. This method, like the phase space distortion methods sketched above, is perturbative in powers of nonlinear element strengths. The reader is referred to the literature for the details of this procedure.

(1) Nonlinear Lattice Functions[27]. In this procedure, nonlinear generalisations of the usual lattice functions are extracted from a Lie algebraic representation of the transfer map for a circular accelerator. This is to be compared to the distortion function method of Collins. The method proceeds in the following manner.

- The transfer map for a circular machine is represented in terms of Lie transformations. These are simply a representation of elements of a Lie group, the group of symplectic mappings generated by the Hamiltonian (discussed

in section II; see Theorem 3); the particular representation employed is discussed below in section IV.B.2).

- Using Lie group methods, the transfer map is decomposed into two parts: 1) a part which, when acting on a "linear" phase ellipse, leaves it undistorted, and 2) a part which contains all terms in the transfer map that act to distort the linear phase space.
- The first (linear) term in the transfer map is viewed as describing (or "causing", if the transformation is considered from the "active" viewpoint) idealised betatron oscillations about the machine.
- The second (nonlinear) term contains the "nonlinear lattice functions" (analogous to Collins' distortion functions), which may be read off directly from the Lie transformation representation of the transfer map.
- Other features include the following:

  The two parts of the transfer map may be recombined and a phase space invariant constructed and used to characterise the amount of phase space distortion.

  As nonlinear lattice functions are obtained, the concept of "matching" is extended to second, third and higher order in x, p_x, y, p_y, T, and P_T.

  Although the method assumes the tune is nonresonant, it is possible to extend the procedure to the resonant case if powers of the transfer map (or, equivalently, multiple-turn transfer maps) are used.

(2) Normal Form Algorithm for Transfer Maps[28]. This algorithm, like the preceeding one, is based on a Lie transformation representation of the transfer map. Using methods of group theory and symmetry principles, the method subjects the transfer map to a series of canonical transformations and reduces it to a particularly simple Lie transformation representation. The resulting transformation (for a third-order Lie transformation description of the machine) contains only linear, cubic, and higher order terms (the second-order having been removed by the canonical transformation).

The final "normal form" of the Lie transformation representation explicitly provides values for tune shifts with amplitude, $\nu_x - \nu_y$ and $2\nu_x - 2\nu_y$ coupling terms, and second order chromaticities. This method therefore immediately yields substantial information about chromatic and geometric properties of accelerators, including nonlinearities through octupole order. The method has been implemented in the analysis and tracking program MARYLIE[31].

4. Limitations On Analytical Methods Analytic methods provide a variety of useful information about accelerator behavior. Using methods such as those outlined above, analyses of machine design and performance can be made which will indicate sources of poor performance, and may give insight as to possible solutions to problems driven by the nonlinear dynamics.

However, the analytical methods are somewhat limited in scope and do not incorporate all of the effects that could lead to difficulties in machine operation. We therefore review some of the restrictions that apply to the methods outlined above.

Resonance Analysis assumes that the tune is in the vicinity of a single isolated resonance. There are no general methods available for handling cases when the tunes lie near two or more strong resonances. Moreover, this technique assumes that there is only a single relevant "slow phase" and that only the slowly varying terms contribute significantly to the motion. This is untrue if the particle motion is stochastic. In this case, even the rapidly varying terms can contribute substantially and must not be neglected.

Phase Space Distortion Analyses all assume the existance of integrals of the motion; this is not necessarily true (the motion may be stochastic, meaning that we cannot think of a particle as moving on a "distorted" phase ellipse; in general, no such ellipse exists). In addition, all these analytic methods are perturbation theories in the strength of the perturbation (e.g. in powers of sextupole field strengths). Such methods are, in general, non-convergent. That is, if we develop higher and higher order (in the perturbation strength) expressions for the invariant, the resulting series can become divergent. It is therefore possible that the "low-order" expressions for the invariant that these approaches produce are not good approximations to the true behavior of the particle, unless the perturbation strengths and/or amplitudes are unrealistically small.

The fundamental problem is that these theories are limited to studying low orders of the perturbing field strength [sextupole or $(\text{sextupole})^2$ terms]. In most modern accelerator designs, it is likely that such terms will be quite large. In superconducting hadron colliders, this is due to persistant current sextupole effects in the superconducting magnets. In electron-positron colliders and synchrotron radiation rings, it is due to strong chromaticity sextupoles that are used to compensate the chromatic aberrations generated by high field quadrupoles (which are necesary to provide small beam spot sizes for high luminosity or high photon-beam brightness). As a result, studying the motion of a particle may require retaining more than just those terms that are linear or quadratic in the sextupole strength.

In this case, it is possible that the Birkhoff transformation procedure will become nonconvergent. If we then look at only the first few terms, the motion may appear well behaved, but if the full series (all orders in the perturbation strength) were used, the motion would appear divergent. We should therefore adopt the attitude that analytic methods provide **necessary but not sufficient** conditions for good machine behavior. If the motion appears irregular when only the lowest-order effects are incorporated into the description, it is most likely that the machine will be badly behaved. The converse, however, is not true: simply because the analytic model predicts regular motion, we are not assured that a more accurate approximation will still demonstrate that the motion is regular. Methods for determining if the motion is well behaved (using numerical techniques) will be discussed in subsequent sections.

Lie Transformation Treatments assume that the motion is nonresonant and that an invariant exists at some level (either in the transfer map or in some power of the transfer map). In addition, these methods are still limited to relatively low orders in the perturbation strength, and low orders of deviations in phase space variables from design values. At present, the methods are limited to octupole-order perturbations, and incorporate only terms through square of the strength of the perturbation. Further developments of these techniques are aimed at relaxing these constraints[32].

The basic philosophy of the machine designer should then be that all available analytic tools must be applied in the planning and optimisation of a design, and all useful analytic tools should be employed in understanding machine operational behavior. Nonetheless, analytic tools are restricted to relatively simplified cases by their very nature. In addition to analytic results, we need models of accelerators that will include arbitrarily high orders of all types of perturbations of all strengths in order to certify machine behavior in the design phase, and understand observed machine behavior. This modeling is possible only through the use of numerical methods, a subject to which we now turn our attention.

B. Numerical Approaches to Dynamic Aperture[33]

1. "Tracking" as Numerical Experimentation The "numerical example" of section III illustrates the fundamental principle employed in the numerical determination of dynamic aperture. The entire process is experimental, and consists of iterating the transfer map on a selected initial condition to determine the stability of that initial condition. This process is referred to as "tracking" the initial condition, and is repeated for several choices of initial condition in order to map out the stable and unstable regions of phase space. The dynamic aperture is then taken to be the maximum stable undistorted phase space volume, as determined by this experimental process.

Example In the numerical example of section III, the maximum stable amplitude was determined to be +8 mm and −16 mm. At these values, separatrices were encountered and the motion became unstable. This is illustrated by Figure 12. In this figure, we see that the maximum amplitude that an inscribed ellipse (the "undistorted phase space") encloses without intersecting a separatrix is limited to 8 mm. In keeping with the definition of dynamic aperture in terms of the undistorted phase ellipse, we take the 8 mm value, rather than the 16 mm, to be the dynamic aperture, because the −16 mm displacement cannot be achieved without distortion of the phase space.

The fundamental principle of numerical determination of dynamic aperture is therefore an experimental one: the transfer map is numerically iterated for a variety of test initial conditions so as to determine empirically the stability of various regions of phase space. The behavior of the iterates of the initial condition (the "intersection of the particle trajectories with the surface of section") are then analysed so as to define a "maximum stable amplitude", which is then taken to be the dynamic aperture.

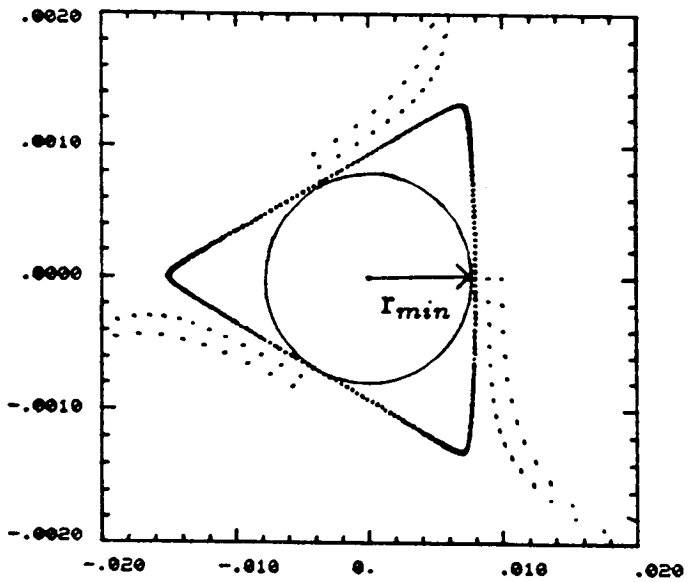

Figure 12. Illustration of definition of dynamic aperture, as experimentally measured in section III. See Figure 6c.

2. Alternative Representations of the Transfer Map We now turn our attention to specific methods that are commonly employed to perform the numerical iteration required for tracking experiments. All the methods described below are alternative ways of representing the transfer map, which describes particle motion around the accelerator. If the approximations made in each representation are valid, the results obtained from any tracking experiment using one method will agree with results from experiments using any other method. (This comparison of results from various programs is in fact a powerful method for determining the range of validity of the approximations made in the numerical models employed by programs, and has proven in practice to be a useful method of improving confidence in the results of particular tracking studies[34]. A sampling of the programs using these methods will also be given with a description of the methods.

a) The "Thin Lens" Model. This model is employed by a variety of programs, including PATRICIA[35], RACETRACK[36], TEVLAT[37], and TEAPOT[38]. The method uses linear transformation matrices to represent the so-called "linear" elements such as dipoles and quadrupoles. "Nonlinear" elements, such as sextupoles, octupoles, or higher multipoles, are treated in the impulse approximation (which was discussed in the numerical example of section III).

In such programs, a normal period of a FODO lattice can be represented and simulated as shown in Figure 13. In this figure, a lens shape superimposed

on an element denoted that a nonlinear "kick" transformation will be evaluated at that location to simulate the nonlinear effect of field errors in dipoles or quadrupoles, or the effect of sextupoles, on the test particles. A test particle is transported from lens to lens by linear transformations appropriate to the dipoles, drifts, or quadrupoles lying between the lens sites.

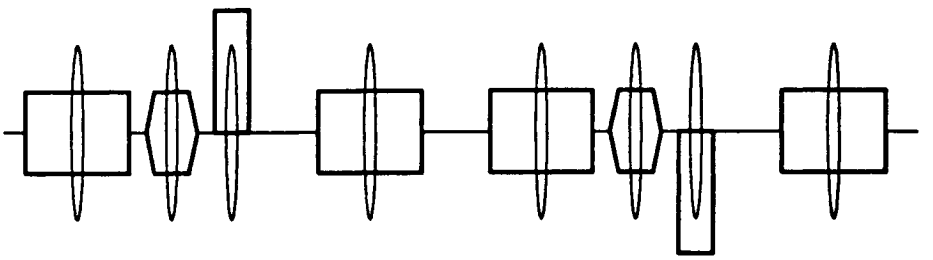

Figure 13. Representation of a FODO structure with thin-lens model. At location of superimposed lenses, a nonlinear kick transformation is evaluted; test particle motion from lens to lens is modelled using linear transformations.

The nonlinear "kicks" are identical to those used in the numerical example of section III. For example, for a sextupole, the transformation is as follows.

$$\bar{x} = x$$
$$\bar{p}_x = p_x + \alpha x^2 .$$

b) Canonical Integration Method[39]. This method, which is to be provided as an option in the program MAD[40], numerically integrates the equations of motion for particles moving through the accelerator. It employs an integration algorithm that is explicitly canonical, that is, one that preserves Poisson brackets, so that each integration step is itself a canonical transformation.

The explicitly canonical nature of the integration algorithm is vital to the tracking process. If the Poisson brackets are not conserved by the integration method, other invariants may also be unconserved. This can lead, for example, to violations of Liouville's theoren, and result in spurious damping or growth of the phase space that is being modeled. This damping or growth is unphysical, as it is a mathematical arifact, but could be misinterpreted as an actual physical effect. In the case of spurious growth, the dynamic aperture could be underestimated, in the case of damping, it could be overestimated.

Example (by L.J. Laslett[41]) In this example, the equations of motion for a model of an alternating gradiant focussing system with a quadratic (sextupole) focussing term were numerically integrated for several thousand periods of the focussing structure. When a fourth-order Runge-Kutta algorithm was employed, the phase space exhibited damping (illustrated in Figure 14), due to the noncanonical character of the integration algorithm. When a third-order Ruth[39] algorithm was used, no damping was observed (Figure 15).

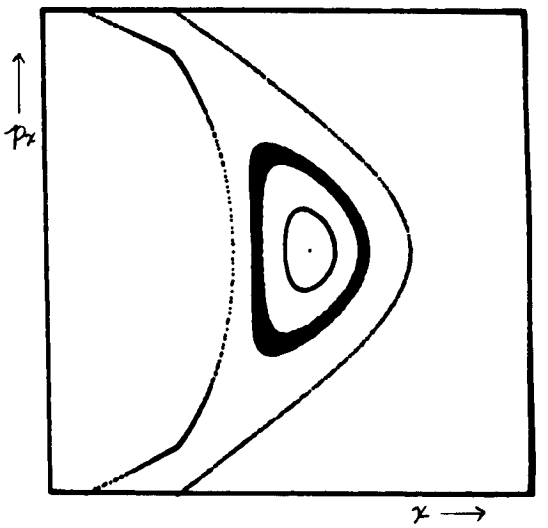

Figure 14. Integration of an alternating-gradient focussing channel with sextupole component for 16000 periods using a Runge-Kutta algorithm. Damping is evidenced by thick band around central island[41].

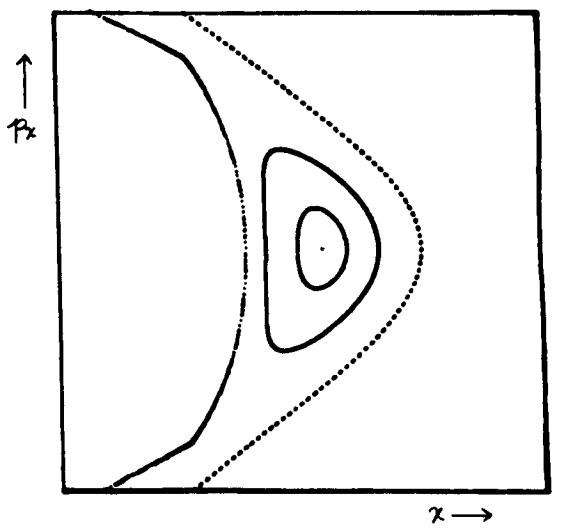

Figure 15. Integration of an alternating-gradient channel with sextupole for 16000 periods using a Ruth algorithm. No damping is visible[41].

Canonical integration may be viewed as an extension of the thin lens method to include terms of higher order in the element length. In the thin lens model, positions are unmodified during the transformation, and the momenta are modified only by terms linear in the element length. Canonical integration algorithms enhance the accuracy of this approximation, by modifying both positions and momenta by terms of linear, quadratic, or higher order in element length. A period of a FODO structure (such as that shown in Figure 13) would then be simulated by this method as shown in Figure 16.

In this illustration, each tick mark represents a single integration step (a single canonical transformation). At the very least, each element (each magnet and drift) must use one integration step; for greated accuracy, the step size is decreased and the number of steps per magnet is increased. This is, of course, at the cost of lower execution speed of the computation.

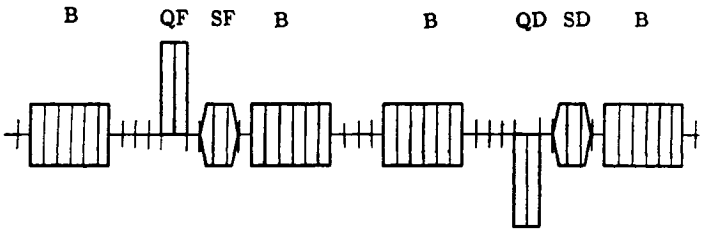

Figure 16. One FODO period, as simulated by canonical integration. Each "slice" is a single integration step.

c) Lie Transformation Techniques[1] This group-theoretic based technique is employed in the programs MARYLIE[31] and MAD[40]. It represents the transfer map in terms of mathematical objects called "Lie transformations". These transformations are operators that are directly generated from Hamilton's equations and that can be easily evaluated numerically.

Using Hamilton's equations (1) and Definition 3 of the Poisson bracket, we can briefly sketch the principles of this method. We define, using the Poisson bracket, a "Lie operator" as follows. Let f and g be two functions on phase space; the Lie operator generated by f is defined by the action it has on any function g in terms of the Poisson bracket:

$$: f : g \equiv [f, g] \ .$$

Hamilton's equations may be written in terms of the Lie operator generated by the Hamiltonian. This is done in the following manner:

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} \\ &= [q_i, H] \\ &= -[H, q_i] \\ &= - : H : q_i \end{aligned}$$

and

$$\begin{aligned}\frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} \\ &= -[H, p_i] \\ &= - : H : p_i \ .\end{aligned}$$

Hamilton's equations may therefore be cast in the following operator form, for any of the six phase space coordinates ξ_i, $i = 1, 2, \ldots, 6$:

$$\frac{d\xi_i}{dt} = - : H : \xi_i \ .$$

This may be formally integrated (in the time-independent case, such as holds within a single beam-line element) to obtain the following solution for the motion of the particle:

$$\xi_i(t) = e^{-t:H:}\xi_i(0) \ . \tag{12}$$

In this expression, the exponential operator is a "Lie transformation" and is defined by the following operator series:

$$e^{-t:H:} \equiv \sum_{n=0}^{\infty} \frac{(-t)^n}{n!} : H :^n \ .$$

Powers of the Lie operator are defined in the natural fashion, using multiple Poisson brackets; for any function g on phase space,

$$\begin{aligned} &: H :^0 g \equiv g, \\ &: H :^1 g \equiv [H, g], \\ &: H :^2 g \equiv [H, [H, g]], \quad etc. \end{aligned}$$

The relationship (12) represents the complete, nonlinear, canonical transformation describing motion through a beam-line element.

This representation has a number of useful features.

- Like the transfer maps they represent, the collection of Lie transformations employed in this method form a group. This group has a well defined product, so that any two such transformations may be combined to give a single transformation.
- All such transformations are symplectic (i.e., canonical).
- Given a Hamiltonian, the transformation is obtained immediately and in all detail.

A single period of a FODO structure can then be represented by a single transformation, through use of the group product on the group of Lie transformations. This is illustrated by Figure 17.

The transformations describing all elements can be "lumped" or, more formally, "concatenated", to give a single transformation. All elements are then treated at once, rather than individually or a few at a time.

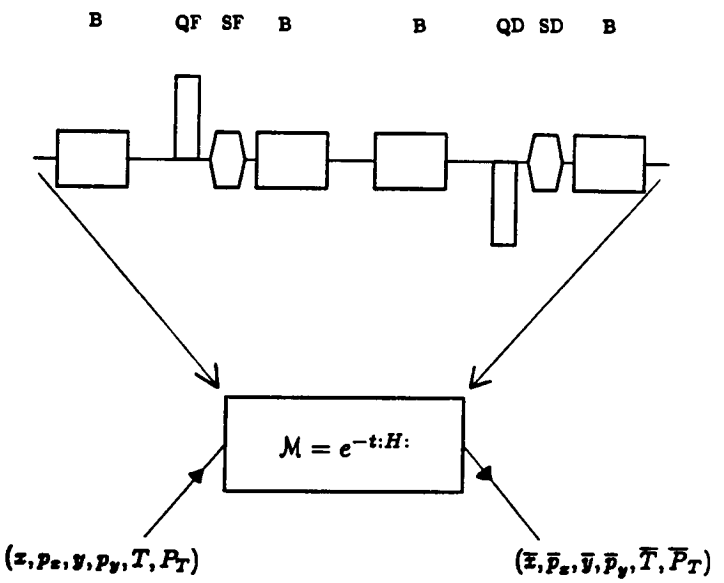

Figure 17. Schematic representation of use of Lie transformation to represent a collection of beam-line elements in a FODO period.

d) Generating Function Methods are based on the fact that the transfer map for an accelerator, or any part thereof, is a canonical transformation. It therefore admits representations in terms of generating functions[2], and may be parameterised in the following way. For $i = 1, 2, 3$,

$$\overline{q}_i = \frac{\partial F(q, \overline{p})}{\partial \overline{p}_i}$$
$$p_i = \frac{\partial F(q, \overline{p})}{\partial q_i} \ ,$$

where $F = F(q, \overline{p})$ is the "generating function".

This technique is employed in certain parts of the programs DIMAD[42], MARYLIE[31], and MAD[40]. It has a number of features to be noted.

- The method is explicitly canonical.
- The generating function may be obtained from the equations of motion, from a Lie transformation, or from matrix methods (which are discussed below).
- The equations are implicit:

$$\overline{q} = \overline{q}(q, \overline{p})$$
$$p = p(q, \overline{p}) \ .$$

When evaluating this transformation, we know only q and p, so we must solve numerically for $\bar{q}$ and $\bar{p}$.

We remark that this method has applicability when other methods, such as the matrix methods to be discussed below and the Lie transformation methods discussed above, are used. Matrix methods do not in general provide canonical transformations, and, for numerical purposes, Lie transformations are in general truncated in the following noncanonical fashion:

$$e^{-t:H:} \cong \sum_{k=0}^{N} \frac{(-t)^k}{k!} :H:^k \ .$$

It is possible, though the use of a generating function, to rewrite such transformations and return them to a canonical form that is convenient to evaluate numerically[43].

e) "Higher-Order" Matrix Methods are employed by the programs MAD[40] and DIMAD[42]. These methods represent the transfer map in terms of a Taylor series representation; for any $i = 1, 2, \ldots, 6$,

$$\begin{aligned}\bar{\xi}_i =& a_1^i x + a_2^i p_x + \ldots + a_6^i P_T \\ &+ A_1^i x^2 + A_2^i x p_x + \ldots + A_{21}^i P_T^2 \\ &+ \mathcal{A}_1^i x^3 + \mathcal{A}_2^i x^2 p_x + \ldots + \mathcal{A}_{56}^i P_T^3 \\ &+ \ldots \ .\end{aligned}$$

This series is truncated, normally to second order[44]. It may be written as follows:

$$\bar{\xi}_i = \sum_k M_{ij}\xi_j + \sum_{j,k} T_{ijk}\xi_j\xi_k \ .$$

The matrix elements M_{ij} and T_{ijk} may be obtained directly from the equations of motion[45]. However, the truncated Taylor series expansion is not a canonical transformation; the Poisson brackets are not preserved. It is possible to bring the transformation to canonical form through use of the generating function technique described above. Like Lie transformation methods, this representation has the advantage that sets of series expansions for a collection of beam-line elements can be multiplied together ("concatenated") so that a group of elements can be described by a single transformation.

f) Capsule Comparison of Methods

We now summarise the above discussion by presenting a brief comparison of all of the aforementioned methods; the strengths and weaknesses of each are contrasted.

Thin Lens

+ Is of "arbitrarily high order" (can describe all order perturbations).

+ Allows a fast ray trace.

+ Can be made canonical (symplectic, Hamiltonian).

- "Thin Lens" approximation is good for short, strong nonlinear elements, but may not be good for long elements with weak, distributed nonlinearities (such as superconducting hadron collider dipoles). For example, beam-beam interaction forces often involve megagauss fields over lengths of millimeters or centemeters. They are therefore well modeled by an impulsive approximation, as are chromaticity sexupoles in electron storage rings, which involve kilogauss fields over tens of centemeter lengths. The model is, however questionable in cases such as superconducting dipoles, where the effect being modeled is due to tens of gauss distributed over tens of meters.

- A single ray trace is fast, but an entire turn can be very slow if many nonlinear elements are present (must treat each nonlinear element as a separate ray trace).

Canonical Integration

+ Is of arbitrarily high order.

+ Is explicitly Hamiltonian.

+ Is "self-checking", that is, the integration step can be subdivided and see if the results are consistant with those generated using a larger integration step.

- Can do at most only one element per ray trace, implying the method is slow.

- Is presently limited to effects of fourth or lower order in the element length; this may not be sufficient to incorporate all important terms in the equations of motion without going to many integration steps per element. For example, a sextupole transfer map has terms cubic in phase space displacements that are seventh order in the sextupole length. Proper representation of such terms would require multiple integration steps in each sextupole.

Lie Transformations

+ Is explicitly Hamiltonian (but need to use generating function to evaluate ray trace).

+ Can combine many nonlinear elements into one map, implying the existence of a fast "global" ray trace.

- Single ray traces are slow.

- "Exact" representation currently limited to octupole or decapole.

- Limits of "concatenation" not well defined; there are no general rules as yet, which govern the number of elements that can be modeled by a single transformation.

Generating Function

+ Is explicitly Hamiltonian.
- Difficult to combine elements.
- Single ray traces are slow.
- High orders are inaccessible (limited to octupole or decapole at present; but can be extended by combining with Lie algebraic techniques).

"Higher-Order Matrix" Methods

+ Can combine elements.
+ Fast single ray trace.
- Limits to combination of elements unexplored: how many elements can be combined?
- Limited to low order (sextupole or octupole, at present), and not easily extended to higher orders.
- Not canonical (but can be made canonical using generating function).

3. Constraints On Tracking Programs The above algorithms place constraints on the details of models that use them; in addition, the physical details of motion in an accelerator place constraints on the approximations that can be made in modeling motion in an accelerator. The first constraint is that the particle motion be Hamiltonian on the time scales of interest. In proton machines, this is a rather long term; in fact the motion, to a good approximation, is Hamiltonian indefinately, insofar as numerical modeling is concerned. In electron machines, the motion is modeled as Hamiltonian on the natural time scale of the machine, which is the damping time. If particles are stable, in a Hamiltonian model, for several damping times, then they may be assumed to be stable more or less indefinately.

The constraint that the model be Hamiltonian implies certain detailed constraints on the model. First of all, the phase space should be modeled as six-dimensional. Four dimensional codes (which transport only x, p_x, y, and p_y) neglect the fact that transverse motion can take energy from longitudinal motion; such programs therefore violate conservation of energy and should be employed with that fact in mind. Secondly, all transformations used in tracking should be canonical transformations; the code must use a symplectic representation of the transfer map. This is because the motion is generated by a Hamiltonian, so an accurate representation of the transfer map will be symplectic.

Example of Problems Arising From Nonsymplecticity In this example, we track a simple lattice of combined function FODO cells, in which chromaticity

correction sextupoles were incorporated in the (combined function) dipole magnets. In a first experiment, tracking was done using a nonsymplectic second-order matrix ray trace; in a second experiment, tracking was done using a symplectic ray trace produced from the second-order matrix transformation by means of a generating function. The matrix (Taylor's series) ray trace falsely indicated an instability. The symplectic ray trace indicates that the particle is stable. See Figure 18.

The nonsymplectic ray trace does not preserve the Poisson brackets of all phase space variables with one another. It therefore leads to a violation of Liouville's theorem which is manifested as a spurious growth in the phase space over the course of the tracking experiment (1000 turns). The symplectic ray trace preserves the Poisson brackets; no phase space growth is evident, even if we extend the tracking to 50,000 turns! (See Figure 19.)

A general tracking program must include those nonlinear effects that potentially limit dynamic aperture. For most accelerators and storage rings, this will mean the following effects can be modeled.

- Chromaticity sextupoles.
- Any other multipole terms (usually, only octupoles; in some cases, it may be desirable to model error multipoles in accelerator magnets. In the case of large hadron colliders (see next section) this is the predominent problem, and special treatments are required.
- Proper modeling of (T, P_T) behavior, that is, the so-called "synchrotron oscillations" must be included. The model should include (or demonstrate that it is not necessary to include, rather than just neglect) the coupling between energy and transverse motion. This will include those terms in the transfer map that link transverse variables to energy deviations (through the rf system) and energy deviations to transverse displacements (through dispersion effects).

4. Special Demands on Codes Used to Study Large Hadron Colliders We first set the scale of the problem. The reference designs study lattice A[16] was the smallest of the proposed SSC lattices, and thus represents a lower limit on the scale of the model. This lattice contained approximately 4000 16.6-m-long 6.5-Tesla dipoles, 1000 quadrupoles, and 750 sextupoles in each of two rings of 90 km circumference. The revolution time was thus 90/c or 3.0×10^{-4} seconds; the synchrotron oscillation period of approximately 1000 turns represents about 0.3 seconds of real time, and, from the revolution time, in one second of real time a particle will execute 3333 turns.

Using a model of this machine, one is called upon to predict behavior on two time scales. The first is that of short-term behavior, over the first few thousand turns, or the first few seconds of real time. Studies on this time scale will provide an understanding of possible injection problems. The second time scale is that for long-term behavior, and involves modeling for minutes or hours

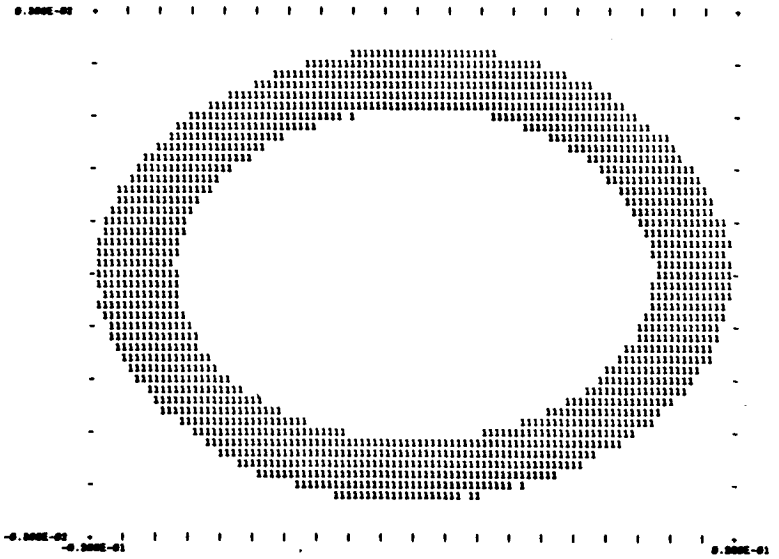

Figure 18a. Phase space with nonsymplectic ray trace; motion appears unstable during 5000 turns in that the amplitude of the test particle grows steadily outward, causing the "smearing" effect.

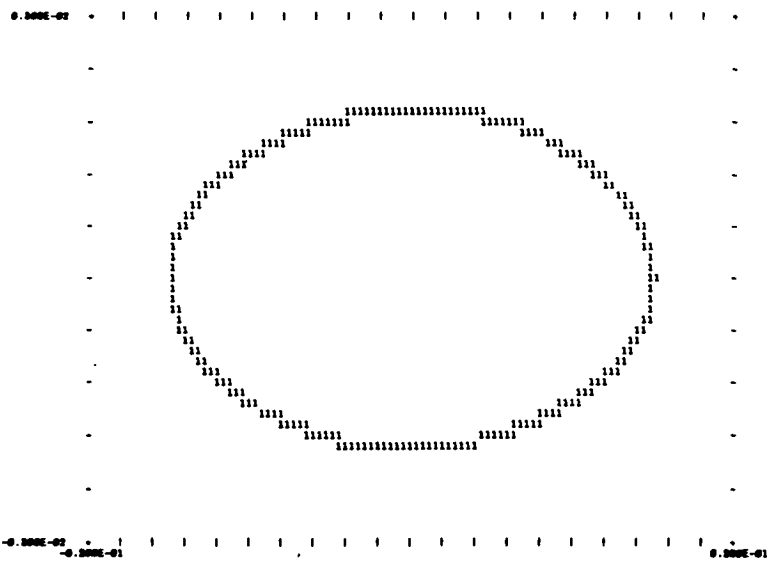

Figure 18b. Phase space with symplectic ray trace; motion appears stable for 5000 turns.

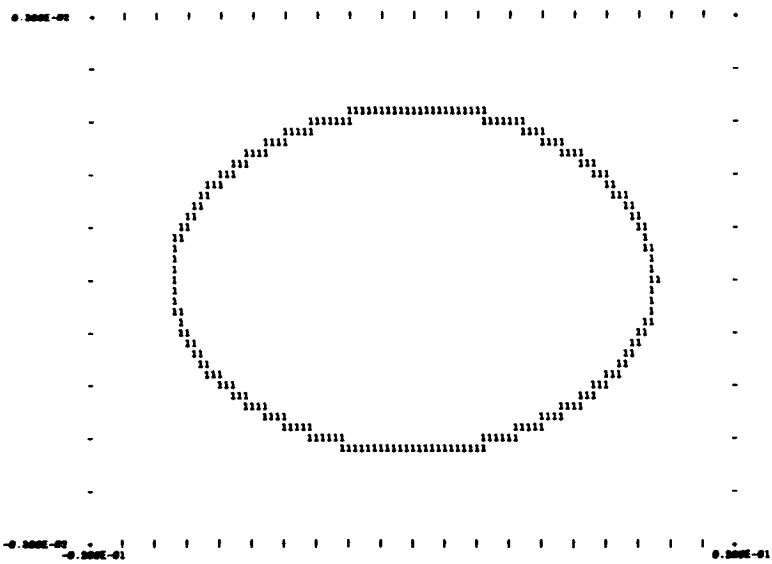

Figure 19. Phase space over 50,000 turns with symplectic ray trace; motion appears stable over entire course of experiment.

of real time (10^6–10^8 turns). Modeling over this time scale will allow predictions of beam lifetime.

The problems of physical interest for a large hadron collider focus entirely on the issue of field quality in superconducting magnets of the type that would be used in such a machine. The critical accelerator physics issue for a large hadron collider is the impact of magnet field quality on dynamic aperture and beam lifetime.

As mentioned in the example of section IV.A.1., superconducting dipole magnets have a variety of "imperfections" so that their fields are not uniform; there are relatively strong nonlinear dependences on position even at the center of the magnet aperture. There are two types of field error that can affect particle stability. The first is systematic errors, which are the same from magnet to magnet. These are due to the design of the coils, shaping of any iron used in the magnet, hysteresis effects such as persistant currents, and so forth. Systematic errors can be well characterised and can be compensated "locally" in the lattice of the machine; that is, compensation can be provided in each cell or in the magnets themselves, through the use of properly designed correction coils.

The second type of errors of importance here is that of random errors in the magnets. Such errors, as the name implies, have different values in each magnet, and are due to sources such as construction errors, variations in iron properties, or variations in quality of superconductor. These errors can be very strong in superconducting magnets, and become stronger when the magnet bore is made

smaller. This is, heuristically, because the beam in a small- bore magnet is closer to the boundary of the magnet, so that it will "see" the error more closely. In order to minimise costs, the magnets for large hadron colliders must be relatively inexpensive to build, implying both lower tolerances on construction errors and less raw material (implying smaller bore). Hence, the random error fields in large hadron collider magnets can contribute significantly to the determination of the machine behavior.

In a superconducting collider, each magnet thus has a vertical field component of the following form:

$$B_y(x,y) = B_0[b_0 + b_1 x + b_2(x^2 - y^2) + b_3(x^3 - 3xy^2) + \ldots + a_1 y + a_2 xy + a_3(3x^2 y - y^3) + \ldots]$$

with a similar expression for the horizontal component $B_x(x,y)$. In this "multipole expansion", each of the coefficients a_n and b_n are large (relative to values expected for conventional magnets), and unique to a specific magnet.

The "multipole moments" a_n and b_n cause substantial focussing (or bending) for particles that do not move on the design trajectory. This focussing changes in a random (though periodic, with period of one turn) way the behavior of such particles and therefore can limit the dynamic aperture. One example (for systematic multipoles only) was given in section IV.A.1., in which the tune variation with momentum was estimated for a particular design of large hadron collider dipole. As a second example of the impact of the multipoles, we compute the contribution to a dipole field from the random sextupole component of a proposed SSC dipole. From the SSC RDS[46], the random sextupole component b_2 for the design B dipoles has a numerical value of $0.7 \times 10^{-4}/\text{cm}^2$. The error dipole component ΔB at displacement x due to this multipole moment is

$$\Delta B = B_0 b_2 x^2 \quad ,$$

so that the random field at 1 cm at 5 Tesla (the design B base field value) is of the following magnitude:

$$\begin{aligned} \Delta B &= (5 \text{ Tesla}) \times (0.7 \times 10^{-4}/\text{cm}^2) \times (1 \text{ cm})^2 \\ &= 3.5 \times 10^{-4} \text{ Tesla} = 3.5 \text{ gauss} \quad . \end{aligned}$$

This is the random fluctuation in each magnet, from the sextupole alone!

It is therefore conceivable that the "multipoles" in such magnets are strong enough to radically influence both the short- and long- term behavior the the machine. Tracking programs that are to model large hadron colliders must be able to model the effect of several orders of multipoles (b_0 to b_n, with n perhaps as large as 20 or 30) with both systematic and random values in each of 4000 or 5000 dipole magnets.

In addition to modeling error magnetic fields, tracking programs used to model large hadron colliders should be able to incorporate other effects in a detailed fashion. One effect of possible importance comprises orbit errors. These may be important for two reasons. On the short-term time scale, the small aperture (1 to 2 cm) of the magnets may impose severe limitations on allowable orbit errors. The nonlinear error fields may be strong enough to increase these orbit errors and thereby decrease the available dynamic aperture. On the long term, the energy of a hadron collider (5 to 40 TeV) is sufficiently high that radiative damping of the beam phase space can occur on a several-hour time scale. Under certain circumstances, radiative antidamping can arise, perhaps also on the same time scale. These circumstances can be provided, for example, by dipole magnets with sufficiently strong quadrupole focussing and defocussing components in the field, a situation which could possibly arise in a dipole with large orbit errors and very strong nonlinear field components. This effect should, therefore, be investigated, to insure that slow radiative excitation of the beam phase space does not occur.

An additional feature that must be provided by codes used to model large hadron colliders is a realistic means of simulating sychrotron oscillations. As the single-particle phase space is six-dimensional, in order to properly represent the dynamical behavior of the system, (T, P_T) motion must be included. Long-term models of collider behavior typically include tracking for 100 (or more) synchrotron periods[47]. For a large hadron collider, with 1000 turns per synchrotron oscillation, this is 10^5 turns or more. High speed is therefore a very clear requirement for tracking programs used to model superconducting collider behavior.

Example In tracking SSC RDS lattice A with PATRICIA, it was found that modeling one particle for one turn required 91 milliseconds cpu time on a CRAY-1S[48]. To track one particle for 100 synchrotron oscillations (approximately 10^5 turns) will take 9100 seconds, or about 2 1/2 hours of cpu time. One may anticipate tracking dozens of particles in such a study, for each of several lattice configurations (different tunes, interaction point beta functions, *etc.*); this implies a need for days or weeks of dedicated CRAY time for each single dynamic aperture study. This indicates the magnitude of the compute-intensive activity represented by a dynamic aperture study of a large hadron collider.

In summary, tracking in a large hadron collider will require programs with the following features:

- o models for high-order nonlinearities
- o very high speed
- o full six-dimensional phase space (including energy oscillations)
- o other effects, such as
  - orbit errors and corrections
  - "noise" sources such as rf noise or magnet power supply fluctuations.

V. Example Tracking Studies

A. A Large Hadron Collider: The SSC Reference Design Lattice A[49]

The tracking studies we outline were performed with a lattice based on 6.5-Tesla 2-in-1 dipoles. They assumed that systematic errors in the superconducting magnets were perfectly compensated, so that the only errors were due to random field imperfections. The energy was held constant in these simulations, so synchrotron oscillations were neglected. Finally, no orbit errors were introduced; it was assumed that the closed orbit was perfectly corrected. The purpose of these studies was therefore to analyse the effect of random multipoles on the single particle dynamics.

Initial tracking studies were performed by Dell[50] to determine the dynamic aperture of RDS Lattice A[51]. In these studies, the aperture of the "bare machine" was "scanned". By "bare machine", we refer to the accelerator including only linear elements and chromaticity sextupoles; no random errors are modeled in the superconducting dipoles. An aperture "scan" refers to the method of selecting test particles for tracking. In the Dell study, particles were launched along the line $x = y$, with $p_x = p_y = 0$, and tracked for 50 turns to determine the stability limit. The small number of turns employed was due to the large size of the machine; cpu time limits prevented longer runs.

Using a version of the program PATRICIA to do the tracking, it was found that the dynamic aperture, with chromaticity sextupoles only, was larger than the physical aperture of the accelerator. In particular, betatron oscillations of large enough amplitude to hit the walls of the vacuum chamber exhibited stable motion. For the particular lattice studied, this means that if a particle were launched at a horizontally focussing quadrupole in a normal FODO cell of the machine, the maximum stable amplitude (limited only by the physical aperture of the machine, not by nonlinear effects) would be approximately 12 mm in x and 5.8 mm in y. If the particle were launched at an interaction point, this amplitude would correspond to a displacement of about 0.66 mm in either x or y.

Energy dependence of the dynamic aperture was investigated by setting the particle energy to values slightly different from the design energy. The aperture scan was then repeated; results are given in Figure 20. These results are exactly as expected for an "ideal" machine. In such an idealised model, an off energy particle oscillates about a displaced closed orbit located at

$$x_{co} = D(\frac{\Delta E}{E})$$

where D is the dispersion function. The amplitude available for betatron oscillations therefore must go down linearly with energy deviation $\Delta E/E$ (provided the machine is well behaved), due to the motion of the closed orbit toward the wall of the vacuum chamber. Thus, the "energy aperture" of the bare machine is, like the "betatron aperture", limited by the physical aperture of the beam pipe rather than by nonlinear effects.

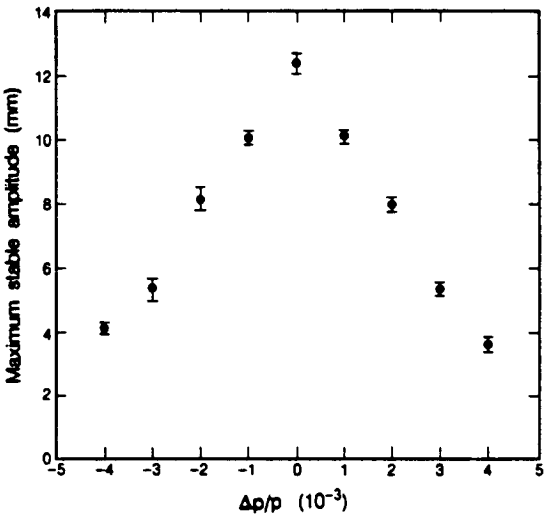

Figure 20. Maximum stable amplitude as a function of energy offset for ideal machine[51].

As implied by earlier discussions (such as the sextupole example of section III and the analytic methods of section IV), the behavior of the particle trajectories in the surface of section, or the "phase space" of the tracking, is a useful diagnostic. In these studies, the phase space of the "bare" machine is fairly undistorted, as illustrated by Figure 21. Plots generated over 400 turns show quite elliptical trajectories even very near the edge of the aperture (at the physical aperture of the machine). The "thickness" of the ellipse is due to coupling of the horizontal and vertical modes of motion through the sextupoles. That is, the sextupoles make x_{final} depend on $y_{initial}$ (and y_{final} depend on $x_{initial}$); this is manifested as a "breathing" effect on the $x - p_x$ and $y - p_y$ phase ellipses, causing a thickening of the phase plots.

Dell included a series of experiments on the machine using random field errors. The procedure was the same, but in addition to chromaticity sextupoles, random magnetic errors in superconducting bends were modeled by appropriate nonlinear transformations (in the impulsive approximation). The results of these experiments demonstrated, through an aperture scan, that random magnetic errors cut the dynamic aperture by more than a factor of 2. The maximum stable amplitude at the interaction point fell from $x = y = 0.66$ mm to $x = y = 0.27$ mm. This corresponds to $x \cong 5$ mm and $y \cong 2.4$ mm at a horizontally focussing quadrupole in an arc FODO cell.

In addition, inclusion of random multipoles had a dramatic impact on the energy dependence of the dynamic aperture. Using the same method as was

Figure 21. Phase space plots for 400 turns in RDS Lattice A, without errors[50].

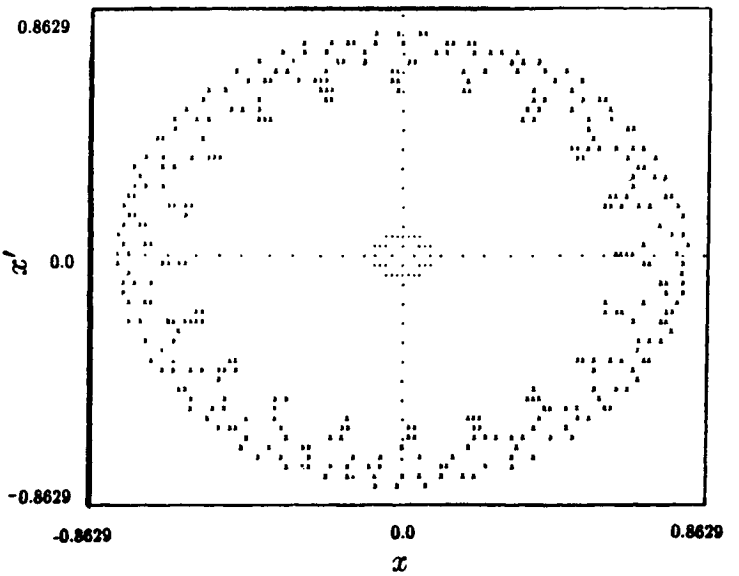

Figure 21a. Horizontal phase space.

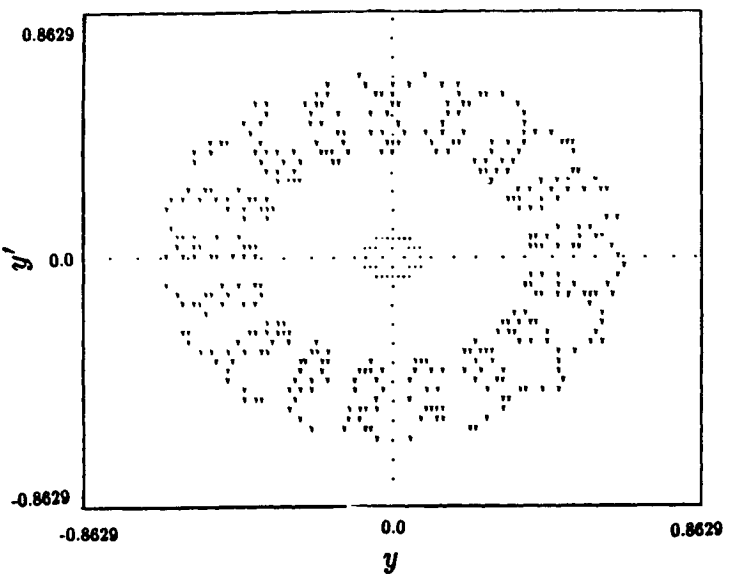

Figure 21b. Vertical phase space.

employed for the bare machine, it was found that the multipoles reduced the maximum stable amplitude for off-energy particles by more than a factor of 2 (Figure 22).

Finally, the phase space became very distorted with the inclusion of random multipoles. This is illustrated in Figure 23. The "diamond" shape indicates possible proximity to a 4^{th} integer resonance (recall dde triangular distortion of the sextupole example in section III), while the thickness of the band around the distorted ellipse indicates strong horizontal-vertical coupling.

The conclusion drawn from this study is that the dynamic aperture of SSC RDS lattice A is seriously limited by the presence of random magnetic errors. Without random errors, the dynamic aperture is larger than the beam pipe; with random errors, it is less than half the beam pipe aperture.

Subsequent tracking studies have repeated and substantiated the results of Dell. Gelfand[52], using the program TEVLAT[53], has provided confirmation of the initial results using an independent computer code. Further studies by Leemann[54] have extended the results of Dell and Gelfand to increased numbers of turns, and have attempted to identify specific magnetic errors limiting dynamic aperture.

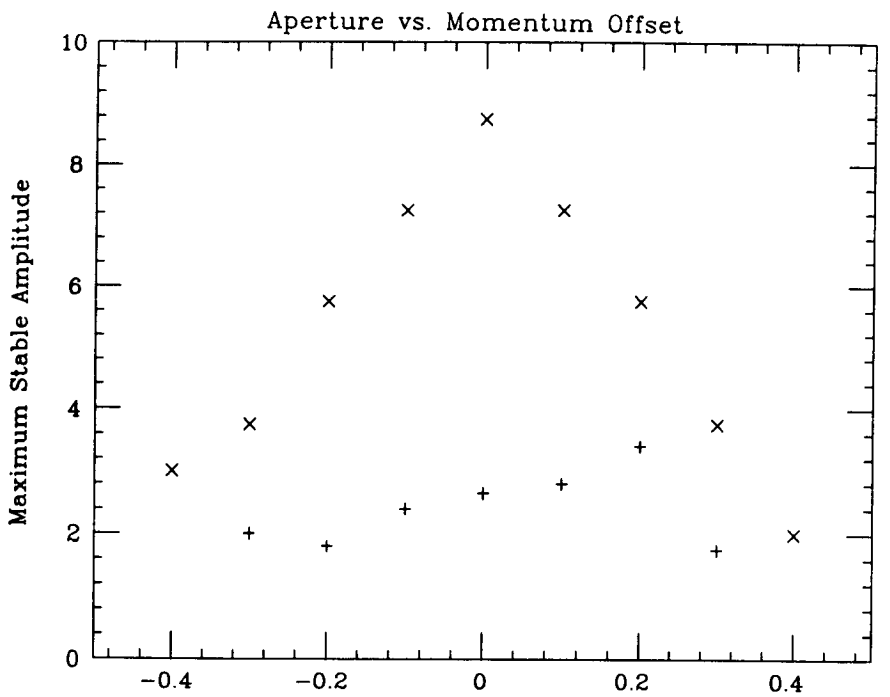

Figure 22. Dynamic aperture with and without random magnetic errors in dipoles[50].

Figure 23. Horizontal and vertical phase spaces for 400 turns in RDS Lattice A, including effects of random multipoles in dipoles[50].

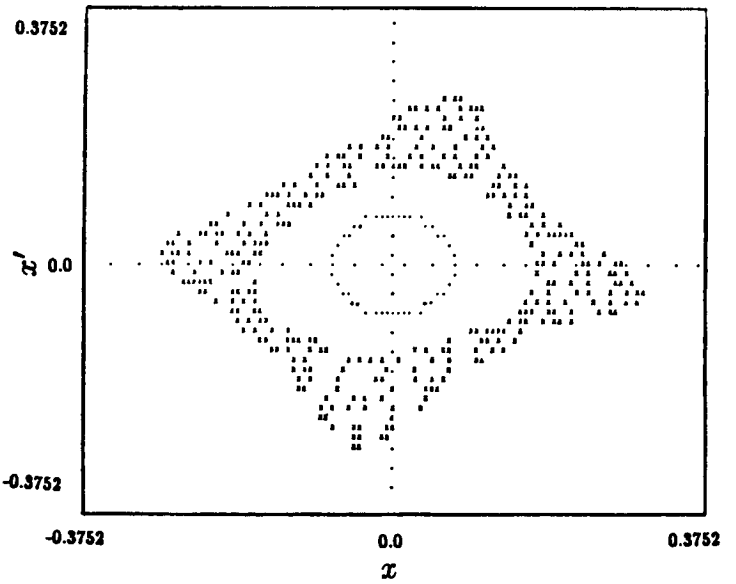

Figure 23a. Horizontal phase space.

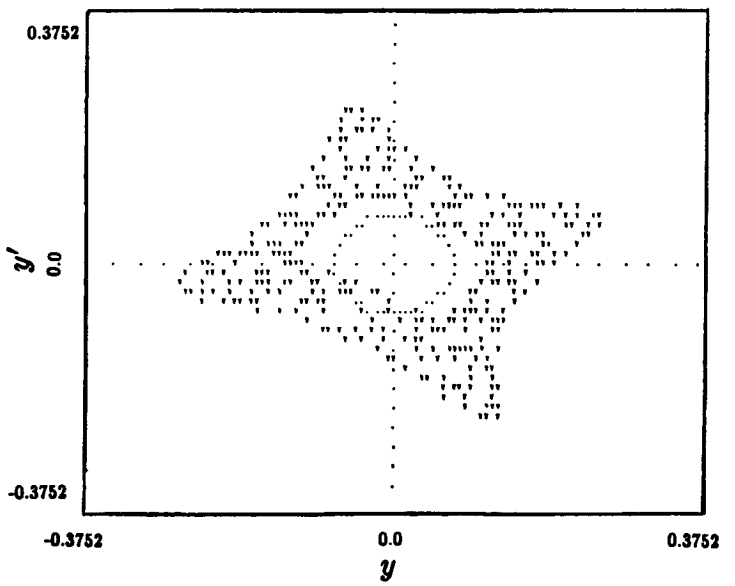

Figure 23b. Vertical phase space.

Early tracking by Dell and Gelfand was done for only 50 to 400 turns; Leemann has extended these studies, using the program PATRICIA, to the 1600-turn regime (and has gone as far as 10000 turns in isolated cases). These studies show that the 1600-turn dynamic aperture is smaller than the 400-turn dynamic aperture; the maximum stable amplitude falls to less than 0.2 mm at the interaction point (which is equivalent to less than 3.6 mm maximum displacement in an arc cell). This represents a 30% decrease. Moreover, these studies are consistant with an important observation of Dell and Gelfand, namely, that the dynamic aperture is dependent in a nontrivial fashion on the tune of the machine.

Some effort was made by Leemann to determine if specific error multipoles (or combinations of error multipoles) are responsible for limits on dynamic aperture. Results were not transparent in this regard. For example, in one case it was found that the aperture was substantially smaller if only random quadrupole and sextupole effects were included than it was when randon quadrupole, sextupole, and octupole were included. Some speculation holds that the results may depend on the way in which random multipoles are generated, but the exact cause of particle loss at the edge of the dynamic aperture remained unclear. Figure 24 illustrates the phase space for two possibilities. The upper plot illustrates the horizontal phase space near the stability limit; the diamond shape is characteristic of motion near a 4^{th} integer resonance. The lower plot illustrates the horizontal phase space just outside the stable limit; the diamond shape has disintegrated and the motion appears irregular. This is characteristic of the stochastic motion discussed earlier. Thus, it is not clear in this case if the dynamic aperture is limited by resonant particle loss alone, or if particles at the stable limit are in a stochastic layer.

All of the above studies support two almost certain conclusions. The first is that the dynamic aperture of SSC RDS Lattice A is nonzero (at least for 10^5 turns or so) but is also noninfinite. The second conclusion is that random multipoles have a serious negative impact upon that dynamic aperture. However, several questions remain unanswered. What is the exact mechanism of particle loss? It is due to phase space distortion, to proximity to a resonance, or to the stochastic limit? What is the dependence of the dynamic aperture on the selection of the "working point" (the choice of tunes ν_x and ν_y)? Are there particular magnetic field errors leading to limits on dynamic aperture, are all dangerous, or are only certain combinations of error multipole responsible for the observed degradation of the dynamic aperture? Will the selection of a different set of random errors cause radically different results? What is the effect of energy (synchrotron) oscillations — what happens when all of the six-dimensional phase space is opened up to particle motion? What are the effects of systematic errors and the compensation thereof? Is the thin lens approximation accurate? (Recall the dipoles in question have error fields of tens of gauss distributed over tens of meters– which is more of a zero strength/infinite length limit than it is a zero length/infinite strength limit!)

Figure 24. Particle motion near the stable limit in SSC RDS Lattice A[54].

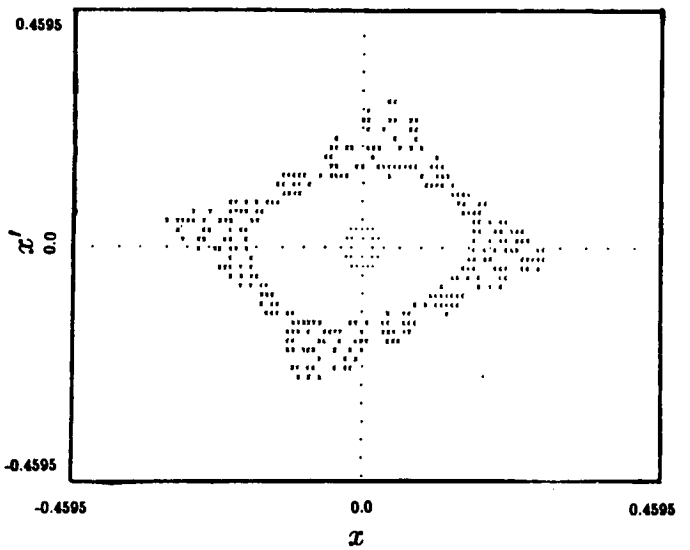

Figure 24a. Horizontal phase space over 400 turns for initial condition just inside the stable limit.

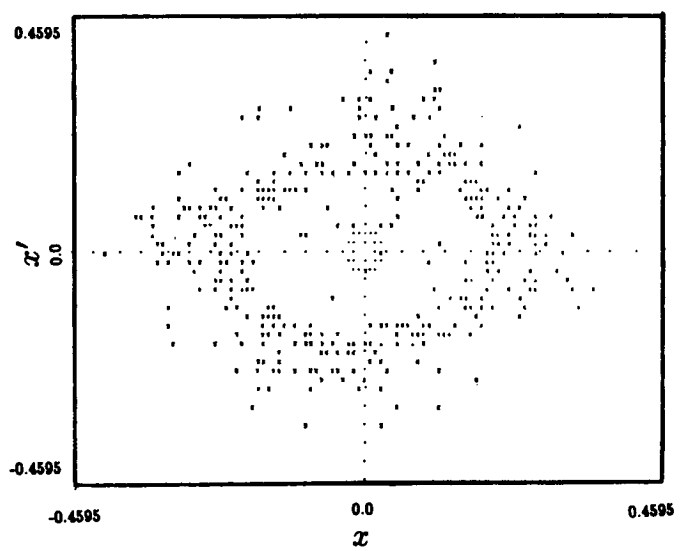

Figure 24b. Horizontal phase space over 400 turns for initial condition just outside the stable limit.

These questions and many related ones have been the subject of intensive study as a result of design projects for large hadron colliders. Many advances have been made in the understanding of nonlinear motion in accelerators as a result of these studies; the reader is referred to the literature for the most recent examples of this work[55].

B. An Electron Storage Ring for the Generation of Synchrotron Radiation

The electron storage ring we will examine is an early design for the Lawerence Berkeley Laboratory Advanced Light Source[56]. This machine is optimised to provide a very small beam spot size, with correspondingly intense photon beams from synchrotron radiation. The ring is extremely strongly focusing in order to provide the small beam spot; the strong quadrupole focusing leads to large chromatic aberrations, requiring correspondingly powerful chromaticity correction sextupoles. One period of the lattice is illustrated in Figure 25; the machine is composed of 12 such periods.

In addition, collective effects requirements (in particular, a requirement for a long Touschek lifetime) lead to a high voltage rf system, meaning that the sychrotron motion is rather high in frequency. The particle motion is, therefore, highly nonlinear, and illustrates several effects very well. These comprise a geometric aberration known as nonlinear detuning (dependence of particle tune on amplitude), suppression of aberrations through use of sextupoles, and two mathematical curiosities: failure of a concatenation process and violation of the symplectic condition. We will discuss each of these in turn, after briefly outlining the basic dynamic aperture calculation for this machine.

The behavior of the machine, as modeled by tracking, is illustrated by Figure 26. This figure shows the $x-p_x$ phase subspace of the machine for several particles launched and tracked for 1000 turns (including energy oscillations). Particles launched with zero transverse momentum and with displacements of $x > 1.5$ cm are unstable and lost. Particles at smaller amplitude have distorted, triangular-shaped phase "ellipses"; near the stable limit, the motion lies on three islands, indicating proximity to a third-integer tune at large amplitude.

In keeping with our notion of dynamic aperture as the maximum stable regularly shaped phase space volume, we have thus determined through tracking that the dynamic aperture of the machine is 6 to 8 mm; however, due to assymetry in the phase space distortion, particles launched with displacements of up to 15 mm along the positive x axis are stable.

The presence of three islands in the phase plot at large amplitude leads us to consider the subject of "nonlinear detuning". The tunes of this storage ring are set to the values $\nu_x = 13.78$ and $\nu_y = 7.78$, which are near the three-quarter integer. However, the phase plot demonstrates that the tune of large amplitude particles lies near a third integer (as indicated by the presence of the three islands). This circumstance comes about because of the nonlinear focussing effect of the chromaticity sextupoles. The magnetic fields in these elements

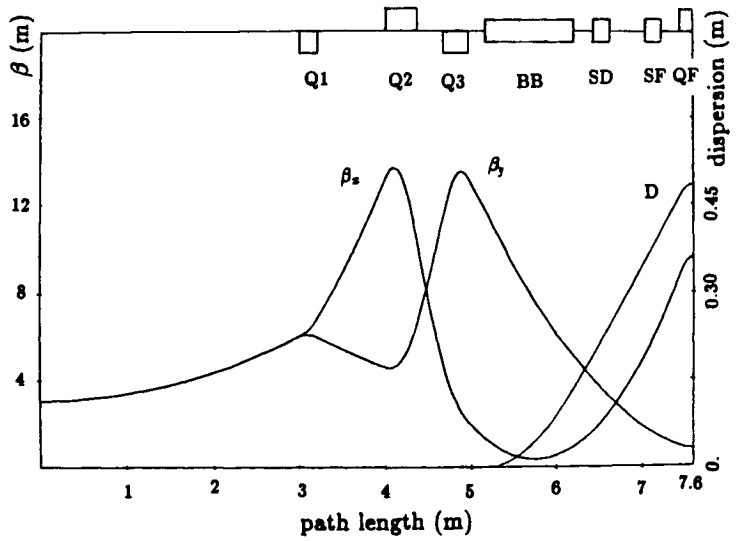

Figure 25. One half period of synchrotron radiation ring, period is reflectively symmetric about QF[57].

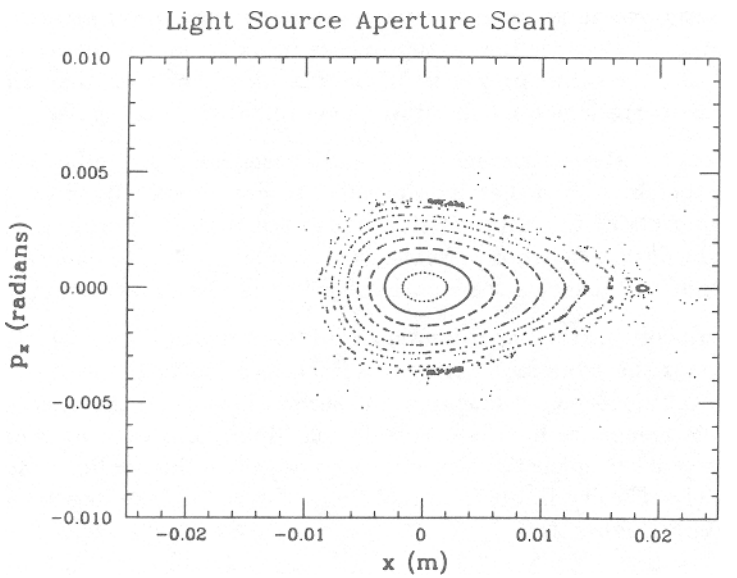

Figure 26. Phase space portrait over 200 turns for 8 particles launched from 2 mm to 16 mm horizontal displacement, including synchrotron oscillations.

are quadratic in particle displacement, so that large amplitude particles are focussed in a substantially different fashion than are small amplitude particles. The number of betatron oscillations executed by a particle is therefore a nonlinear function of the amplitude of the particle; this is the so-called "nonlinear detuning" effect we have mentioned. The effect is referred to as a "geometric aberration" (in analogy with optical effects of the same name) because it is an amplitude-dependent effect.

In the ring under consideration, the effect is quite strong, due to the presence of high field chromaticity correction sextupoles. This effect was first diagnosed by Wiedemann[58], who determined that the horizontal tune falls from 13.78 to 13 2/3 at $x = 7.5$ mm; the tune then continues to fall to 13 1/3 at about $x = 14$ mm. The 13 2/3 value is well within the dynamic aperture of the ideal machine, and could lead to resonant particle loss if error fields exist at a sufficiently high strength.

The geometric aberration in question can, however, be compensated. One method that has been investigated is the use of geometric aberration compensation sextupoles in the long straight sections of the machine. This method is based on chromatic correction schemes proposed by Brown and Servranckx[59] and consists of placing "correction sexupoles" in the dispersion-free straight sections, exactly one-half betatron wave length away from the chromaticity correction sextupoles. Because the sextupoles are in a dispersion-free region, they do not affect the chromatic properties of the machine. However, they do cancel the geometric aberrations introduced by the chromaticity sextupoles.

Exercise 7. Read References 63, and, based on the methods described therein, prove that it is possible to cancel exactly those geometric effects of the chromaticity sextupoles that are linear in sextupole strength. (It is not possible, in a simple fashion, to cancel the effects that are quadratic or higher order in the sextupole strengths.)

The results of this type of correction are illustrated in Figure 27. The first figure illustrates the dynamic aperture and the location of the 13 2/3 integer resonance in the uncorrected machine; the second illustrates the aperture and location of the resonance for the corrected machine. As is evident from these figures, the geometric correction sextupoles compensate the nonlinear detuning in an extremely effective fashion. In addition, they reduce the degree of phase space distortion (Figure 27c).

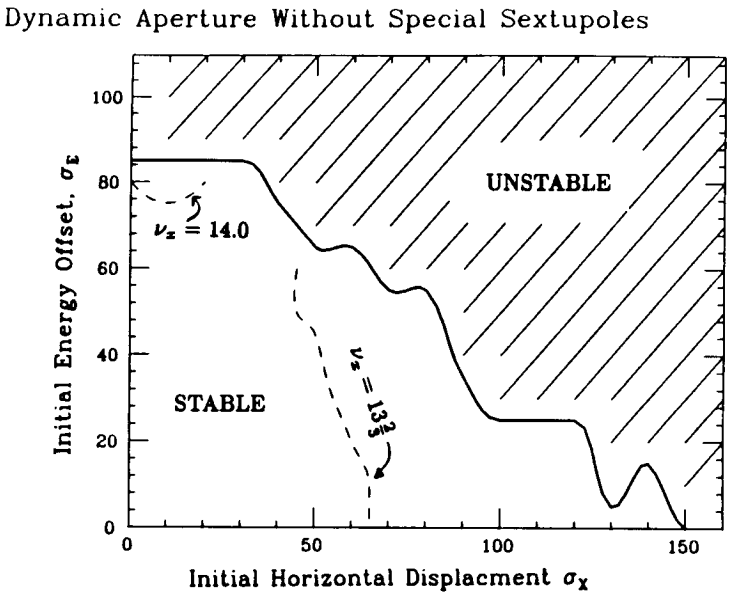

Figure 27a. Stability limits and resonance locations as functions of horizontal location and momentum offset for lattice without geometric compensation sextupoles.

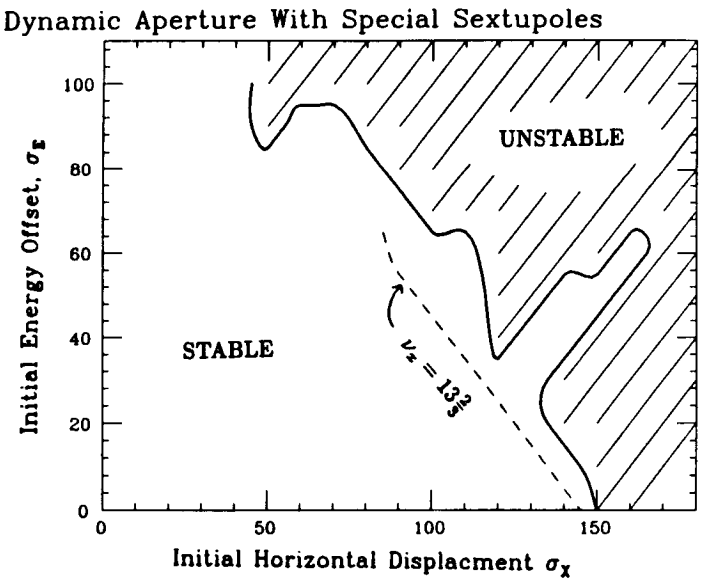

Figure 27b. Stability limits and resonance locations as functions of horizontal location and momentum offset for lattice with geometric compensation sextupoles.

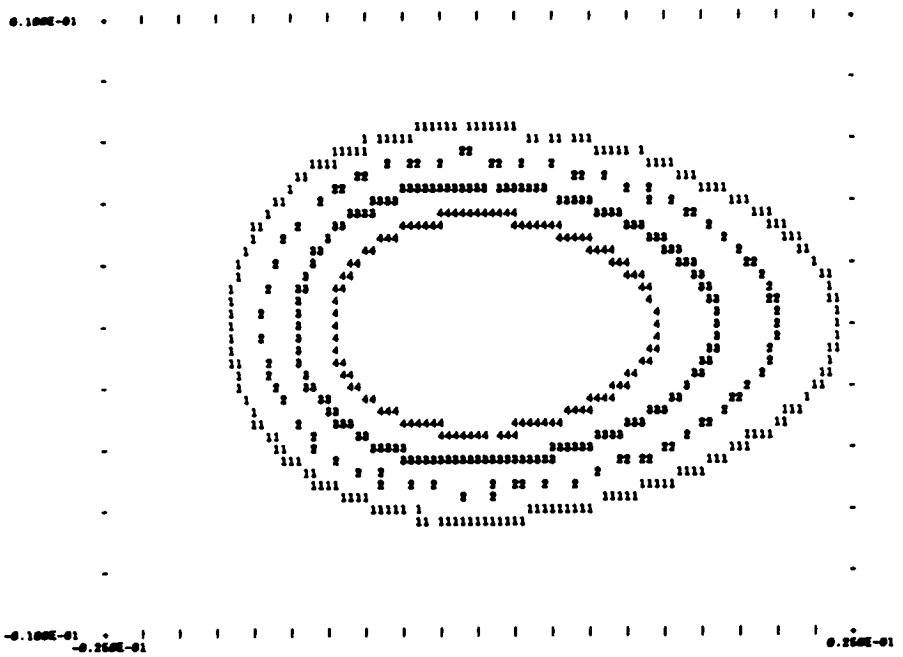

Figure 27c. Phase space portrait with geometric compensation sextupoles. Comparison to Figure 26 indicates the degree of distortion has been reduced by including sextupole compensation.

The strong nonlinearities present in this machine also provide illustration for two mathematical points that we have discussed in this paper. The first is the need for symplecticity in a tracking code. Figure 28 illustrates the stable limit without synchrotron oscillations, computed by DIMAD. The first plot was generated using a nonsymplectic ray trace, the second using a symplectic ray trace. The nonsymplectic result implies the stable limit is substantially smaller than it is found to be using the canonical method. This demonstrates that the need for symplecticity arises in realistic machine lattices; it is not confined to somewhat artificial test cases such as the one discussed in section IV.B.3).

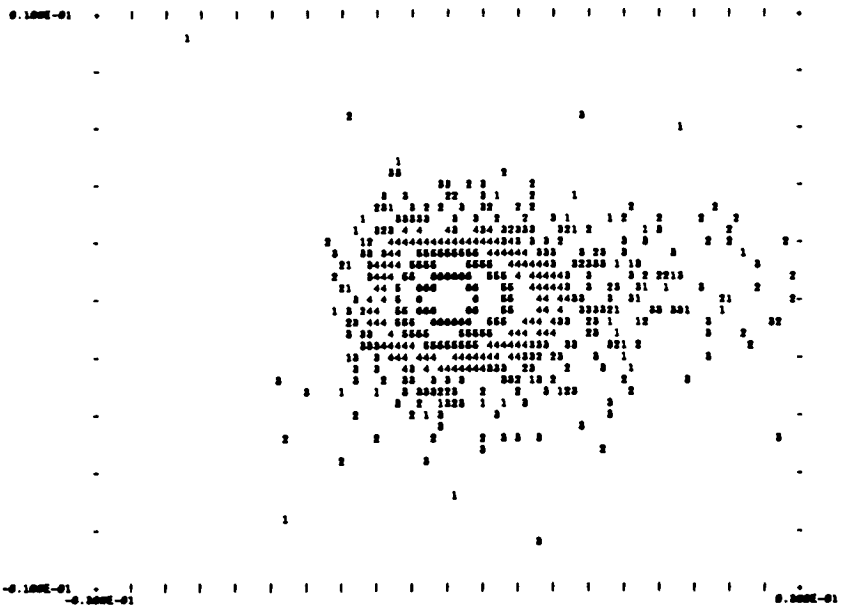

Figure 28a. Stable limits as computed without symplectic ray trace.

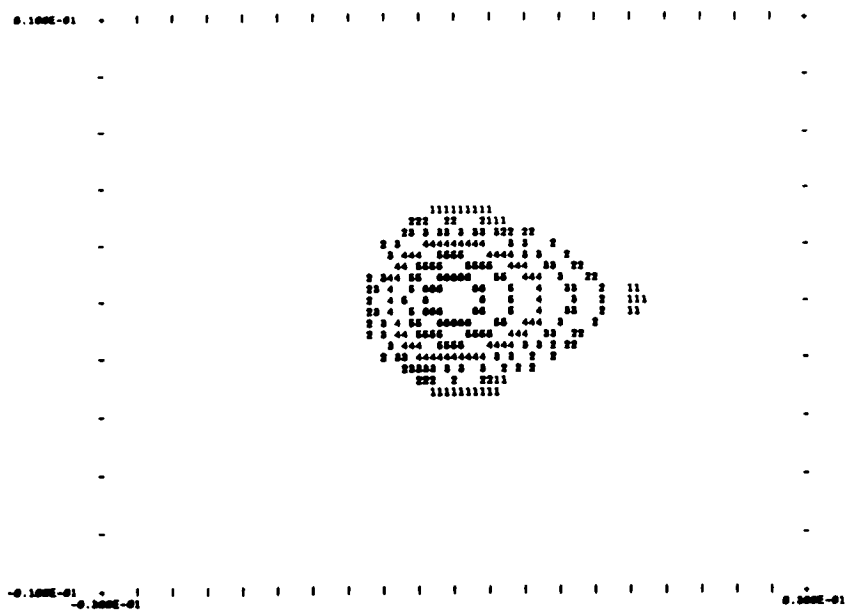

Figure 28b. Stable limits as computed with symplectic ray trace.

The second point of mathematical interest has to do with the issue of lumping several beam-line elements, or "concatenation", discussed in relation to Lie algebraic and Taylor series methods. As noted previously, there are no clear guidelines for determining how many elements may be represented by a tranformation that is accurate to only a specified order. That this is a relevant issue is verified by Figure 29. This illustrates the single particle phase space of the storage ring as computed using three different methods in the program MARYLIE.

The first plot is similar to that in Figure 26; it was generated by tracking for 200 turns without any concatenation of "nonlinear" elements (*i.e.* sextupoles) and without synchrotron oscillations. This may be considered to be the "benchmark" plot; it represents the numerically most accurate method of representing the motion as it uses single transformations to model motion through individual elements. The second plot was generated by tracking for 200 turns without synchrotron oscillations, using a map representing an entire period of the machine. This map thus includes concatenation of four strong sextupoles. It is different than the first plot, and implies that the motion at large amplitude is better behaved than the accurate computation predicts. The final plot was generated by concatenating three periods of the machine and tracking for 200 turns without synchrotron oscillations. It implies the motion is drastically less regular than expected.

Caution must therefore be exercised in employing concatenated representations of machine lattices. Benchmark runs using the most highly subdivided representations possible should be employed to provide a basis against which the more time efficient, but less accurate, concatenated results may be judged.

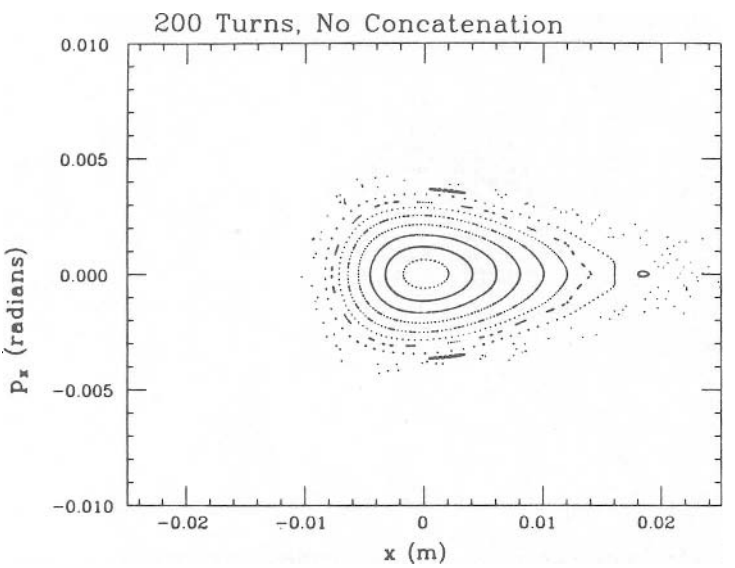

Figure 29a. Phase space without concatenation.

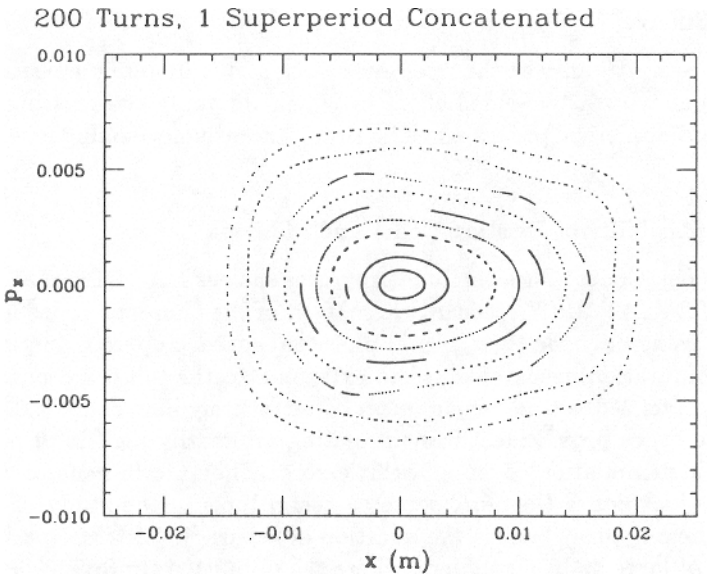

Figure 29b. Phase space with concatenation of one period.

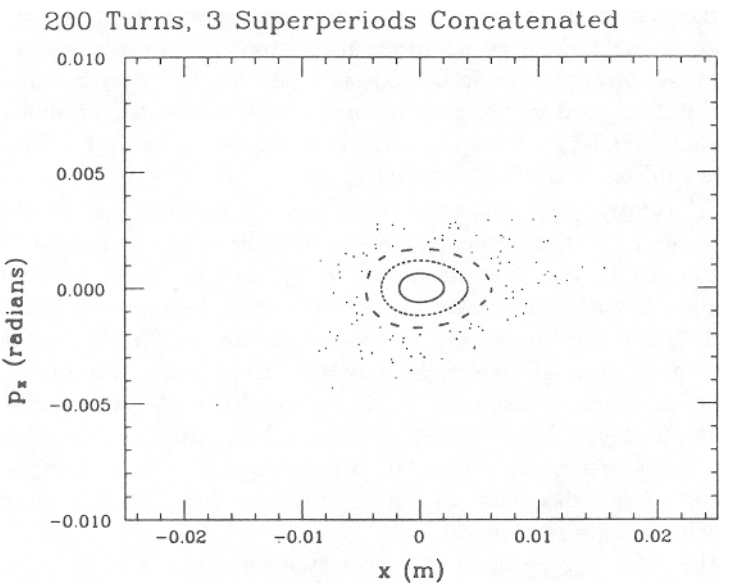

Figure 29c. Phase space with concatenation of three periods.

VI. The Future

How are we to answer the open questions about dynamic aperture? This must be done with a two-sided effort in which we apply the existing analytic methods and numerical programs while simultaniously developing new methods and codes.

A. Applicability of Existing Codes and Methods

Thin Lens Codes These include programs such as RACETRACK, PATRICIA, and TEVLAT. Until recently, these most of the thin-lens codes have been "4×4", or using only the (x, p_x, y, p_y) phase subspace. Superconducting supercollider computation needs have led to extension to the full 6×6 phase space, and have in fact led to the development of an entirely new code, TEAPOT[38]. All of these codes provide features simulating arbitrarily high multipoles, and many provide simulation of closed orbit errors. They are in general very high speed. The kick codes therefore serve as useful tools in the study of dynamic aperture; the only drawback is the question of accuracy. This is especially true in the case of large hadron colliders, where the impulsive approximation is used to simulate the effect of very weak, highly distributed nonlinearities.

Canonical integration methods are not yet fully implemented, though they may be available in the program MAD. These are extremely accurate, but it is not *a priori* clear how large an integration step may be employed. If an entire magnet can be simulated in a single step, the method may be sufficiently fast that it could be used for long-term studies; if several integration steps are required in each magnet, the method may be too slow to be useful for general studies and would be restricted to "benchmark" runs.

Lie transformation methods are available in the programs MARYLIE and MAD. They are typically limited to low-order multipoles (though MARYLIE has been extended to model multipoles through fifth order). Simulation of magnet misalignments and mispowerings have recently become available[60]. Associated analytical tools[28] have been incorporated into MARYLIE, so that such programs can perform analytical as well as numerical tasks. The major drawback to use of Lie transform methods is the coupled questions of speed and concatenation. If several beam-line elements can be lumped, the programs are quite fast; if concatenation is not possible (as in section IV.B.) the programs are, relative to other codes, slow, by virtue of the mathematical and numerical complexity of the single ray trace.

Generating function methods have been typically used as auxiliary calculational tools for symplectification in programs that are based on other techniques (MAD, MARYLIE, and DIMAD). They are thus undeveloped as independent simulation tools. As with the Lie transformation techniques, the mathematics and numerical evaluation is, relative to other techniques, complicated, so that concatenation must be accepted for high speed. Analytic and numerical treatments using these methods are being developed[61] and may prove to be powerful

tools for studying the aperture question.

Taylor series expansion methods such as those used in DIMAD and MAD must be coupled to generating functions to be canonical. They are limited to low-order effects (sextupole, octupole); higher-order effects are typically simulated through thin lens methods. The programs using these methods contain, however, detailed models of a variety of effects including closed orbit errors and mispowerings.

Of these methods, only canonical integration, Lie transformation, and generating function methods can investigate the effects of high-order distributed nonlinearities in an accurate fashion. (It is, in principle, possible to do this with thin lens methods. However, this requires breaking each element up into many thin lenses, which will degrade program performance in a significant fashion.) However, of the existing programs, only the thin lens codes, MAD, and DIMAD, are fully implemented and closely coupled to standard design codes. Others are more analytical in structure and do not provide information that is useful for diagnostic purposes (such as momentum dependences of the tunes and beta functions, lattice functions throughout the machine, *etc.*). Therefore, further code development is warrented. The programs should be extended to include synchrotron oscillations, closed orbit errors, and magnet mispowerings. Extension of the methods (*e.g.* Lie tranformations, generating functions, *etc.*) to describe high-order nonlinearites should be carried out.

B. Computing Hardware Requirements

The basic requirements are high speed and large memory. The necessity of high speed was discussed earlier, when in the example of section IV.B.4) we noted that PATRICIA would require about 2 1/2 CRAY-1S cpu hours to simulate motion for 100 synchrotron oscillations (10^5 turns). Use of tracking simulations for studies of modern large accelerators therefore consumes substantial amounts of cpu time. In fact, most tracking studies are cpu time-limited; investigators perform studies for just as long (but only as long) as their computing budget allows! The utility of a high speed program thus becomes clear.

Large memory is also needed, both to store the many parameters (such as families of multipoles) defining each of the $\sim 10^5$ elements comprising a large accelerator, and to allow use of "double precision". Extended precision (at least 64 bits) is required to avoid accumulation of roundoff errors; such errors would destroy the accuracy of a computation after only a few thousand turns[62].

Both of these requirements imply a need for considerable access to large computers or supercomputers. The availability of such high speed computing power, as well as a desire for even higher speed computation at ever lower cost, has led many to consider the use of novel computation archetectures. The most readily available novel archetecture is that of the vector or "pipeline" processor. This is the structure implemented on large supercomputers such as the CYBER-205 or the CRAY family. These machines are well suited to Lie transform and generating function methods, as these models involve processing of relatively

large numerical arrays (a few hundred elements long). A. Dragt and associates have achieved factors of ten speed increases using vector processing capabilities on a CRAY[63]. For example, the program MARYLIE, executing on a CDC-7600, requires approximately 3 millisecond for a single ray trace; executing on a CRAY-1S, a vectorised version of the same code requires about 0.3 millisecond. Similarly, a version of the same program, heavily optimised to make use of the four-instruction pipeline and floating point accelerator of a VAX-8600 will perform a single ray trace in approximately 4 to 4.5 milliseconds, even though the basic VAX-8600 processor is about a factor of 8 slower than the CDC.

A second archetecture that has been investigated for use in tracking studies involves the use of multiple processors. In this arrangement, several cpu's are used to do tracking; for example, one processor is assigned to each particle and several particles are tracked simultaneously. In this manner, it is possible to achieve a speed increase by increasing the number of cpu's employed. One example of this archetecture was generated at the Lawerence Berkeley Laboratory's now defunct Advanced Computer Archetecture Laboratory, where six DECNETed VAX cpu's were used to run the code MARYLIE, with multiple simultanious ray traces and load-balancing after each ray trace[64]. The single ray trace speed on one VAX cpu, after program optimisation, was 9 milliseconds; simultaniously running on all six cpu's, the single ray trace speed was 9 milliseconds for six particles. This is a factor of two faster than the CDC-7600 speed, where serial processing of six particles, at a timing of 3 milliseconds per ray trace, requires 18 milliseconds. The VAX system is therefore faster, but only when tracking of multiple particles is desired.

The requirement of multiple particles is a general one made to insure that special purpose "tracking engines" are computationally efficient. Most schemes using multiple computers or micro-processors invoke the use of many (tens, hundreds, or even thousands) particles in order to compensate the inherently slow single ray trace of the small cpu's involved. The main requirement for special purpose archetectures designed for tracking is that the number of particles be optimal and the speed high enough for efficient numerical studies, while the main constraint is that the cost (both hardware and software) be low enough to make the choice of a special process machine reasonable, given the availablity of larger, higher speed general purpose mainframes.

C. Experimental Tests of Tracking Codes

Given the substantial effort expended in generating descriptions of charged particle motion in an accelerator, one is driven to ask the obvious question, namely, do the results these models provide match any behavior observed in an actual machine? The answer to this question is, as expected in the real world, an unequivocal maybe. Historically, the performance of an accelerator most typically was limited by collective effects or the beam-beam interaction. It has been only in the most recent generation of machines that single particle

dynamics has arisen as a major issue in machine performance. Thus, only recently have experimental programs directed toward tests of single particle effects been initiated.

Certain statements may be made. The theories describing nonlinear motion in accelerators do provide qualitatively correct descriptions of machine behavior even for the highly nonlinear beam-beam interaction. The observed performance of a variety of machines admits convincing explanations using many of the tools that have been outlined above[65]. Furthermore, considerable effort has been made to corrolate models with the observed behavior of accelerators. Methods for direct measurement of the dynamic aperture of accelerators have been developed[66]. In essence, these allow a machine to act as an "analog computer" performing an ultra-high speed tracking experiment; the results are presented in much the same manner as are the results from a tracking experiment. These methods are presently being applied to existing machines, with somewhat mixed agreement with theory. In one example, an application to SPEAR did not agree well with modeling[66]; in a second, TEVATRON studies seemed to corollate well with theoretical understanding[67]. It is anticipated that additional machine studies will assist in our understanding of the theoretical tools and provide a basis on which increasingly accurate predictions of machine performance may be confidently made.

D. Analysis and Interpretation of Tracking Results: Moving From the Surface of Section to the Torus

Advances in interpreting dynamic aperture computations must come not only from direct tests of the fidelity of such computations to observation but also from a deeper understanding of what the computations represent (so as to know what to observe!). In the above discussion, we limited consideration of the dynamical system's time-evolving trajectory in state space to "snapshots" at particular times. Specifically, we looked only at the intersection of the trajectory of the system point with a surface of section. However, the surface of section is simply a "slice" of the motion; in well behaved dynamical systems, particles actually move on a multidimensional torus. Our phase space plots, generated during tracking, are projections of the motion into various subspaces of the phase space, made in an effort to determine if the entire torus exists, or if the motion is ill-behaved.

Recent developments in nonlinear dynamics have expanded the accelerator physicist's view of the problem of particle stability to include a more global outlook on the phase space. Various individual workers and groups are generating new diagnostics which, when applied to dynamic aperture computations of the type discussed here, provide new insight into the behavior of particles in accelerators[68]. An extensive body of knowledge is developing that will provide future generations of accelerator physicist with those tools necessary to design and build the instruments to be used on the forefronts of scientific research.

VII. Acknowledgements

I would like to thank Drs. Al Garren, Max Cornacchia, L. Jackson Laslett, and Roger Servranckx for many illuminating discussions on the topics presented here. I would also like to thank Dr. Mel Month for continual encouragement and boundless patience. This work was supported by the United States Department of Energy, in part by contract numbers DE-AC03-76SF00098 and DE-AC05-84ER40150.

VIII. References and Notes

1. Theorem and proof given by A. Dragt, in **Physics of High Energy Particle Accelerators** (Fermilab Summer School, 1981), A. I. P. Conf. Proc. 87, R. A. Carrigan, F. R. Huson, and M. Month, editors (1982).
2. H. Goldstein, **Classical Mechanics** (Addison-Wesley, Reading, MA, 1950).
3. See, *e.g.*, V. I. Arnol'd, **Mathematical Methods of Classical Mechanics** (Springer-Verlag, New York, 1978); E. T. Whittaker, **Analytical Dynamics** (Dover, New York, 1944); A. J. Lichtenberg and M. A. Liebermann **Regular and Stochastic Motion** (Springer-Verlag, New York, 1983); or H. Goldstein, *op. cit.* (Ref. 2).
4. A. Dragt, *op. cit.* (Ref. 1), or D. Douglas, "Lie Algebraic Methods for Particle Accelerator Theory", Ph.D. thesis, U. Maryland (unpublished, 1982).
5. The reader is referred to any of the proceedings of the High Energy Particle Accelerator Summer Schools (A. I. P. Conf. Proc. 87, 105, and 127), especially those articles on linear beam transport theory.
6. E. D. Courant and H. S. Snyder, **Ann. Phys. 3**, 1 (1958).
7. See, *e.g.*, the appendix to "Introduction to Accelerator Theory" by E. D. Courant *et al.*, p. 62 of **Physics of High Energy Particle Accelerators**, A. I. P. Conf. Proc. 87, R. A. Carrigan, F. R. Huson, and M. Month, editors (1982).
8. L. J. Laslett, **Proc. 1974 Int. Conf. on High Energy Accelerators**, p.394 (Stanford, 1974); L. J. Laslett in **Nonlinear Dynamics Aspects of Particle Accelerators**, J. M. Jowett, M. Month, and S. Turner, editors, (Springer-Verlag, New York, 1985); L. J. Laslett, **Particle Accelerators 19**, No. 1–4, 1–8 (1986).
9. The reader is referred to the literature for discussions of features of single particle motion specific to electrons. Two standard introductions are M. Sands, "The Physics of Electron Storage Rings — An Introduction", SLAC Report SLAC-121 UC-28 (ACC) (Nov. 1970), and A. Renieri, "Problems in Single-Particle Dynamics Specific to Electrons", in **Theoretical Aspects of the Behavior of Beams in Accelerators and Storage Rings**, CERN 77-13 (19 July 1977).

10. See, for example, H. Winick and S. Doniach, **Synchotron Radiation Research** (Plenum, New York, 1980), Chap. 2; J. N. Galayda, **I.E.E.E. Trans. Nuc. Sci. NS-30**, No. 4, 3109 (1983), J. Bisognano *et al.*, **Particle Accelerators 18**, No. 4, 223 (1986); or "National Center for Advanced Materials Conceptual Design Report", Lawrence Berkeley Laboratory publication PUB-5084 (March 1983).

11. For an example comparing specific machine lattice designs, see A. Jackson, "A Comparison of the Chasman-Green and Triple Bend Achromat Lattices", Lawrence Berkeley Laboratory Preprint LBL-21279 (March 1986; submitted to **Particle Accelerators**).

12. For state of the art designs of small emittance synchrotron radiation rings, see G. Vignola *et al.*, **I.E.E.E. Trans. Nuc. Sci. NS-32**, No. 5, 3391 (1985), or A. Jackson, *op. cit.* (Ref. 11). See the example in section V.B. for a conceptual discussion of a scheme using sextupoles to correct dynamic aperture limitations.

13. H. Wiedemann, "Chromaticity Correction in Large Storage Rings, PEP-220 (Sept. 1976); M. H. R. Donald, P. L. Morton, and H. Wiedemann, **I.E.E.E. Trans. Nuc. Sci. NS-24**, No. 3, 1200 (1977); M. H. R. Donald, "Chromaticity Correction in Circular Accelerators and Storage Rings", PEP Note 311 (July 1979); R. Brinkmann and F. Willeke, "Chromatic Corrections and Dynamic Aperture in the HERA Electron Ring", DESY 86-079 (July 1986); G. Guignard and J. Hagel, **Particle Accelerators 18**, No. 3, 129 (1986); P. Morton, "Derivation of Nonlinear Chromaticity by Higher-order 'Smooth Approximation' ", PEP-221 (August 1976); a number of the papers in the "Proceedings of the Workshop on Accelerator Orbit and Particle Tracking Programs", Brookhaven National Laboratory report BNL 31761 (August 1986) address this and related issues.

14. "Conceptual Design of the Superconducting Super Collider", SSC Central Design Group Report SSC-SR-2020 (March 1986).

15. D. Douglas, in "Accelerator Physics Issues for a Superconducting Super Collider", M. Tigner, editor, U. Michigan Report UM HE 84-1 (Dec. 1983).

16. Report of the Reference Designs Study Group on the Superconducting Super Collider (May 8, 1984).

17. See, e.g. L. Michelotti, **Proc. 1984 High Energy Particle Accelerator Summer School** (Fermilab), for a thorough pedagogical treatment of resonance problems.

18. G. Guignard, "The General Theory of All Sum and Difference Resonances in a Three-Dimensional Magnetic Field in a Synchrotron, CERN 76-06 (23 March 1976), and G. Guignard, "A General Treatment of Resonances in Accelerators" CERN 78-11 (10 November 1978).

19. E. D. Courant, "Periodic Passage Through Nonlinear Resonances During Acceleration in ISABELLE", BNL ISABELLE Tech. Note No. 163 (Feb. 12, 1980).

20. L. R. Evans and J. Gareyte, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2234 (1985), L. R. Evans, A. Gaugier, R. Schmidt, *Ibid.* 2209.

21. M. Cornacchia and L. Evans, **Particle Accelerators 19**, No. 1–4, 125 (1986).

22. T. Collins, " Distortion Functions", FNAL Tech. Note Fermilab-84/114 (Oct. 23, 1984).

23. F. Willeke, in "Accelerator Physics Issues for a Superconducting Super Collider", M. Tigner, editor, U. Michigan Report UM HE 84-1 (Dec. 1983).

24. A. Ando, **Proc. 12^{th} Int. Conf. on High-Energy Accelerators** (Fermilab, 1983).

25. R. D. Ruth and W. T. Weng, in **Physics of High Energy Particle Accelerators** (Fermilab Summer School, 1981), A. I. P. Conf. Proc. 87, R. A. Carrigan, F. R. Huson, and M. Month, editors (1982).

26. M. H. R. Donald, "Chromaticity Correction in Circular Accelerators and Storage Rings, Part I. A User's Guide to the HARMON Programs", PEP Note 311 (July 1979).

27. A. J. Dragt, "Nonlinear Lattice Functions", in **Proc. 1984 Summer Study on the Design and Utilization of the Superconducting Super Collider**, R. Donaldson and J. G. Morfín, editors (Snowmass, 1984).

28. E. Forest, "Lie Algebraic Methods for Charged Particle Beams and Light Optics", Ph.D. thesis, U. Maryland (unpublished, 1984), and "Normal Form Algorithm on Nonlinear Symplectic Maps", Superconducting Super Collider Central Design Group Report SSC-Report-29.

29. L. Michelotti, **Particle Accelerators 16**, 233-252 (1985), and **Particle Accelerators 19**, 205-210 (1986).

30. A. Deprit, **Celestial Mechanics 1**, 12 (1969).

31. A. J. Dragt *et al.*, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2311 (1985).

32. At the time of this writing, methods of incorporating higher-order effects (both in strength of perturbation and multipole order of perturbation) have been developed and are being incorporated into MARYLIE and tested (A. Dragt, private communication, Sept. 1986).

33. R. Servranckx, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2186(1985); D. Douglas, **Particle Accelerators 19**, No. 1-4, 119 (1986).

34. B. Leemann, E. Forest, and D. Douglas, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2300 (1985).

35. H. Wiedemann, "Chromaticity Correction in Large Storage Rings, PEP-220 (Sept. 1976).

36. A. Wrülich, "RACETRACK, A Computer Code for the Simulation of Nonlinear Particle Motion in Accelerators", DESY 84-026 (March 1984).

37. A. D. Russell, "TEVLAT", unpublished. An example application of the code is given by N. M. Gelfand, in "Accelerator Physics Issues for a Superconducting Super Collider", M. Tigner, editor, U. Michigan Report UM HE 84-1 (Dec. 1983).

38. L. Schachinger and R. Talman, "TEAPOT, A Thin Element Accelerator Program for Optics and Tracking", SSC Central Design Group Note SSC-52 (Dec. 1985).

39. R. Ruth "A Canonical Integration Technique (A Symplectic Map)", Lawrence Berkeley Laboratory note LBL-14770 (Aug. 1982), and **I.E.E.E Trans. Nuc. Sci. NS-30**, No. 4, 2669 (1983).

40. F. Ch. Iselin, **Proc. 12^{th} Int. Conf. on High-Energy Accelerators** (Fermilab, 1983).

41. L. J. Laslett, private communication (July 1984).

42. R. Servranckx *et al.* "User's Guide to the Program DIMAD", SLAC Report 285 UC-28 (A) (May, 1985).

43. D. Douglas, E. Forest, and R. Servranckx, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2279 (1985).

44. Third-order expansions have now or soon will become available in a new version of TRANSPORT (F. Rothacker, private communication, June 1986).

45. K. L. Brown, "A First- and Second-Order Matrix Theory for the Design of Beam Transport Systems and Charged Particle Spectrometers", SLAC Report 75. For a pedigogical treatement, see K. L. Brown and R. V. Servranckx, in **Physics of High Energy Particle Accelerators** (BNL/SUNY Summer School, 1983), A. I. P. Conf. Proc. 127, M. Month, P. F. Dahl, and M. Dienes, editors (1985).

46. Report of the Reference Designs Study Group on the Superconducting Super Collider (May 8, 1984), p. 358.

47. S. Peggs, **Particle Accelerators 17**, No. 1–2, 11 (1985).

48. G. F. Dell, private communication (July 1984).

49. Information for this example was provided by G. F. Dell, B. Leemann, and N. Gelfand through informal reports and private communications during the Superconducting Super Collider Reference Designs Study (1984); some of their work was documented by G. F. Dell, F. Willeke, and B. Leemann in **Proc. 1984 Summer Study on the Design and Utilization of the Superconducting Super Collider**, R. Donaldson and J. G. Morfín, editors (Snowmass, 1984).

50. G. F. Dell, "Tracking Results", informal note distributed during the SSC Reference Designs Study (April 13, 1984, unpublished), and G. F. Dell, F. Willeke, and B. Leemann in **Proc. 1984 Summer Study on the Design and Utilization of the Superconducting Super Collider**, R. Donaldson and J. G. Morfín, editors (Snowmass, 1984).

51. Report of the Reference Designs Study Group on the Superconducting Super Collider (May 8, 1984), p. 128.

52. N. Gelfand, private communication (4 June 1984).

53. A. D. Russell, *op. cit.* (Ref. 37).

54. B. T. Leemann, "Preliminary Summary of the SSC Tracking Studies (May - June 1984)" (June 11, 1984, unpublished).

55. A good starting point for a literature search would be the Superconducting Super Collider Conceptual Design Report, *op. cit.* (Ref. 46); the SSC Aperture Workshop Summary, SSC-TR-2001; and the Proceedings of the most recent Snowmass workshops on the design and utilization of the SSC.

56. For details of the design, see the "National Center for Advanced Materials Conceptual Design Report", Lawrence Berkeley Laboratory publication PUB-5084 (March, 1983); results to be discussed in this example have been supplied by A. Garren and H. Wiedemann, private communications (1983) and the next two references, and by the author (unpublished, 1983).

57. A. Garren, "ALS Storage Ring Design and Parameters", Lawrence Berkeley Laboratory NCAM Note-6, LBID-715 (May 10, 1983).

58. H. Wiedemann, "Some Comments on the ALS Design", June 14, 1983 (unpublished).

59. K. L. Brown, **I.E.E.E Trans. Nuc. Sci. NS-26**, No. 3, 3490(1979), R. V. Servranckx and K. L. Brown, *Ibid.* 3598.

60. L. M. Healy, "Lie Algebraic Methods for Treating Lattice Parameter Errors in Particle Accelerators", Ph.D. thesis, U. Maryland (unpublished, 1986).

61. R. D. Ruth, T. Raubenheimer, and R. L. Warnock, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2206 (1985); R. L. Warnock and R. D. Ruth, "Invariant Tori Through Direct Solution of the Hamilton-Jacobi Equation", SLAC-PUB-3865, LBL-21709 (Jan. 1986) (submitted to **Physica D**); R. L. Warnock and R. D. Ruth, "The Study of Invariant Surfaces and Their Break-Up by the Hamilton-Jacobi Method", SLAC-PUB-4044, LBL-21979 (Aug. 1986), presented at 13^{th} Int. Conf. on High Energy Accelerators, Novosibirsk, USSR (August 1986).

62. A. Wrülich, in **Nonlinear Dynamics Aspects of Particle Accelerators**, J. M. Jowett, M. Month, and S. Turner, editors (Springer-Verlag, New York, 1985); P. Wilhelm and E. Lohrmann, **Particle Accelerators 19**, No.1–4, 99 (1986).

63. A. Dragt, private communication (April 1985).

64. W. Greiman, private communication (July 1984).

65. See, *e.g.*, L. Evans, *et al.*, *op. cit.* (Ref. 20, 21), S. Peggs, *op. cit.* (Ref. 47), or S. Peggs, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2249 (1985).

66. P. L. Morton *et. al.*, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2291 (1985).

67. D. A. Edwards, R. P. Johnson, and F. Willeke, **Particle Accelerators 19**, No.1–4, 145 (1986).

68. An extensive and rapidly growing literature exists in this field. We list a few sources with which the interested reader may begin an investigation of the topic. H. Mais, A. Wrülich, and F. Schmidt, **I.E.E.E Trans. Nuc. Sci. NS-32**, No. 5, 2252(1985) provide a discussion of long-term tracking in HERA, and present useful diagnostics. A useful means of distinguishing chaotic motion from quasi-period motion is given by J. M. Jowett in "A Method for Distinguishing Chaotic From Quasi-Periodic Motions in Orbit Tracking Programs", CERN LEP-TH/83-51, and applied by M. E. Biagini, S. Guiducci and A. Renieri in "An Application of a Method for Distinguishing Chaotic From Quasi-Periodic Motions to the ESRP Lattices", European Synchrotron Radiation Project internal report ESRP-IRM-8/83. The first of these was presented at the Europhysics Conference on Computing in Accelerator Design and Operation, Berlin, Sept. 1983; the proceedings of this conference are available in **Computing in Accelerator Design and Operation**, W. Busse and R. Zelazny, editors, Lecture Notes in Physics No. 215 (Springer-Verlag, New York, 1984). Several recent conferences and workshops have been devoted to this subject. In addition to the proceedings just mentioned, the following two sources may prove useful to the interested reader: **Nonlinear Dynamics Aspects of Particle Accelerators** (Proceedings, Sardinia 1985), Lecture Notes in Physics Number 247, J. M. Jowett, M. Month, and S. Turner, editors (Springer-Verlag, New York, 1986); "Proceedings of the Workshop on Orbital Dynamics and Applications to Accelerators", **Particle Accelerators 19**, No. 1–4, 1 (1986).