Maps for distributions and their time evolution

Étienne Forest
High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan

David S. Robin
Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720

(Received 10 February 1999)

Many dynamical stochastic processes occur ''on top'' of a deterministic process. We present a method which uses the trajectory of the deterministic process as basis functions for quasiarbitrary distributions. A map for the stochastic process can then be computed. This may have applications in electron storage rings or other devices perturbed by a small stochasticity. In this paper we will look only at the most elementary applications of the method. [S1063-651X(99)05310-6]

PACS number(s): 29.20.−c, 02.50.Fz, 02.50.Ng

I. INTRODUCTION

Predicting the time evolution of a distribution, even when all the mathematical properties of the system are known, can be a very difficult task. In this paper we present a method which assumes that the stochastic map is ''near'' the symplectic map and thus uses the invariants of the symplectic map as support functions for a quasiarbitrary distribution.

We limit the core of the discussion to one degree of freedom. We do this for two reasons. First it is simpler to introduce a new method in a simpler setting and, in this setting, it is even beyond the scope of this introductory paper to address all the potential applications.

The second reason is more problematic: we are not capable at this point of extending the method ''correctly'' in a multidimensional setting. Of course in accelerators this might be the most important case. We hope the reader will nevertheless get useful ideas from this paper, if not for ring dynamics, maybe for other fields.

We acknowledge at the onset that this method was first proposed by Gerasimov [1] to Pauluhn and Mais at DESY and that Pauluhn [2] did a preliminary testing in her thesis. This was fortunately unknown to us when we started the work because it might have discouraged us from going any further.

II. THEORETICAL DESCRIPTION OF THE METHOD

To introduce this technique it is best to limit ourselves first to one-degree-of-freedom processes, that is to say, to two-dimensional systems. We will also present a true one-dimensional example to illustrate the relative merits of the brute force, Irwin [3], and map methods of this paper.

One has to be honest here: although some two-degree-of-freedom systems can be handled easily by this method, in general there are severe complications. This is *not* the case of the brute force or Irwin's method, which extend to higher dimensions rather easily.

Furthermore, we will assume that the stochastic process under consideration does not depend on the distribution itself. This invalidates processes involving intrabeam scattering such as Touschek although they can be included in the method.

A. The map for the distribution

First consider an initial distribution f_0 in the phase space variables $\mathbf{z}=(x,p)$. Let us assume that, in the absence of any stochastic process, it evolves under the action of a symplectic one-turn map $\mathbf{m}$:

$$\text{if } \mathbf{z}_1=\mathbf{m}(\mathbf{z}_0)$$

$$\text{then } \forall \mathbf{z}\ f_1(\mathbf{z})=f_0(\mathbf{m}^{-1}(\mathbf{z})). \tag{1}$$

Since we will deal explicitly with functions, it is worth noticing that the action of the map $\mathbf{m}$ on the function f is equivalent, in the symplectic case, to the action of the symplectic map $\mathcal{M}^{-1}$ as defined by Dragt (see Ref. [4] for explanations),

$$f_1=f_0\circ\mathbf{m}^{-1}=\mathcal{M}^{-1}f_0. \tag{2}$$

In the literature the map $\mathcal{M}$ is called the ''Koopman'' operator. Furthermore, the process of finding a basis for $\mathcal{M}$ which is then usually truncated is called ''Koopmanization.'' Standard perturbation theory whether expressed in terms of Lie operators or other terms amounts to representing the map $\mathcal{M}$ in a monomial basis truncated at certain degree. Hamiltonian perturbation theory does the same thing on the generator of the map $\mathcal{M}$, the Hamiltonian of the system. The nonperturbative methods of Warnock *et al.* also amount to representing $\mathcal{M}$ in a special basis more suitable to problems near the short term dynamic aperture (very nonlinear). In the case of perturbation theory, the truncation of the matrix representing $\mathcal{M}$ leads in general to an asymptotic series as one tries to find invariants of the motion, that is to say, eigenvectors of unity. For these reasons the choice of the basis for the expansion of $\mathcal{M}$ is critical to the search of well-behaved quasi-invariants.

The operator which transforms a distribution is $\mathcal{M}^{-1}$ in the symplectic case. In the case of an arbitrary map, it is called the Frobenius-Perron [5] operator (FP) and it acts in a

 © 1999 The American Physical Society

space dual to the Koopman operator. It should be said that an expansion of the Koopman $\mathcal{M}$ in terms of monomials is equivalent to a representation of the (FP) operator in terms of moments. It is well known in accelerator physics that a moment expansion is notoriously inadequate for the study of lifetime or even equilibrium issues. Therefore in this paper, in effect, we will propose a better expansion set for the FP operator in the presence of stochasticity and small damping.

Thus, let us imagine that the ray $\mathbf{z}$ is modified by a stochastic variable Δ,

$$\overline{\mathbf{z}} = \mathbf{z} + \Delta, \tag{3}$$

and that Δ is governed by a distribution $\rho(\Delta, \mathbf{z})$. Then the distribution after one turn can be written as

$$\mathcal{U} f_0(\mathbf{z}) = f_1(\mathbf{z}) = \int f_0(\mathbf{m}^{-1}[\mathbf{z}\Delta])\rho(\Delta; \mathbf{z}\Delta) d\Delta. \tag{4}$$

This map is exact so far. It should be noted that, if the stochastic process happens all along the ring, one can rigorously integrate (or sum) along longitudinal positions where stochastic kicks occur:

$$\mathcal{U} f_0(\mathbf{z}) = f_1(\mathbf{z}) = \oint_{s=0}^{C} \int f_0(\mathbf{m}_{0,s}^{-1}[\mathbf{z}\Delta])\rho_s(\Delta; \mathbf{z}\Delta) d\Delta\, ds. \tag{5}$$

In this paper we will use Eq. (4) without loss of generality because the function ρ can represent the integrated effect of ρ_s over one turn or more.

B. The expansion set for the map

This is very nice and completely useless so far. The FP map is truly a monstrous integro-differential equation; so what can we do? Traditionally, it is customary to convert this equation into a differential equation neglecting higher order moments of the random variable Δ. The resulting equation is the Fokker-Planck equation. Indeed, if we include first and second order moments of Δ, we will get a differential equation for the function f or equivalently for the FP operator $\mathcal{U}$. The Fokker-Planck equation will look like the Vlasov equation except for additional terms proportional to second partial derivatives in the distribution. These terms will lead to diffusion.

The Fokker-Planck equation is neither simple [6] to set up nor to solve in the general case. However, the dimensionality of the problem can be further reduced if one assumes that the final distribution will sit on the invariants [7] of the symplectic map. This is done, formally, by introducing a new set of variables (ϕ, J) to replace the original vector $\mathbf{z} = (\mathbf{x}, \mathbf{p})$. It is assumed that the set (ϕ, J), consisting of one angle and one action, totally normalizes the map $\mathbf{m}$ or equivalently the Hamiltonian which gave rise to this map. Thus we have

$$\mathcal{M} J = J \circ \mathbf{m} = J. \tag{6}$$

Moreover, if we denote by $\mathbf{a}_s$ the canonical transformation connecting (x,p) to (ϕ, J) at position ''s,'' then the resulting map $\mathbf{n}_{0,s}$ in the new variables has the form

$$\text{if } \mathbf{n}_{0,s} = \mathbf{a}_s^{-1} \circ \mathbf{m}_{0,s} \circ \mathbf{a}_0$$

$$\text{and } (\phi, J)_s = (a_{s;1}(x,p), a_{s;2}(x,p))$$

$$\text{then } \begin{cases} n_{0,s;1}(\phi, J) = \phi + \mu_{0,s}(J) \\ n_{0,s;2}(\phi, J) = J \end{cases}. \tag{7}$$

This map propagates a normalized ray from an arbitrary position $s = 0$ to a final position s by advancing the phase ϕ an amount $\mu_{0,s}(J)$. Obviously if ''s'' is a continuous length along some reference curve, this will induce changes in the Hamiltonian or the associated Fokker-Planck equation. Indeed that Hamiltonian part of the Fokker-Planck equation (Vlasov) will leave invariant the function of J. If we further average over the canonical angle ϕ, then the resulting Fokker-Planck equation depends on J only (see Sec. IV C 3). Obviously, by construction, the resulting distribution will depend on the invariant function $J(x,p)$ only. This is justified when the average phase advance coming from the stochastic term Δ is small compared to the tune μ. Generally this means that the damping must be small compared to the tune.

There are three problems associated with this formulation. First of all, as usual, we still have to solve a messy differential equation while our most reliable model for the symplectic (and nonsymplectic) dynamics comes in the form of a tracking code. Hence just writing down an equation for (x,p) is already a problem of major proportions. Secondly we must still find the invariant coordinates (ϕ, J); this is no small task especially if they are to be fed into a differential equation in s, the longitudinal position around the ring. Thirdly, and most fundamentally, in the presence of islands the variables (ϕ, J) are, at best, only locally defined. While we can check by tracking a large number of rays (or a ray for a very long time) that the equilibrium distribution will sit very nearly on invariant trajectories, we cannot find a canonical transformation $\mathbf{a}$ that will work over the entire phase space. Within each island (and within tiny islands inside the islands) one could get an approximate canonical transformation $\mathbf{a}$, but these transformations cannot be smoothly interpolated. This means that any method based on a differential equation that tries to take advantage of the local existence of an invariant J will not work.

In the case of a differential equation, it is important to have a grid which is indexed in a continuous manner. For example, if the invariant J represents concentric curves around the origin, then a discretization may work as follows: first we decide that the largest action will be $\hat{J}$. This value is picked by assuming that the distribution is almost null at this large amplitude. Secondly, we decide to slice phase space into N orbits for the purpose of discretizing the differential equation. The concentric curves will be labeled by an integer i ranging from 1 to N. For example, J_i, the ith trajectory, can be given by $i\hat{J}/N$. In this case, the derivatives with respect to J that are needed in the solution of the Fokker-Planck equation will be obtained by taking divided differences. In the case of an integral equation, we will see that this is not necessary. We are free to label values of J used in the discretization scheme in a completely arbitrary manner.

Figure 1 displays an example with islands. A continuous indexing scheme is not possible for the same reasons that a

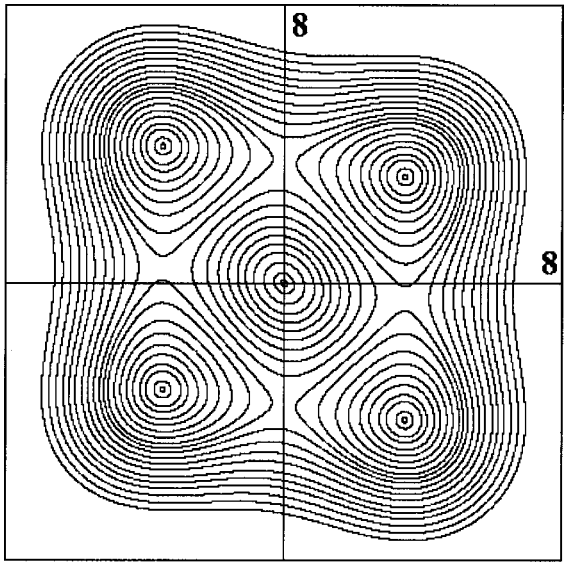

FIG. 1. Support grid with three regions.

globally defined J is topologically impossible. The phase space is divided into three regions: below the islands, in the islands, and above the islands. The trajectories are labeled as follows: $i=1,10$ are below the islands, $i=11,20$ in the islands, and finally $i=21,30$ are the trajectories above the island chain. It is important to realize that this labeling is now totally arbitrary. One could randomly scramble the relationship between the index i and the topological features (trajectory in this case) on the plot.

Now we are in a position to describe a distribution whose contours are the symplectic trajectories. Let us again denote by N the number of trajectories plotted. Then an arbitrary distribution f can be projected onto this contour by a yet undefined projection operator π:

$$\pi(f)=\sum_{i=1}^{N} v_i J_i . \tag{8}$$

The quantities J_i represent unit distributions concentrated on the ith trajectory. (They are Dirac δ functions in the variable J.) In other words, the distribution $f=1000J_4$ is a distribution of 1000 particles evenly distributed in the local phase ϕ on the fourth trajectory. Thus a distribution of 10 000 particles with 1000 on the first trajectory, 8000 on the second, and 1000 on the fourth would be given by

$$1000\ J_1+8000\ J_2+1000\ J_4 \tag{9}$$

or, equivalently by the component vector $\mathbf{v}=(1000,8000,0,1000,0,0,\ldots,0)$. Again we must repeat that, while it is convenient that the index i progresses more or less monotonically as we go away from the origin, it is not necessary.

Finally we must say a few words about the projection π. If we take any distribution f that depends only on J, then due to the coarseness of the grid, it will not be possible to represent it exactly. The best we can do is to project f by some approximate scheme onto the grid $\{J_i|i=1,N\}$. This projection is done with a prescription denoted by π. Since an arbitrary distribution can be written in terms of sharply peaked (Dirac δ functions) distributions, we only need to define π on an individual phase space point. Thus if a particle lands at coordinates $\mathbf{z}_0$ between the trajectories J_i and J_k, we will then assume that this ray contributes to both J_i and J_k using a particle number preserving algorithm. For example, suppose we determine that the ''distance'' between the trajectories J_i and J_k is d, then a ray (Dirac δ distribution) falling a ''distance'' d_i from J_i will result in a new distribution

$$\pi(\delta(\mathbf{z}-\mathbf{z}_0))=\frac{[d-d_i]J_i+d_iJ_k}{d}. \tag{10}$$

The details of the projection π will turn out to be quite irrelevant. If the reader needs a concrete model, he can imagine that in the case of the J_i's being concentric circles, the distance between them could be measured along the radial direction.

In summary, we have a rather coarse phase plot as our support grid and a projection operator π which allows us to project a δ function distribution on the neighboring trajectories.

C. The map restricted to the grid

It is instructive to make an analogy with quantum mechanics. The symplectic map divides the phase space into energy levels. Indeed the Bohr-Sommerfeld quantization rules would involve precisely the action J. However, this J is only locally defined, thus we need several ''quantum'' numbers: first we need an integer describing the region of phase space we are sitting in. In the example of Fig. 1, the index takes the values 1, 2, and 3. Secondly, within each region, a second index ranges from 1 to 10 in the example of Fig. 1. Thus we could label our support set $\{J_i|i=1,N\}$ by an equivalent set $\{J_{kl}|k=1,\hat{k};\ l=1,l_k\}$. In our example, $\hat{k}$ is 3 and l_k is 10 for all k's. This new set is useful when discussing connection with the continuous case. The index k specifies a canonical transformation $\mathbf{a}_k$ which locally assigns a numerical meaning to the action variable. If we follow this analogy further, we can translate the entire computation of the map into a quantum equivalent. Our goal will be to compute the probability that a state J_{kl} jumps to a new state $J_{k'l'}$. The probability of the jump is governed by the function $\rho(\Delta;\mathbf{z})$ or, more appropriately, by the projection of this function onto a space consisting of J-dependent distributions only.

Physically this is not a bad approximation even far from equilibrium. Suppose we inject a beam off axis, then two things will happen. The beam, due to nonlinear shear (filamentation), will evolve rapidly towards a J-dependent distribution. However, even in the linear case when filamentation is null, the beam will visit the entire length of the trajectory due to the phase advance. This means that the ergodic average over a few turns will mimic the effect of an evenly distributed density in the phase ϕ associated to the local action J. Thus two effects rapidly combine so as to vindicate our choice for the grid.

We now proceed to restrict the map $\mathcal{U}$ on the grid using our quantumlike notation. This will temporarily permit us to retain a connection with the continuous case. It will be

dropped later. We start with an arbitrary distribution on the grid as in Eq. (8),

$$f_0(J,j)=\sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k} v_{kl}\delta(J-J_{kl})\,\delta_{jk}. \tag{11}$$

In this expression, the variable J is a true action in the jth region of phase space. Thus the function $\delta(J-J_{kl})$ is a Dirac delta function while δ_{jk} is a Kronecker delta. We substitute this result into Eq. (4) for the operator $\mathcal{U}$:

$$\mathcal{U}f_0(\phi,J,j)=\int \sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k} v_{kl}\delta(J-J_{kl} -\Delta J)\,\delta_{j-\Delta jk}\rho(\Delta\phi,\Delta J,\Delta j;\phi-\Delta\phi,J\Delta J,j -\Delta j)\,d\Delta. \tag{12}$$

We can now sum/integrate in the variables $(\Delta J,\Delta j)$,

$$\mathcal{U}f_0(\phi,J,j)=\sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k}\int v_{kl}\rho(\Delta\phi,J-J_{kl},j-k;\phi -\Delta\phi,J_{kl},k)\,d\Delta\phi. \tag{13}$$

Let us change the integration variable from $\Delta\phi$ to a quantity $\phi_{kl}=\phi-\Delta\phi$:

$$\mathcal{U}f_0(\phi,J,j)=\sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k} v_{kl}\int \rho(\phi-\phi_{kl},J-J_{kl},j -k;\phi_{kl},J_{kl},k)\,d\phi_{kl}. \tag{14}$$

Finally, under the assumption of this paper, a distribution which is uniform along a trajectory stays uniform. Thus we must have approximately

$$\frac{\partial}{\partial\phi}\int \rho(\phi-\phi_{kl},J-J_{kl},j-k;\phi_{kl},J_{kl},k)\,d\phi_{kl}\approx 0, \tag{15}$$

$$\int \rho(\phi-\phi_{kl},J-J_{kl},j-k;\phi_{kl},J_{kl},k)\,d\phi_{kl}\approx U(J,j;J_{kl},k).$$

We substitute the operator U in Eq. (14),

$$\mathcal{U}f_0(\phi,J,j)=\sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k} U(J,j;J_{kl},k)\,v_{kl}. \tag{16}$$

Equation (16) is extremely simple to interpret: The kernel element $U(J,j;J_{kl},k)$ connects a discrete distribution on the support grid denoted here by the double index (J_{kl},k) to the potentially continuous set of trajectories (J,j). In reality we must project this information on the grid and this is where the projection π comes into play. The result

$$\pi(U(J,j;J_{kl},k))=\sum_{\kappa=1}^{\hat{k}}\sum_{\lambda=1}^{\lambda_\kappa} U_{\kappa\lambda;kl}\delta(J-J_{\kappa\lambda})\,\delta_{j\kappa} \tag{17}$$

is now substituted into Eq. (16),

$$\pi\mathcal{U}f_0(\phi,J,j)=\sum_{k=1}^{\hat{k}}\sum_{l=1}^{l_k}\sum_{\kappa=1}^{\hat{k}}\sum_{\lambda=1}^{\lambda_\kappa} U_{\kappa\lambda;kl}\delta(J-J_{\kappa\lambda})\,\delta_{j\kappa}v_{kl}. \tag{18}$$

At this stage the pretense of a continuous theory can be dropped and we can replace the double indices kl and $\kappa\lambda$ by single indices a and b. Thanks to the projection π, the map $\mathcal{U}$ has been restricted to the grid, and we can extract a transfer matrix U_{ab} for the map:

$$w_a=\sum_{b=1}^{N} U_{ab}v_b \quad \text{or} \quad \mathbf{w}=U\mathbf{v},$$

$$\text{where } \mathbf{w}=\sum_{i=1}^{N} w_iJ_i \quad \text{and} \quad \mathbf{v}=\sum_{i=1}^{N} v_iJ_i. \tag{19}$$

D. Actual computation of U_{ab}

Let us start with the main process leading to an equilibrium distribution, namely, radiation. We will assume that a computer code exists which can provide us reliable simulations without radiation (symplectic), with classical radiation, and with Monte Carlo stochastic corrections to the radiation process. We further assume that this code has a graphical user interface (GUI) interface so that the user can select the grid and erase undesirable trajectories as he goes along.

The first challenge is to map out the symplectic trajectories on a relatively coarse grid and to select points on these trajectories which are equally spaced in phase. Suppose we are looking at the bth trajectory in an attempt to compute U_{ab}. This trajectory is started by clicking on an initial ray $\mathbf{z}^0$ on the screen. We keep tracking until the trajectory has finished painting an area on the screen. At this point, after T_b iterations, the trajectory is recurrent as far as the pixel resolution of the screen is concerned. It has completely filled the invariant torus on which it is assumed to sit. The total trajectory is given by a set $J_b=\{\mathbf{z}^i|i=0,T_b\}$ where iterates are listed consecutively. Now consider the subset $I_b=\{\mathbf{z}^{mk}|k=1,N_b;\,mk<T_b\}$ which represents N_b particles equally spaced in phase. Furthermore, by plotting these separately or using a very different color, we can ensure that they also populate the torus more or less evenly. The computation of the matrix element U_{ab} will use these particles. How does this work? Let us assume that the map U represents a q-turn operator. Furthermore, we now introduce the map $\mathbf{m}_\Sigma$, the stochastic extension of the symplectic map $\mathbf{m}$, and we build a new set $I_b^{q,M}$ using the subset I_b. At every point $\mathbf{z}\in I_b$ we launch the map $\mathbf{m}_\Sigma$ for q iterations and we repeat this stochastic experiment M times. We then obtain a new set $I_b^{q,M}$, the union of all our experiments,

$$I_b^{q,M} = \bigcup^M \left\{ \mathbf{z} \left| \mathbf{z} = \underbrace{(\mathbf{m}_\Sigma \circ \cdots \circ \mathbf{m}_\Sigma)}_{q \text{ times}} (\mathbf{z}^{mk}) \;\; ; \;\; k = 1, N_b \; ; \; mk < T_b \right. \right\}. \tag{20}$$

To evaluate U_{ab} we use the projection π on a single ray (Dirac δ distribution) as we hinted in the preceding section. We take a ray $\mathbf{z} \in I_b^{q,M}$ and project it on the grids using Eq. (10). In general, a ray will distribute itself onto (two) neighboring grid points as in Eq. (10):

$$\pi[\mathbf{z}] = \sum_{i=1}^{N} d_i^{mk} J_i, \quad \text{with} \quad \sum_{i=1}^{N} d_i^{mk} = 1. \tag{21}$$

We then sum all contributions from the set $I_b^{q,M}$:

$$U_{ab} = \frac{1}{M} \sum_{e=1}^{M} \frac{1}{N_b} \sum_{k=1}^{N_b} [d_{a;e}^{mk}]_{\text{results of } e^{\text{th}} \text{ experiment}}. \tag{22}$$

The process is repeated for all support curves so that b runs from 1 to N. In the example of Fig. 1 we would repeat this process 30 times.

Let us say a few words about the curve with the largest amplitude. Usually we can select a trajectory at a large amplitude which encompasses all the other support grid points. We can call this curve the Nth trajectory. A particle landing outside this trajectory poses a boundary condition problem. In this paper we opted to make this last boundary absorbing, that is to say, that rays falling outside are assigned to J_N.

III. SOME PROPERTIES OF THE MAP

In this section we examine some properties of the map. Rigorous statements will be made only for the map restricted to the grid. The reader will notice that many of the statements will apply to a general grid instead of a grid made of symplectic trajectories.

A. The existence of a fixed point

The map restricted to the grid must have a least one fixed point. We know that the particle number $Q(\mathbf{v})$ is preserved by construction,

$$Q(\mathbf{v}) = \sum_{i=1}^{N} v_i = Q(U\mathbf{v}). \tag{23}$$

In addition, we also know that U will preserve the positive nature of the component v_i's. Thus if we start with a physical distribution $v_i \geqslant 0$, not only will the particle number stay constant but so will the positive definite nature of the distribution. Thus the set of all distributions with a given number of particles Q_0,

$$\mathcal{D}(Q_0) = \left\{ \mathbf{v} \left| \mathbf{v} = \sum_{i=1}^{N} v_i \, ; v_i \geqslant 0 \right. \right\}, \tag{24}$$

is also preserved under the action of U.

It is easy to show that this set is a convex compact set. By Brouwer's fixed point [8] theorem the map U, provided it is continuous, must have a fixed point.

Finally we will (almost) prove later that the map is contracting; it will converge to the equilibrium distribution although it may not be unique. To illustrate this consider a strange process for which radiation is present only on the inner trajectories. For example, looking at Fig. 1, we assume that damping is present on trajectories 1–10 around the origin. We also assume that fluctuations are present from trajectory 1 to 9 with a zero probability of making a jump to the tenth. It is easy to construct such an imaginary system. It is clear that the inner particles would settle on some equilibrium orbit and that the outer orbits would be on the symplectic trajectories. Essentially the outside would be a proton beam with infinite degeneracy while the inside would be ''electron'' in behavior.

The reader must remember that the existence of an equilibrium orbit is proven here on the grid; in some cases the result is totally unacceptable. This usually will reflect the fact that the grid is no longer adequate. For example, if only diffusion is present, the beam will slowly crawl out until it bounces around the maximal grid trajectory J_N. This is unphysical; in the real system it keeps diffusing forever (well, not really...there has to be a beam pipe somewhere).

All of this implies that if we spend time proving theorems for our grid system, it may not be applicable to the real thing. Nevertheless we will give proofs of the various assertions since they teach us a lot about the nature of the system we constructed.

The existence of a fixed point has already been mentioned and is the result of Brouwer's fixed point theorem. This result extends to the case of a ρ-dependent $\mathcal{U}$, i.e., a nonlinear U. (We mentioned that the map $\mathcal{U}$ was dual to the map $\mathcal{M}$ in the symplectic case. One may ask how a matrix of infinite dimension can become ''nonlinear'' once more...of course this happens because the original problem had a greater dimensionality than first contemplated since the nonlinearities in U came from the interactions between the particles—intrabeam scattering.) Indeed if we add density-dependent diffusion and construct the projection π so as to preserve the particle number, the conditions for the application of the theorem will still be satisfied and thus there must be a fixed point. The next important question is whether the map is contracting. From now on the discussion focuses on the linear case.

B. Contraction: All eigenvalues are $\leqslant 1$

The issue is this: will an initial distribution settle down on an equilibrium or oscillate? First we prove that the eigenvalues have a modulus less than or equal to one. (One notes that $\lambda = -1$ prevents contraction from taking place. This was pointed out by Bob Warnock.)

Suppose that v is an arbitrary real vector. It can be written as

$$\mathbf{v}=\mathbf{v}_+-\mathbf{v}_-, \quad v_{+i}=\max(0,v_i), \quad v_{-i}=\max(0,-v_i). \tag{25}$$

The two vectors are orthogonal, $(\mathbf{v}_+,\mathbf{v}_-)=0$, and non-negative: $\mathbf{v}_+\geqslant 0,\mathbf{v}_-\geqslant 0$. Nonzero components of $\mathbf{v}_+$ correspond to zero components of $\mathbf{v}_-$ and vice versa. Also,

$$\sum_i (Uv)_i=\sum_i [(Uv_+)_i-(Uv_-)_i]=\sum_i v_i,$$

because U on non-negative vectors conserves particle number.

Now suppose that v is an eigenvector with eigenvalue λ. Then by the preceding result, $\Sigma_i v_i=\lambda\Sigma_i v_i$. Either (a) $\lambda=1$, or (b) $\Sigma_i v_i=0$. Let us pursue (b) in the case of real λ. The corresponding eigenvector may be assumed to be real, since real and imaginary parts, if linearly independent, would simply amount to two real eigenvectors with the same eigenvalue. Now we can write

$$Uv_+=v'_+ +\Delta v'_-,$$

where $v'_{+i}\neq 0$ if and only if $v_{+i}\neq 0$ and $\Delta v'_{-i}\neq 0$ if and only if $v_-i\neq 0$. If $v_i=0$ we arbitrarily set $\Delta v'_{-i}=0$, and notice that v'_{+i} may possibly be nonzero. That is, Uv_+ may have a nonzero component where neither v_+ nor v_- does. Now, since $Uv_+\geqslant 0$, we have $v'_+\geqslant 0$ and $\Delta v'_-\geqslant 0$. Of course, we also make the analogous breakup of Uv_-.

Now we have

$$\lambda(\mathbf{v}_+-\mathbf{v}_-)=(\mathbf{v}'_+-\Delta v'_+)-(\mathbf{v}'_--\Delta\mathbf{v}'_-). \tag{26}$$

We cannot, however, always draw the conclusion that

$$\lambda v_+=v'_+-\Delta v'_+$$

since, as mentioned above, v'_+ may have nonzero components in positions where v_+ has none. Such components must be cancelled, however, by corresponding components of v'_-, in order that Eq. (26) be satisfied. Let us then define $\hat{v}_{+i}=v'_{+i}$ if $v_i\neq 0$ and $\hat{v}_{+i}=0$ if $v_i=0$. Then we have

$$\lambda v_+=\hat{v}_+-\Delta v'_+,$$

and the corresponding equation for particle numbers,

$$\lambda N_+=\lambda N=\hat{N}_+-\Delta N_+.$$

But by the above we also have

$$N'_+ +\Delta N_-=N\geqslant\hat{N}_+ +\Delta N_-$$

hence by that

$$\lambda N\leqslant N-\Delta N_+-\Delta N_-\leqslant N. \tag{27}$$

Thus, we conclude that the modulus of the eigenvalues must be less than or equal to one.

Now it is obvious that the map will contract to the subspace spanned by the eigenvalues of unit modulus. Secondly, only the eigenvectors corresponding to $\lambda=1$ can be physical stationary distributions. Unfortunately it is not possible to reject the possibility that a linear combination of the other eigenvectors is present. In that case, the final distribution will oscillate in time so that its ergodic average will be the ''equilibrium'' distribution. Consider the following case:

$$U=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{28}$$

for which all the eigenvalues are on the unit circle. The eigenvectors are

$$\mathbf{v}_1=\left(1,\frac{-1+i\sqrt{3}}{2},\frac{-1-i\sqrt{3}}{2}\right),$$

$$\mathbf{v}_2=\left(1,\frac{-1-i\sqrt{3}}{2},\frac{-1+i\sqrt{3}}{2}\right), \tag{29}$$

$$\mathbf{v}_3=(1,1,1).$$

This map could represent a symplectic map with a third order island chain. Obviously if we start with one particle in island number 1,

$$\mathbf{v}_0=(1,0,0), \tag{30}$$

then it will jump to island 3 and then island 2 and finally back to 1.

$$U^n\mathbf{v}_0=\tfrac{1}{3}\{e^{i2\pi n/3}\mathbf{v}_1+e^{-i2\pi n/3}\mathbf{v}_2\}+(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}). \tag{31}$$

Therefore the ergodic average will be given by the component along the eigenvector of unit eigenvalue.

It should be pointed out that we will not see anything like that in our simulation since we purposely lump the islands together. Therefore we can safely assume (on physical grounds) that there are no eigenvectors of modulus 1 other than equilibrium distributions, in which case the map is contracting. The existence of $\lambda=-1$ eigenvectors can always be checked on the map itself.

C. Uniqueness of the equilibrium distribution: Connected grid

To prove the uniqueness of the equilibrium distribution, we must assume that the grid is connected. We say that the grid is fully connected by the map U if a distribution on the grid J_k will eventually diffuse to an arbitrary grid J_m after forward or backward propagation under the action of U. If we denote the nth power of U by U^n then we say that the map connects all of phase space if

$$\forall k\forall m\exists n\neq 0 \ \text{ such that } \ U^n_{mk}\neq 0. \tag{32}$$

The prime example of a totally disconnected grid is the original symplectic map described on the J_k's. In this case the map U is just the identity and thus there is no communication between the various support distributions.

Our goal is to prove that the equilibrium distribution is unique when the map U satisfies condition (32). First we prove that an equilibrium distribution $\mathbf{v}$ of the form given by

Eq. (25) must be such that $\mathbf{v}_+$ or $\mathbf{v}_-$ vanishes, i.e., the equilibrium is physical. We start by assuming that condition (32) is satisfied and thus there exists a power of U for which a part of the vector $\mathbf{v}_+$ gets mapped into $\mathbf{v}_-$ and/or vice versa. Thus, for this map, Eq. (27) applies with $\lambda=1$:

$$N \leqslant N-\Delta N_+ -\Delta N_- \leqslant N. \tag{33}$$

Since at least one of the vectors is assumed to make connection with the other (let us say $\mathbf{v}_+$), then it must be true that

$$\Delta N_+ > 0, \tag{34}$$

from which we can immediately conclude that

$$N \leqslant N-\Delta N_+ , \tag{35}$$

which is obviously a contradiction.

Thus, since we assumed that the map is connected, it follows that a fixed point must be physical (all positive components).

D. Uniqueness in the connected case

Uniqueness is proven by assuming the existence of two fixed points and constructing a situation contradicting what was just proven in the preceding section. We assume that there exist at least two equilibria $\mathbf{v}$ and $\mathbf{w}$.

1. Case $v\cdot w=0$

This leads to an immediate contradiction: it suffices to consider the new equilibrium $\mathbf{v}-\mathbf{w}$ which has exactly the form given by Eq. (25). We know that such a distribution cannot exist as proven in Sec. III C.

2. Case $v\cdot w\neq 0$

We construct a vector $\mathbf{g}$ perpendicular to $\mathbf{v}$,

$$\mathbf{w}=\mathbf{g}+\frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}, \tag{36}$$

and notice that $\mathbf{g}$ itself must be an equilibrium. From the results of the preceding section, the components of $\mathbf{g}$ must be all positive or all negative. Thus either $\mathbf{g}$ or $-\mathbf{g}$ is physical. But notice that $\mathbf{g}\cdot\mathbf{v}=0$, this leads to a contradiction with the first case. Therefore the equilibrium is unique. Q.E.D.

We conclude that in the case of the map U, there must be at least one fixed point. In the linear case, the map must be contracting and thus reach an equilibrium. Finally, if the map is connected as defined in Sec. III C, the equilibrium is unique.

These results are in agreement with our intuition. As we said before, we do get an unphysical result when only diffusion is present or equivalently when our last grid $\hat{J}$ is not located far enough so as to enclose most of the beam. The reader might wonder why the ''connectedness'' includes backwards propagation: a linear map with damping alone does have one equilibrium, the origin, but connectedness is realized only with backwards propagation. The same is true with maps which have limited diffusion. If there is an upper limit to the magnitude of the random kick a particle receives, then it is possible to show that the equilibrium distribution vanishes at a certain amplitude. This amplitude corresponds to a region where the maximum change due to a stochastic kick is always less than the damping towards the origin. In the purely damped case, the backwards connectedness is essential to prove the uniqueness of the distribution.

Finally it is intuitively obvious that, for a disconnected case, the number of connected regions will determine the degeneracies in the final distribution. Again the purely damped case with islands is a good example to think about.

E. Conclusion on the theoretical digressions

We have set down the basic features of the map we intend to construct. Nevertheless this paper is truly worthless without *reproducible* examples. One needs to get a feel for the adequacy of the method and indeed its failures.

As mentioned before, we restrict ourselves to systems in one degree of freedom or approximately reducible to one degree of freedom. We believe that there is a lot to learn about this technique and its relationship with other methods even in this restricted environment.

IV. EXAMPLES AND RESULTS

In this section we will present first a simple beam-beam map where we, the authors, control all the aspects of the simulation. We will then simplify even further and produce a ridiculous (unphysical) one-dimensional map. This map will permit us to understand the relationship between different methods of calculation, namely, the brute force method, the Irwin method, and the method presented here.

Next we will present an example extracted from the Berkeley Advanced Light Source (ALS). This example presents some challenges even when restricted to one degree of freedom.

A. A little beam-beam map

We will show here some tracking results of the original beam-beam map suggested by Hirata. This map consists first of a rotation $\mathbf{r}$ which mimics the ring:

$$\begin{aligned} r_1(x,p) &= \cos(\mu_0)x+\sin(\mu_0)p, \\ r_2(x,p) &= \cos(\mu_0)p-\sin(\mu_0)x. \end{aligned} \tag{37}$$

Then it is followed by a one-dimensional beam-beam kick $\mathbf{b}$:

$$\begin{aligned} b_1(x,p) &= x, \\ b_2(x,p) &= p+8\pi\xi\frac{\exp(-x^2/2)-1}{x}. \end{aligned} \tag{38}$$

The variable ξ measures the first order linear tune shift (in revolution unit) resulting from the beam-beam force.

The complete map for the symplectic system denoted by $\mathbf{m}$ in the previous sections is just

$$\mathbf{m}=\mathbf{b}\circ\mathbf{r}. \tag{39}$$

The ''radiative'' map was constructed by adding a damping map $\mathbf{d}$ given by

$$\mathbf{d}(x,p)=\lambda(x,p)=e^{-\alpha}(x,p), \tag{40}$$

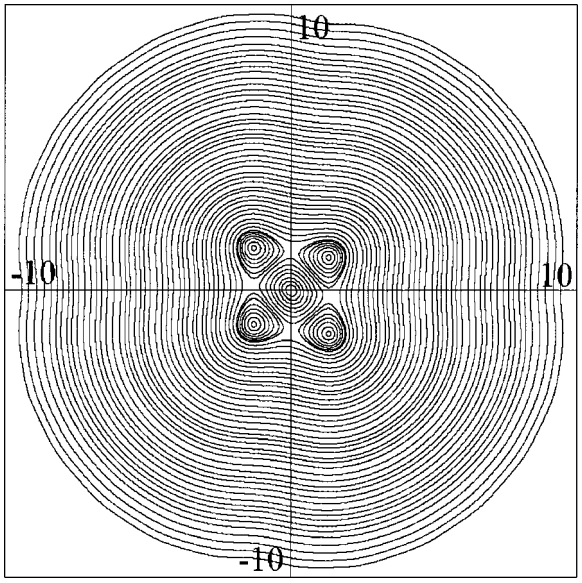

FIG. 2. Support grid.

followed by a stochastic map Σ,

$$\Sigma(x,p)=(x+\Delta,p). \tag{41}$$

The variable Δ was chosen naively to take the values $\pm\delta$ with equal probability. The constant δ can be related to the nominal beam sizes of the linear map. For example, the equilibrium emittance of the linear map is

$$\varepsilon_{\mathrm{eq}}=\frac{\delta^2}{1-\lambda^2}, \tag{42}$$

and the nominal beam size (or nominal sigma) is just

$$\sigma=\frac{\delta}{\sqrt{2(1-\lambda^2)}}\approx\frac{\delta}{2\sqrt{\alpha}}. \tag{43}$$

The complete stochastic map $\mathbf{m}_\Sigma$, entering in Eq. (20), is the composition of all these maps,

$$\mathbf{m}_\Sigma=\Sigma\circ\overbrace{\mathbf{d}\circ\underbrace{\mathbf{b}\circ\mathbf{r}}_{\mathbf{m}}}^{\mathbf{m}_\lambda}. \tag{44}$$

In Eq. (44) we also introduce the damped deterministic map $\mathbf{m}_\lambda$. This map, which contains the main component of classical radiation, is deterministic. It allows us to check the assertion that the full map does not destroy the symplectic structure too much. For example, in order to increase the speed of the simulation, one is tempted to increase the damping and the fluctuation components of $\mathbf{m}_\Sigma$, this cannot be done if the damped map $\mathbf{m}_\lambda$ dramatically changes the island size. In practice it means that big islands, under a small damped forced, become basins of attraction for their respective elliptic points. This can be totally destroyed for larger damping decrements.

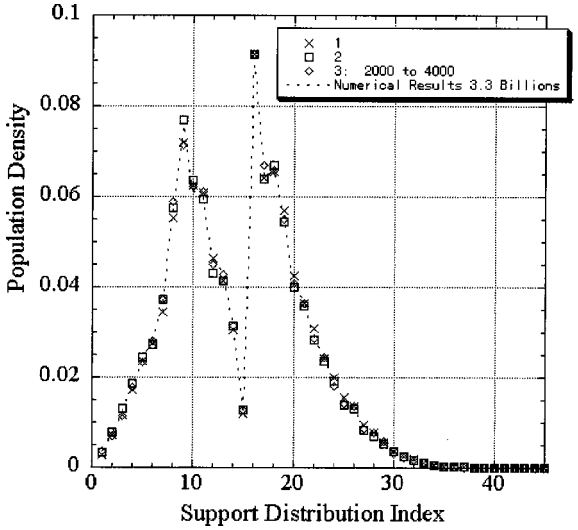

FIG. 3. Comparison of brute force and map calculations.

B. Where is the distribution?

Let us now turn to some real simulations using the little map of Sec. IV A. The parameters of the map were selected as follows:

$$\nu_0=\frac{\mu_0}{2\pi}=0.23,$$

$$\xi=1.5\times10^{-2},$$

$$\lambda=1-10^{-3}, \tag{45}$$

$$\delta=9\times10^{-2}. \tag{46}$$

The first stage in our method consists in producing a phase space portrait of the support distribution.

In Fig. 2 we see 55 trajectories which will be used to support the distribution. The reader will appreciate that the computation of a one-turn map is extremely inaccurate because the actual map would depend critically on the projection operator π that projects an arbitrary distribution on the chosen grid. The reason for this is simple: in one turn particles do not migrate far as they stay in the immediate neighborhood of the support distribution from which they originate. For this reason we selected here a 200-turn map, i.e., a map representing about 20% of a damping time.

In the calculation of the map U we arbitrarily launch $MN_b=1000$ trajectories twice. These are denoted by the labels ''1'' and ''2'' in Fig. 3. We also did a ''manual computation.'' By this we mean that we watched the various coefficients U_{ab} for a fixed b being plotted as they are computed; we then stopped the execution when the logarithmic bar plot seemed to settle down. We found out that this took between 2000 and 4000 trajectories. This is shown in curve number 3. The agreement is quite good and, in particular, the agreement is equally good (or bad) in the core and the tail. In fact the core presents a special problem: at small amplitude the damping will always be too small and thus the details of its effect depend on the projection algorithm π.

We can see the tail results in the logarithmic plot of Fig. 4.

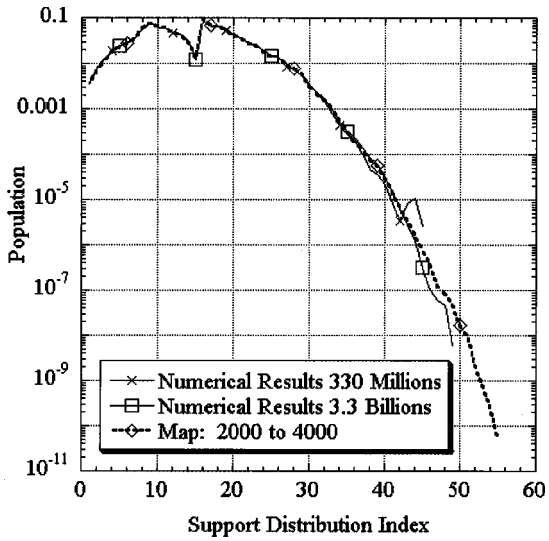

FIG. 4. Logarithmic plot of the tail.

The reader notices that we made two brute force simulations: one simulation with 3.3×10^{9} turns and a simulation with 330×10^{6} turns. We present both numerical results so as to gauge the accuracy of the brute force simulation. The results appear to be quite good.

In the next section we investigate the big island case; as we will see, it is very hard to obtain reliable results.

C. Big islands: Study with the beam-beam map

When the nominal beam size becomes small compared to the size of the islands we can expect the beam to split into what appears to be two independent distributions: one distribution in the main island centered around the origin and one distribution around the secondary islands—the fourth order islands in Fig. 2. In such a case the equilibrium distribution may be very hard to compute (see Sec. IV E). In particular, brute force tracking may fail as the test particle spends an eternity in one island. Indeed the separatrix is now situated in the ''tail'' of this distribution and thus particles seldom reach it. It is also our belief that the Irwin method may also fail unless it is modified to include two cores; this will be examined in Sec. IV D.

Indeed the brute force and the plain Irwin method will fail qualitatively unless a prohibitive number of turns and/or particles are used. On the other hand, our method will always produce good qualitative results but will have a hard time nailing down the exact ratio if the islands are truly very big.

Let us look at two cases of the beam-beam map. In both cases the deterministic parts of the maps are identical,

$$\nu_0=\frac{\mu_0}{2\pi}=0.24,$$

$$\xi=1.5\times10^{-2}, \tag{47}$$

$$\lambda=1-10^{-3},$$

but the fluctuation is different,

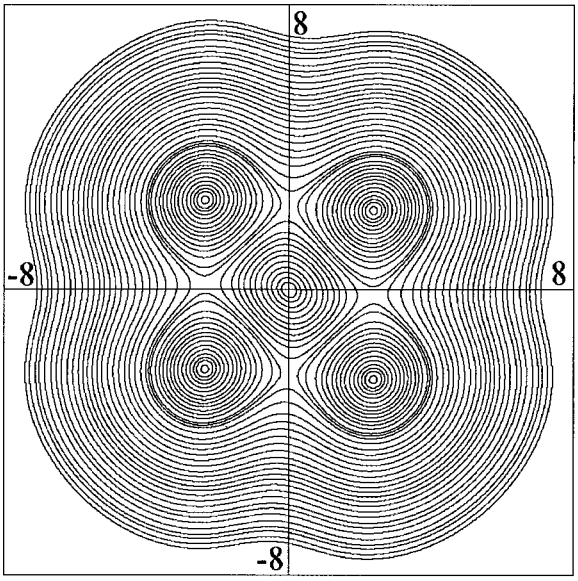

FIG. 5. Support grid for the big island case.

$$\delta=9\times10^{-2},\quad \sigma=1.42\ \text{(case 1)},$$
$$\delta=3\times10^{-2},\quad \sigma=4.74\times10^{-1}\ \text{(case 2)}. \tag{48}$$

The support grid for our simulations is given by Fig. 5.

The next plot shows a tridimensional plot of the brute force simulations for both values of the fluctuation parameters.

The important number to consider here is the transfer rate between islands. Since the brute force method tracks a particle for a long time, it is important that it visits all the allowable phase space. Once the islands start separating as the fluctuation strength diminishes, then the probability of crossing from one island to the other also diminishes. In the case of $\delta=0.03$ we see that a particle transferred on average once every 370 000 turns, while in the case of $\delta=0.09$ the transfer rate is greatly enhanced at one transfer per 5300 turns.

1. Back of the envelope explanation of the results

To understand what is going on here let us reverse the problem. Let us think of two distributions sitting at equilibrium very far from each other; we assume that each one sits around its own equilibrium center. This can be realized mathematically by inventing a system with a discontinuous behavior around a boundary which separates phase space into two identical linear regions. (We will study such a fake system in Sec. IV D.) This is a little bit what happens in the presence of islands. Furthermore, if the random kick has a finite maximum value (as in our case), then the distribution will fall to zero at some maximum amplitude. Let us call this amplitude $\hat{a}$. In addition, if we assume that the center of each distribution is separated by a distance greater than $2\hat{a}$, then there is no contact possible between each region and the two equilibria are centered around their respective fixed points. The map U for such a system would have two eigenvectors with unit eigenvalues.

If we reduce the distance between the fixed points below the critical value of $2\hat{a}$, then we expect the distribution to

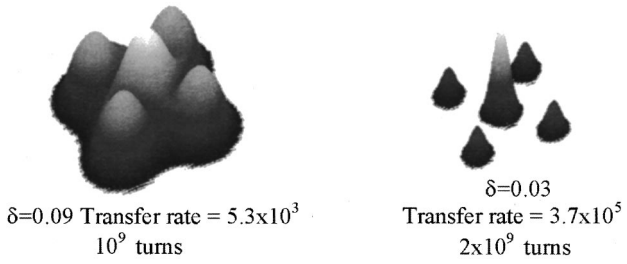

FIG. 6. Brute force simulations.

start mixing until a single equilibrium is reached. Because of the symmetry, we expect the equilibrium distribution to have an identical number of particles, namely, $(N_1+N_2)/2$ where N_i is the initial population of the ith region. Parenthetically it is interesting to see how the two independent eigenvectors must combine so as to produce one physical distribution with eigenvalue one and one unphysical distribution with an eigenvalue less than one and a population of zero. Remember that all the eigenvectors of the map U must have a population of zero unless they are equilibrium distributions. Thus before contact, the two independent vectors are centered around their respective fixed points. For an infinitesimal contact, a single equilibrium distribution is formed as the sum of the two previous equilibria with equal population. The other eigenvector, with an eigenvalue infinitesimally below one, is made of the difference of the two previous equilibria with equal population. In actual physical cases, such as those displayed in Fig. 6, one does indeed get a second eigenvector, with components of opposite signs in each island as is shown in Fig. 7.

The first question to ask is how big is $\hat{a}$? For our simple stochastic distribution it can be computed exactly in the case of a linear map. However, let us just say that it is bigger than the nominal σ of Eq. (43) and that it is typically of order $\sigma/\sqrt{\alpha}$. Therefore for a damping of $\alpha=0.001$ we expect that $\hat{a}$ will be about 30 times larger than σ. The exact ratio [9] depends on the details of the linear map and the stochastic process.

Calling the distances between the two fixed points $2D$, then we can say that for values of D a bit smaller than $\hat{a}$ it

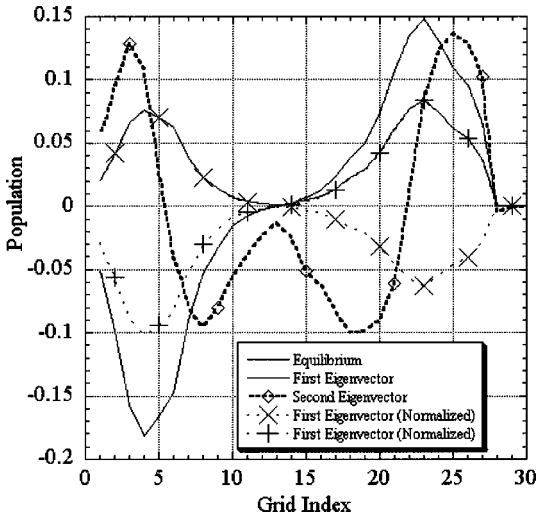

FIG. 7. The three leading eigenvectors for $\delta=0.03$.

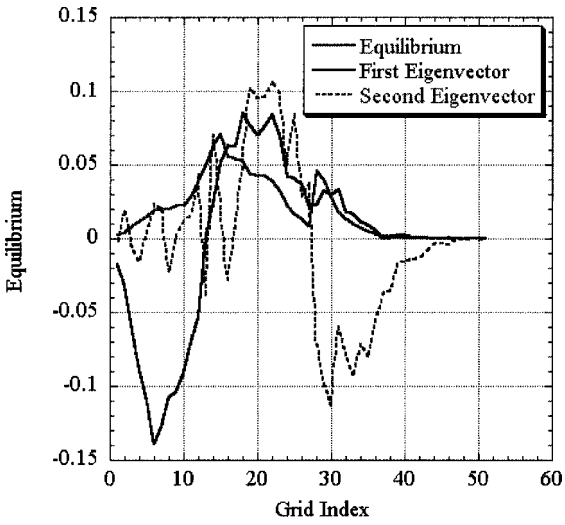

FIG. 8. The three leading eigenvectors for $\delta=0.09$.

would take an astronomical time to reach equilibrium. However, if we bring them suddenly into closer contact, we can assume that the two distributions are nearly Gaussian and estimate the relaxation time. The rate of change (per turn) for the population of a Gaussian distribution is given at large amplitude by the approximate formula [10,7]

$$\frac{dN}{dn}=N\frac{2\alpha\varepsilon}{\varepsilon_0}\exp(-\varepsilon/\varepsilon_0)=N\frac{\alpha\varepsilon}{\sigma^2}\exp\left(-\frac{\varepsilon}{2\sigma^2}\right). \quad (49)$$

For example, at a radius $\sqrt{\varepsilon}=5\sigma$, the relaxation time (approximately N/dN) is

$$\mathcal{T}=\frac{N}{dN}=\frac{1}{10^{-3}25}\exp\left(\frac{25}{2}\right)\approx 1.07\times 10^7. \quad (50)$$

This large number indicates that the two distributions will take more than 10×10^9 turns to reach equilibrium. This is at the limit of our calculational abilities. Let us look at the plot of Fig. 5 and estimate the surface enclosed by the separatrix. In the units of the graph, we get an area of 10.25 for the quasirectangle around the origin. This is equal to $\pi\varepsilon$, and thus for $\delta=0.03$

$$\frac{\varepsilon}{\sigma^2}\approx\frac{10.25}{\pi(0.414)^2}\approx 19, \quad (51)$$

leading to a relaxation time

$$T=N\left(\frac{dN}{dn}\right)^{-1}\approx 7\times 10^5. \quad (52)$$

This number is not inconsistent with the transfer rate found for the case $\delta=0.003$ despite the handwaving we used: we assume that the islands are of equal size and we neglected distortions of the linear dynamics. In any event we just wanted a ''ballpark'' number.

2. Explanation of the results using the eigenvectors

It is very interesting, in light of the discussion of Sec. IV C 1, to plot the eigenvectors of the map. In particular, let us look at the case $\delta=0.03$. It is given in Fig. 7.

The first curve depicts the equilibrium distribution which goes nearly to zero between the islands. The second curve shows the second eigenvector, the one with an eigenvalue closest to one. (Please notice that the equilibrium is thus the ''zeroth'' eigenvector in Figs. 7 and 8.) As expected, it is made of a linear combination of the first island and the second island distribution. The fourth and fifth curves show the same eigenvector with two different normalizations. The reader will notice that the fourth curve follows the first bump of the equilibrium distribution very well while the fifth curve is almost on top of the second bump; thus this calculation vindicates what was said in the preceding section.

The third curve represents the second eigenvector. The reader will notice that it looks very much like a derivative of the equilibrium distribution. This eigenvector would have been the leading eigenvector in a more conventional monotonically decreasing distribution. The local maxima of the distribution have a population close to zero with a high gradient while the regions of high gradient are now flat. In the case of a simple system and under the Fokker-Planck approximation, it can be proven that the eigenvectors are related to the Laguerre polynomial of order zero and do have an increasing number of zeros as their eigenvalue decreases.

We can therefore say that the rate of transfer between islands is controlled by the eigenvalue λ_1 while the mixing within each island is controlled by λ_2. In a case without a big isolated island, then λ_2 becomes λ_1. In Fig. 8 we display again the first three eigenvectors. The reader will notice that, despite the clear existence of a populated island, the second eigenvalue has a rather messy shape.

In the case of $\delta=0.03$, the first eigenvalue (for one turn) was computed to be $\lambda_1=1-(1.147\times10^{-5})$. Therefore, if we assume (incorrectly) that the probability of jumping from one island to the next is given by $\mu_1=1-\lambda_1=1.147\times10^{-5}$ and that the number of jumps obeys a Poisson distribution, then we can expect a jump every μ_1^{-1} turns, that is, every 87 000 turns.

In the next section, we show how the Fokker-Planck treatment gives a similar result in the linear case.

3. The Fokker-Planck treatment

In this paper, we have avoided the Fokker-Planck treatment completely. Nevertheless it is interesting to see how the eigenvectors come out of the linear problem in cases which we can solve analytically.

The idea here is to derive a stochastic operator for U in differential form rather than the more correct integral form. Let us look at the effect of the map $\Sigma\circ\mathbf{d}$ of Eq. (47) on an arbitrary distribution f. It is given by the exact relation

$$\overline{f}=\int_\Delta e^{2\alpha}f(e^\alpha x-\Delta_1,e^\alpha p-\Delta_2)\rho(\vec{\Delta})d\vec{\Delta}. \tag{53}$$

Equation (53) refers only to the stochastic part (damping included) of the map. To this we must add the map $\mathbf{m}$ which is purely deterministic:

$$\mathcal{U}f=\overline{f}\circ\mathbf{m}=\left\{\int_\Delta e^{2\alpha}f(e^\alpha x-\Delta_1,e^\alpha p-\Delta_2)\rho(\vec{\Delta})d\vec{\Delta}\right\}\circ\mathbf{m}^{-1}. \tag{54}$$

Let us concentrate on the stochastic part. If we assume that the damping and the stochastic kick are small, we can expand the integral as follows:

$$\overline{f}=\int_\Delta\left\{f+2\alpha f-\nabla f\cdot\vec{\Delta}+\alpha\vec{x}\cdot\nabla f+\frac{1}{2}\Delta_i\Delta_j\partial_i\partial_j f\right\}\rho(\vec{\Delta})d\vec{\Delta}. \tag{55}$$

We can now perform the integral over the stochastic kick:

$$\overline{f}-f=\frac{\partial f}{\partial n}=\underbrace{2\alpha f+\alpha\vec{x}\cdot\nabla f}_{deterministic\ part}+\underbrace{\frac{1}{2}\langle\Delta_i\Delta_j\rangle\partial_i\partial_j f}_{stochastic\ part}$$

where we assumed $\langle\vec{\Delta}\rangle=0$. (56)

We have introduced the timelike parameter ''n'' to denote the turn number. Obviously we can integrate this equation with respect to ''n'' and this will not change the results to leading order:

$$\overline{f}=\exp(\hat{\Sigma})f,$$

$$\text{where}\quad\hat{\Sigma}=\alpha\{2+\vec{x}\cdot\vec{\nabla}\}+\frac{1}{2}\langle\Delta_i\Delta_j\rangle\partial_i\partial_j. \tag{57}$$

We can now express the full one-turn map U as

$$Uf=\{\exp(\hat{\Sigma})f\}\circ\mathbf{m}^{-1}$$

$$\text{or}\quad U=\mathcal{M}^{-1}\exp(\hat{\Sigma})\quad\text{where}\quad\mathcal{M}^{-1}f\overset{def}{=}f\circ\mathbf{m}^{-1}. \tag{58}$$

The map $\mathcal{M}^{-1}$ is the usual symplectic map which can be expressed in terms of Poisson bracket Lie operators.

Now let us perform a phase average. Since we have gone through the trouble of writing a propagator for our map, let us perform a canonical transformation on it:

$$\begin{aligned}\mathcal{A}U\mathcal{A}^{-1}&=\mathcal{A}\mathcal{M}^{-1}\exp\left(\hat{\Sigma}\right)\mathcal{A}^{-1}\\&=\underbrace{\mathcal{A}\mathcal{M}^{-1}\mathcal{A}^{-1}}_{\mathcal{R}^{-1}}\mathcal{A}\exp\left(\hat{\Sigma}\right)\mathcal{A}^{-1}\\&=\mathcal{R}^{-1}\exp\left(\mathcal{A}\hat{\Sigma}\mathcal{A}^{-1}\right)\quad.\end{aligned} \tag{59}$$

By assumption the canonical transformation turns the symplectic map into a rotation. Now let us compute the effect of $\mathcal{A}$ on the stochastic operator $\hat{\Sigma}$. Let us start by examining the transformation properties of a regular deterministic vector field. They are given by the formula (repeated indices are summed over)

$$\mathcal{A}F_i\partial_i\mathcal{A}^{-1}=G_i\partial_i\Rightarrow G_i=(F_a\partial_a a_i^{-1})\circ\mathbf{a}. \tag{60}$$

Next we should examine the transformational properties of a diffusion operator of the type $D_{ij}\partial_i\partial_j$. In general, if the canonical transformation $\mathcal{A}$ is nonlinear, then the diffusion operator will transform into the sum of a diffusion operator and a deterministic vector field. The formula is given by

$$\mathcal{A}D_{ij}\partial_i\partial_j\mathcal{A}^{-1}=\overline{D}_{ij}\partial_i\partial_j+\overline{D}_i\partial_i,$$

$$\overline{D}_{ij}=\{D_{ab}\partial_a a_i^{-1}\partial_b a_j^{-1}\}\circ\mathbf{a}, \qquad (61)$$

$$\overline{D}_i=\{D_{ab}\partial_a\partial_b a_i^{-1}\}\circ\mathbf{a}.$$

Let us assume that the map $\mathbf{m}$ is just a linear map and thus it can be normalized by a regular Courant-Snyder transformation:

$$a_1^{-1}(x,p)=\frac{1}{\sqrt{\beta}}x,$$

$$a_2^{-1}(x,p)=\frac{\alpha}{\sqrt{\beta}}x+\sqrt{\beta}p. \qquad (62)$$

If we substitute this transformation into Eq. (61), we get the following results:

$$\mathcal{A}\hat{\Sigma}\mathcal{A}^{-1}=\frac{1}{2}\left\{\underbrace{\frac{1}{\beta}\left\langle\Delta_1^2\right\rangle}_{D'_{11}}\partial_1^2+\underbrace{\left[\frac{\alpha^2}{\beta}\left\langle\Delta_1^2\right\rangle+\beta\left\langle\Delta_2^2\right\rangle+2\alpha\left\langle\Delta_1\Delta_2\right\rangle\right]}_{D'_{22}}\partial_2^2\right.$$
$$\left.+\underbrace{\left[\left\langle\Delta_1\Delta_2\right\rangle+\frac{\alpha}{\beta}\left\langle\Delta_1^2\right\rangle\right]}_{D'_{12}}\partial_1\partial_2\right\}+\alpha\{2+\vec{x}\cdot\nabla\}\quad . \qquad (63)$$

This operator of Eq. (63) is two dimensional and completely equivalent to the original operator. However, we would like to average over the phase of the symplectic motion. This is best done by introducing another (nonlinear) transformation into the action angle variables (ϕ,J). The transformation analogous to that of Eq. (62) is just

$$c_1^{-1}(x_1,x_2)=-\tan^{-1}\frac{x_1}{x_2},$$

$$c_2^{-1}(x_1,x_2)=\frac{x_1^2+x_2^2}{2},$$

or equivalently

$$c_1(x_1,x_2)=\sqrt{2x_2}\cos x_1,$$

$$c_2(x_1,x_2)=-\sqrt{2x_2}\sin x_1. \qquad (64)$$

Since we are going to ignore any angle dependence, we need only to compute the $\overline{D}_2$ and $\overline{D}_{22}$ terms of Eq. (61). After some algebra, we obtain the results

$$\overline{D}_2=D'_{11}+D'_{22},$$

$$\overline{D}_{22}=2J(D'_{11}\cos^2\phi+D'_{22}\sin^2\phi-D'_{12}\sin 2\phi). \qquad (65)$$

We may now substitute the results of Eq. (63) into Eq. (65) and average over the phase ϕ. The total averaged map $\langle U\rangle$ becomes

$$\langle U\rangle=\exp\left(\frac{\partial}{\partial J}\left\{2\alpha J+\frac{D}{2}J\frac{\partial}{\partial J}\right\}\right),$$

$$D=\gamma\langle\Delta_1^2\rangle+2\alpha\langle\Delta_1^2\Delta_2\rangle+\beta\langle\Delta_2^2\rangle. \qquad (66)$$

The symplectic part of the map $\mathcal{R}^{-1}$ leaves any function of J invariant and thus is just the identity once we average over ϕ. It is interesting to investigate the eigenfunctions of the operator $\langle U\rangle$:

$$\frac{\partial}{\partial J}\left\{2\alpha J+\frac{D}{2}J\frac{\partial}{\partial J}\right\}w_k=\lambda_k w_k. \qquad (67)$$

Let us first look at the equilibrium distribution, which must obviously obey

$$\frac{\partial}{\partial J}\left\{2\alpha J+\frac{D}{2}J\frac{\partial}{\partial J}\right\}w_0=0, \qquad (68)$$

and have a finite integral. The answer can be easily found to be

$$w_0(J)=J_0^{-1}\exp(-J/J_0),\quad \text{where}\ \ J_0=\frac{D}{4\alpha}. \qquad (69)$$

The next step consists in writing a general eigenvector as the product of an unknown function p and the equilibrium w_0. The resulting equation for p is

$$J\ddot{p}+(1-J/J_0)\dot{p}=\frac{2\lambda_k}{D}p. \qquad (70)$$

The solutions are expressible in terms of the Laguerre polynomials:

$$w_k(J)=w_0(J)\mathcal{L}_k(J/J_0),$$

$$\lambda_k=-2\alpha k,\quad k=0,1,2,\ldots. \qquad (71)$$

The Laguerre polynomials can be obtained from the recursion relation

$$\mathcal{L}_k(x)=\frac{1}{k!}e^x\frac{d^k}{dx^k}\{e^{-x}x^k\}. \qquad (72)$$

The reader will notice that the eigenvector w_1 has a single zero at J_0 and thus looks like the derivative of w_0 with respect to the radius $r=\sqrt{2J}$. (Not exactly, since the zero of the derivative is at $J_0/2$.) Finally, we see that the eigenvalues of the exponential operators are given by the formula $\exp(\lambda_k)$ or $\exp(-2\alpha k)$.

Therefore as one would expect in a one-island situation, the modes damp at a rate consistent with the damping of the deterministic motion. This is why we often track a distribution of the order of a few damping times to get a general idea of the core equilibrium distribution. Obviously this must fail in a large island case.

D. Big islands: A one-dimensional fake map

The following question arises. When the islands are separated beyond the ability of brute force tracking to work, what can we expect from our map method, what can we expect from the Irwin method?

To understand the relative merits of all the methods, we will create a true one-dimensional problem. Consider the following system defined for positive x only:

$$\forall x>0 \quad m(x)=x,$$

$$d(x)=\begin{cases}\lambda x & \text{if } x<x_{\text{mid}}\\ x_{2D}-\lambda(x_{2D}-x) & \text{if } x_{2D}>x>x_{\text{mid}}\end{cases} \tag{73}$$

$$\Sigma(x)=x+\Delta, \quad \Delta=\pm\delta$$

$$\text{if } x+\Delta<0 \text{ then } \Sigma(x)=-\Delta-x$$

$$\text{and if } x+\Delta>\hat{x} \text{ then } \Sigma(x)=2\hat{x}-x-\Delta. \tag{74}$$

First let us look at the case $x_{\text{mid}}=x_{2D}=\hat{x}=\infty$. This corresponds to an attractive fixed point at the origin. Thus in the absence of a stochastic force, the equilibrium distribution will be a δ function at the origin as all particles are inexorably attracted towards $x=0$.

If we turn on the fluctuation, then the beam will settle near the origin. The reader can check that all the even equilibrium moments can be exactly computed. For example, we have

$$\langle x^2\rangle_\infty=\frac{\delta^2}{1-\lambda^2}, \tag{75}$$

while the odd moments depend on the distribution

$$\langle x\rangle_\infty=\frac{\delta}{1-\lambda}\left(\int_0^\delta \rho_\infty(x)\frac{(\delta-x)}{\delta}dx\right)<\frac{\delta}{1-\lambda}. \tag{76}$$

This simple model can be simulated easily by brute force, Irwin method, and our method. We should point out that our method is (unfairly) rigorous in this one-dimensional problem because it must be true that the distribution sits on the trajectories of the unperturbed problem (points on the real axis). However, we selected this simple problem because all methods are easily implemented.

For the case $x_{\text{mid}}=x_{2D}=\infty$, we will see that all the methods ''work'' in the sense that they produce good results. The Irwin method will actually produce extremely good results. In fact, the Irwin method can be shown to be a rigorous extension of the brute force method in such a case.

Now what can be said of the case $x_{\text{mid}}<x_{2D}<\hat{x}=n\langle x_\infty^2\rangle^{1/2}$? The reader notices that this system has two fixed points: one at zero and one at x_{2D}. We will find that the Irwin method has serious troubles if the islands are far apart; it is reduced to brute force. Let us describe in words the implementation of these three methods.

1. Brute force

Let us discretize the space into N points,

$$x_j=j\frac{x_{2D}}{N}, \quad j=1,N. \tag{77}$$

Now we simply track a ray for a large number of turns M and record the visiting frequency of each interval $[x_j,x_{j+1}]$.

Obviously this method has the usual problem: to know the tail with a good accuracy one needs to waste an enormous amount of time building the core to a ridiculous accuracy. Therefore the number of turns needed becomes prohibitive.

2. The Irwin method

The Irwin method is a clever trick which allows one to keep a constant accuracy from core to tail. Suppose we determine that to know to the core with a 1% accuracy or better, it is necessary to track 10 000 particles. (Let us define the core as the region near the origin containing 90% of the particles.) Then the Irwin method proceeds as follows.

Step 0: Track 10 000 particles until they seem to settle to an equilibrium.

Step 1: Locate the surface x_{90} that determines the limit of the core; the position below which 90% of the particles are. Record the population distribution of that core; this will be the equilibrium core.

Step 2: (useless in one dimension) Now track the 10 000 for a few more damping times and record where the particles cross the x_{90} boundary. Obviously in one degree of freedom, they all cross at the same place. Store this flow information for future usage.

Step 3: Start 10 000 particles in the region $x>x_{90}$. Track as in step 0. However, each time a particle falls below x_{90}, put a particle back on the surface x_{90} using the flow information computed at step 2. (In one degree of freedom, simply put back at x_{90}.) This step clearly assumes equilibrium.

Step 4: Go back to step 1. Now the word core applies to this new region.

This algorithm is looped around as many times as necessary. In the case when 90% is selected, we slowly advance towards the deeper part of the tail, each iteration looking into a region nine times less populated than the previous iteration. The accuracy, it is hoped, remains approximately the same, and more importantly, the number of particles needed for each region is the same for a constant accuracy.

This method, originally due to Irwin, has been very successful and is not limited to one or two dimensions. Furthermore, unlike brute force, it permits the exploration of tails for most problems (but not all, as we will see) with a good degree of accuracy.

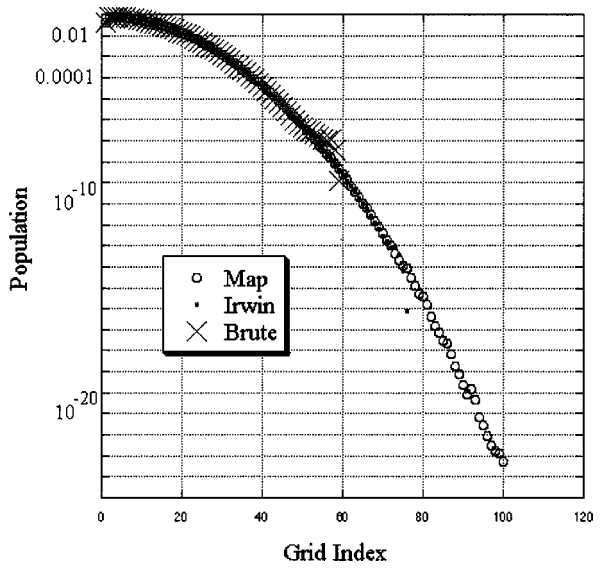

FIG. 9. Comparison in the case of $x_{\mathrm{mid}}=x_{2D}/2=\infty$.

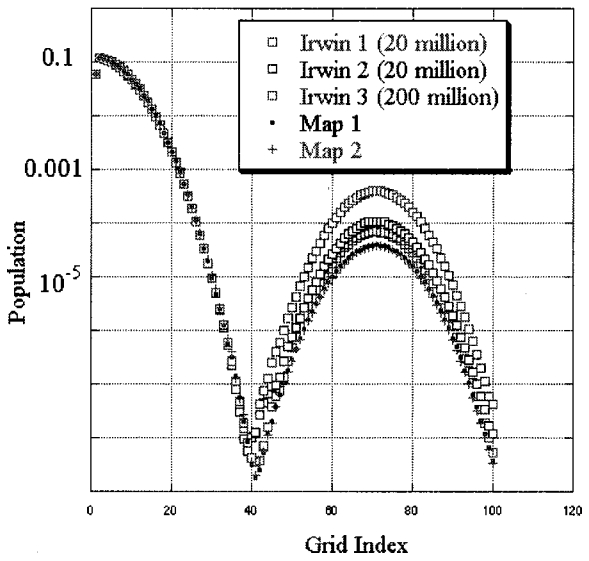

FIG. 10. Comparison for the separated islands case. $x_{\mathrm{mid}}=6\langle x^2\rangle^{1/2}$, $x_{2D}=10.5\langle x^2\rangle^{1/2}$, and $\hat{x}=15\langle x^2\rangle^{1/2}$.

It should be pointed out that the ergodic hypothesis, which is behind Irwin's method and brute force, allows us to replace the tracking of a core (10 000 particles, for example) by the tracking of a single particle as in the brute force method. This point, apparently not in Irwin's original paper, greatly simplifies the work. It also permits a unification of the brute force method and the Irwin method. The Irwin method is a clever extension of the brute force method rather than a competing algorithm. We will use it on the one-dimensional map of Eq. (73).

3. The map method

The map method is trivially applicable to this one-dimensional problem. The support trajectories are simply the grid points x_j. The π projection can be selected as in Eq. (10). As usual a number of turns equal to a substantial fraction of the damping time must be used. The end result will be a map U propagating a distribution supported by the ''points'' x_j.

4. No islands: $x_{mid}=x_{2D}/2=\infty$

Here we performed a brute force simulation with 280 000 000 particle turns. We also performed an Irwin calculation with 14 steps of 20 000 000 particle turns. At each step we advance by increments of 90% deeper in the tail. Finally we did a map calculation with 101 entries. A 100-turn map was computed with 28 000 stochastic experiments for each entry. In the language of Eq. (20), the map calculation was done with

$$q=100, \quad N_b=1, \quad M=28\,000. \tag{78}$$

The parameters for the map were selected as follows:

$$\lambda=0.999 \quad \text{and} \quad \delta=0.09,$$

with a total grid over the region

$$x\in[0,100\langle x^2\rangle^{1/2}], \quad \langle x^2\rangle^{1/2}\approx 2.013.$$

This interval was divided into 100 small intervals and thus into 101 points for the map computation. For the sake of comparison, the binning algorithm used the projection π of the map algorithm both in the brute force and in the Irwin calculation. We can see in Fig. 9 that the map method agrees in the tail with the Irwin method up to the point where the Irwin method starts to fall apart.

5. With islands: $x_{mid}<x_{2D}<\hat{x}=n\langle x_\infty^2\rangle^{1/2}$

In this case we considered the following ''tough'' case (see Fig. 10):

$$x_{\mathrm{mid}}=6\langle x^2\rangle^{1/2}, \quad x_{2D}=10.5\langle x^2\rangle^{1/2}, \quad \text{and} \quad \hat{x}=15\langle x^2\rangle^{1/2}. \tag{79}$$

The Irwin method was tried with 20×10^6 as well as 200×10^6 particle tracking per iteration. Each iteration attempted to build the distribution in increments of 50%, which is, of course, conservative.

The map method was also used and the results were far more constant, indicating that it probably settled to an answer closer to the correct one. Why do we get these bad results? It has to do with the sudden appearance of a corelike region at large amplitude. As we built the tail, this region is invisible; no particles whatsoever have yet entered it. Intrinsically there is nothing wrong with this because the same is true in a regular monotonically decreasing distribution. However, in the case of a big island, the center of ''gravity'' of the tail abruptly jumps as particles start populating this island. In our example it was typical to see a jump of the last Irwin surface from 25% of $\hat{x}$ to about 70%. In other words, the Irwin algorithm suddenly discovers the existence of a second beam orbiting around the island centered around x_{2D} and starts building the outer tail of that island. The inner tail between the central island and outer island is badly represented.

One way to fix this problem would be to increase the number of particle turns, but this is a return to brute force. The second possibility would be to devise a two-core Irwin method and match at the interisland boundary.

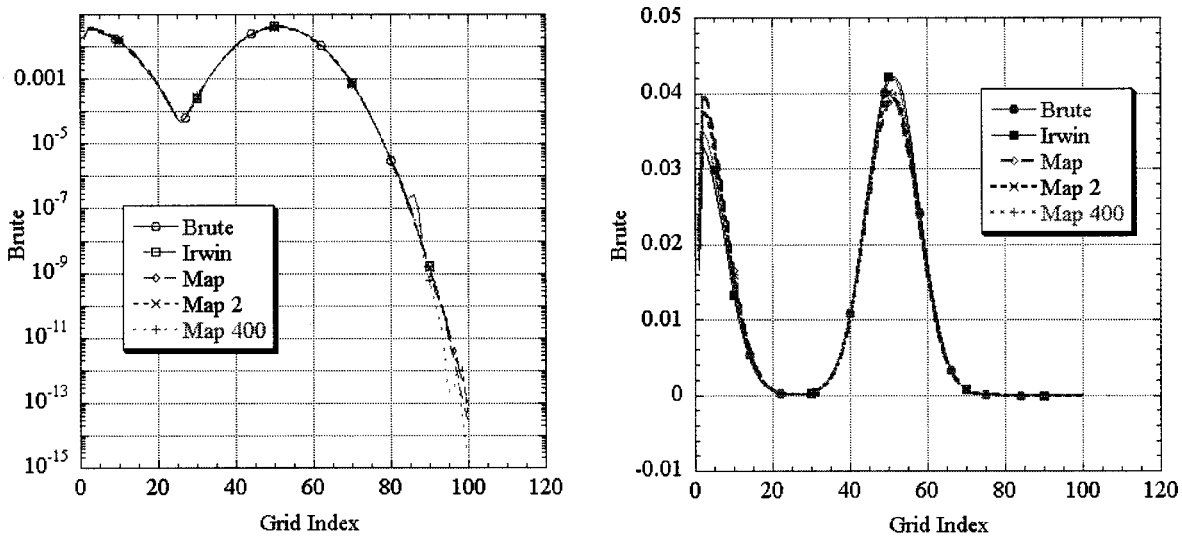

FIG. 11. Logarithmic and linear plots for the closer islands case. $x_{\mathrm{mid}}=3.75\langle x^2\rangle^{1/2}$, $x_{2D}=7.5\langle x^2\rangle^{1/2}$, and $\hat{x}=15\langle x^2\rangle^{1/2}$.

To show how things get better as the islands approach each other, we also simulated the following case (see Fig. 11):

$$x_{\mathrm{mid}}=3.75\langle x^2\rangle^{1/2}, \quad x_{2D}=7.5\langle x^2\rangle^{1/2}, \quad \text{and} \quad \hat{x}=15\langle x^2\rangle^{1/2}. \tag{80}$$

We simulated ''brute'' force with 280×10^6 particle turns. The particle visited each region 105 times and thus the brute force results must be sensible. The reader will notice that there is an excellent agreement between the brute force simulation and the Irwin method. Of course in the extreme tail, as the logarithm plot shows, the Irwin method agrees much better with the map method. Let us look at the case

$$x_{\mathrm{mid}}=5\langle x^2\rangle^{1/2}, \quad x_{2D}=10\langle x^2\rangle^{1/2}, \quad \text{and} \quad \hat{x}=10\langle x^2\rangle^{1/2}. \tag{81}$$

This case must have by symmetry equal population in both islands. So, for the purpose of making things a little asymmetric, two grids were chosen. In one case 100 intervals separated $11\langle x^2\rangle^{1/2}$ rather than 10.

This places the midpoint between the 45th and 46th grid point, at 45.45 to be precise. We also tried to put the center point at 45.25 using 100 intervals over a distance of about $11.0497\langle x^2\rangle^{1/2}$—it is interesting to see if this affects the total population of the two islands. The results are displayed in Fig. 12. The distribution is very symmetric as one would expect. Actually the percentages of the total population occupying the first island are found to be 51.4%, 48.4%, 47.9%, 50.0%, and 49.8% for the five simulations shown in Fig. 12.

Our last example, with this one-dimensional map, concerns the case of a large island with a small population. The parameters are

$$x_{\mathrm{mid}}=5\langle x^2\rangle^{1/2}, \quad x_{2D}=7.5\langle x^2\rangle^{1/2}, \quad \text{and} \quad \hat{x}=10\langle x^2\rangle^{1/2}. \tag{82}$$

We selected 100 grid points over 11.0497 (for no particular reason in this case). The results are in Figs. 13 and 14.

This case shows how the Irwin method, which works beautifully in a monotonic case, seems to have trouble with even small islands. The basic problem is the sudden population increase as one reaches the outer island.

E. An advanced light source case

The Berkeley Advanced Light Source is a 12-fold triple bend achromat ring. An experiment was conducted ''to gauge'' the amount of beam present in the main island as well as in the third order islands shown in Fig. 15.

At first we thought that these islands were perhaps populated by gas scattering or other long range phenomena. It appears that the ordinary radiation process is sufficient to cause the outer islands to populate. In any event, this is a good example of the method with its pluses and minuses.

The first hurdle is the dimensionality of the problem. The full problem certainly requires a multidimensional calculation. If we ignore the vertical direction, there still remain the synchrotron oscillations. However, if we go to a dispersion free point right at the symmetry point where insertion devices are normally put, the equilibrium beam will be ''one dimensional.'' By that we mean that the equilibrium qua-

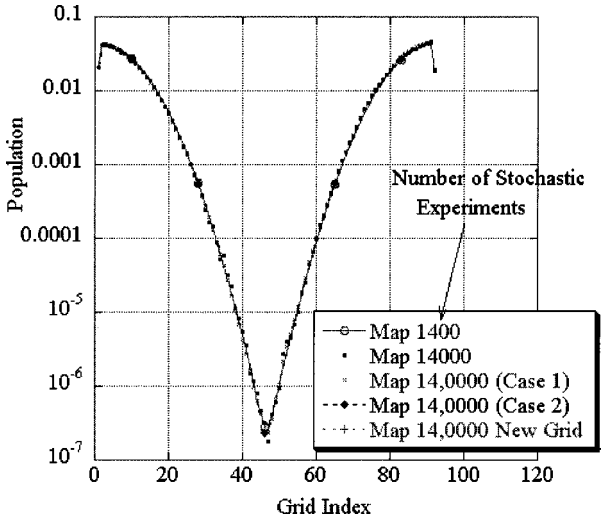

FIG. 12. Symmetric case (200-turn map). $x_{\mathrm{mid}}=5\langle x^2\rangle^{1/2}$ and $x_{2D}=\hat{x}=10\langle x^2\rangle^{1/2}$.

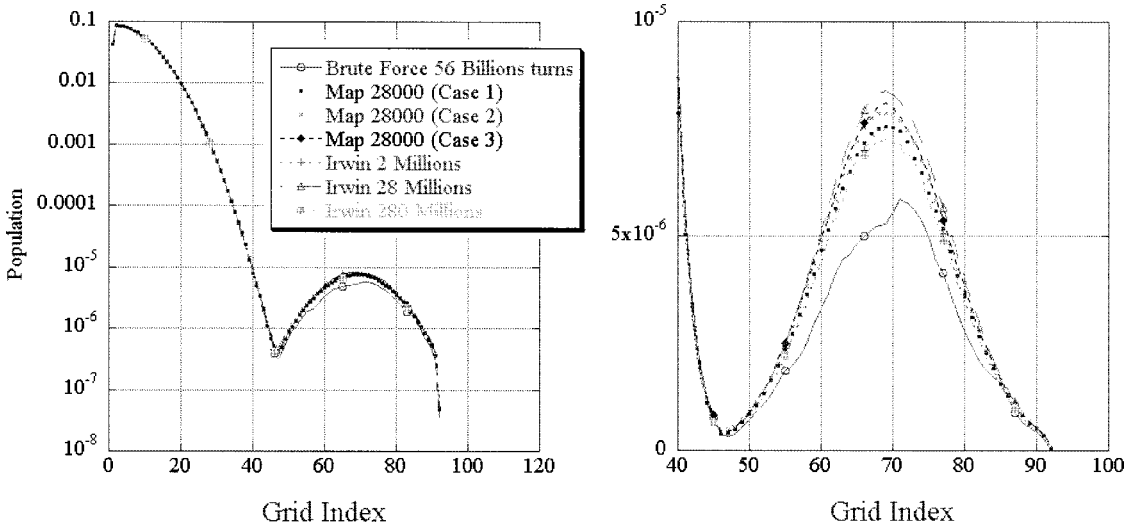

FIG. 13. Small island calculation with tail results magnified. $x_{\mathrm{mid}}=5\langle x^2\rangle^{1/2}$, $x_{2D}=7.5\langle x^2\rangle^{1/2}$, and $\hat{x}=10\langle x^2\rangle^{1/2}$.

dratic moments will be proportional or consistent with the existence of an equilibrium emittance whose shape is given by the one-dimensional Courant-Snyder theory. Therefore, using the beam envelope [11,12] formalism, we check this assertion and concoct a one-dimensional stochastic kick which will give us the correct horizontal moments at the symmetry point. There is a lot of hand waving, but it does reduce the dimensionality of the problem to a one-degree-of-freedom case.

For better or worse here is how we proceeded. First let us assume a one-dimensional problem. It is easy to show that the average moments are given by

$$\langle x^2\rangle=\frac{1}{2}\beta\langle\varepsilon\rangle,\ \langle p^2\rangle=\frac{1}{2}\gamma\langle\varepsilon\rangle,\quad \text{and } \langle xp\rangle=-\frac{1}{2}\alpha\langle\varepsilon\rangle. \tag{83}$$

The quantity $\langle\varepsilon\rangle$ is the average value of the Courant-Snyder invariant. Therefore if the beam sizes at the dispersion free point ''appear'' one dimensional, we expect

$$\frac{\langle x^2\rangle}{\beta}\approx\frac{\langle p^2\rangle}{\gamma}\approx-\frac{\langle xp\rangle}{\alpha}. \tag{84}$$

The moments were computed using a six-dimensional beam envelope formalism. We found these ratios to be constant within better than one part in a thousand. The next step consists in generating a random kick in the x direction which will reproduce the correct moments. The formula for the quadratic moment of that kick $\langle\delta^2\rangle$ is obtained by equating the stochastic change in the Courant-Snyder invariant with the change due to damping:

$$\langle\delta^2\rangle=\frac{4(1-\lambda)\langle x^2\rangle}{\beta\gamma}. \tag{85}$$

The stochastic kick as well as the damping parameter λ are input to a one-dimensional tracking of the ALS.

At this stage we can link our favorite tracking code with our GUI program and track element by element as we normally do. Of course since the beam is rather small we are tempted to generate a Taylor series map using truncated power series algebra (TPSA) and perform symplectic tracking with an adequate method. It should be said that this is a perfect problem on which to try the TPSA map tracking. We found out that the phase space of Fig. 15 can reproduce with

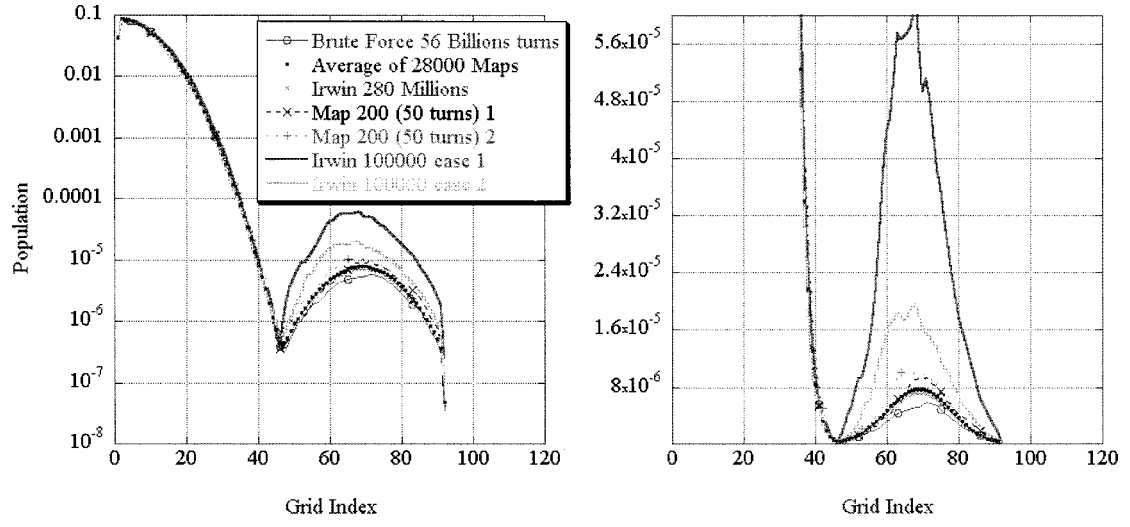

FIG. 14. Some good results for small islands using very approximate maps.

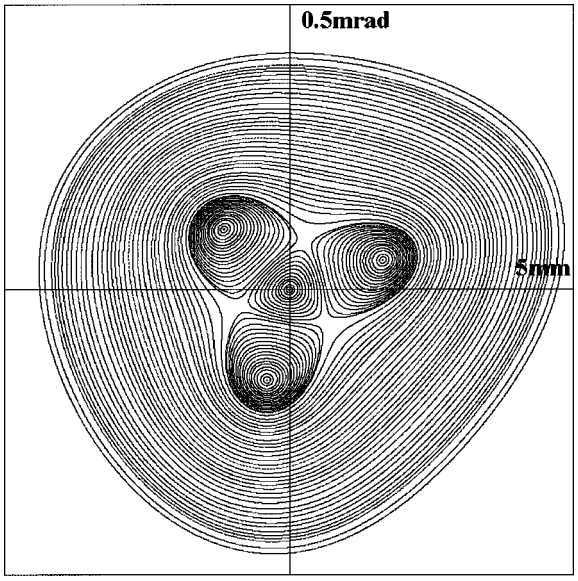

FIG. 15. Support grid for the ALS example.

a fifth degree Taylor series map. The map was factorized and its nonlinear part was tracked using a characteristic function.

Damping was added by a general multiplicative factor of

$$\lambda = \exp(-4.143\times 10^{-5}) \tag{86}$$

acting on the matrix part of the map. The stochastic kick was added as before on the variable x. To speed up tracking the damping was increased to $\lambda = e^{(-1.14\times 10^{-4})}$. We checked that the deterministic damped map had similar basins of attraction. The stochastic fluctuation was increased so as to keep the nominal beam sizes constant.

For the case of Fig. 15 the islands were close enough to be obtainable by brute for tracking. We tracked 500×10^6 turns and recorded the visitation frequency. We also recorded the number of times the particle moved from the middle of the main island to the middle of the outer island. We found that this happened 350 times. This implies that the relative population is probably known quite accurately by brute force tracking. The result is plotted in Fig. 16 where both linear and logarithmic plots are displayed. Numerically we found the population of the center island to 3.9% of the total population.

We found out that the map method does a particularly bad job at estimating the relative population. We are not comparing with the experiment but simply comparing brute force with the map method on a given representation of the experiment. For example, if we select the grid support displayed in Fig. 15, we can estimate the center island population to be at about 7.5% using a 10 000 turn map. Reducing the number of turns to 5000 increased this estimate to 9%. Pushing it to 20 000 turns gave us the more correct 4.5%. What is going on here?

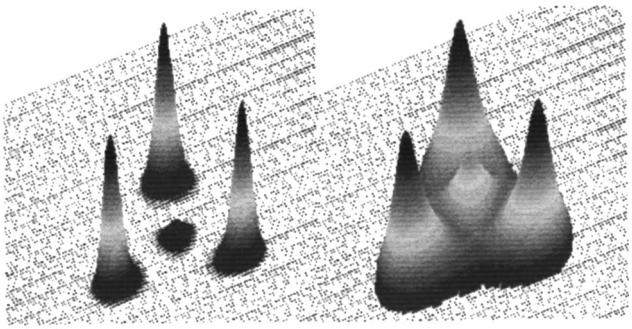

FIG. 16. Results of 500 000 000 turns of brute force tracking of ALS.

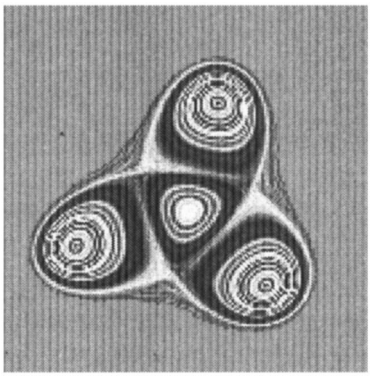

FIG. 17. Another view of the logarithmic plot for ALS.

First of all it is clear that if we push the number of statistical experiments towards brute force numbers (per trajectory), we will obtain the brute force result. The map, of little value, will simply indicate that all initial distributions tend towards the equilibrium distribution. Therefore we should not be surprised to see the results improving as the number of turns represented by the map U increases.

Why are the results so hard to obtain? We think that the large size of the islands creates a substantial region of slow tune. Figure 17 shows again the logarithmic plot where we have enhanced colors and shadows to reveal the asymmetry of the equilibrium distribution. Around the fixed points the equilibrium distribution does not follow the symplectic trajectories very well. Furthermore, this region is rather large, but more disturbing to our method is the importance of that region in a weakly populated tail. Indeed the relative population of the islands is determined by the transfer rate between support trajectories in the very close neighborhood of the separatrix if a small number of turns is used. Clearly this can be a serious problem if the support trajectories have no relationship with the equilibrium distribution. Using a larger number of turns is a crude way to still get a map of some value.

We presented this example of the ALS to show that things are not always simple even under the best circumstances. Nevertheless the map method does produce reasonable numbers.

V. CONCLUSION

We presented a method which uses the symplectic trajectories as support functions for representing a quasiarbitrary distribution. It is quite clear that there are serious limitations preventing a total generalization of the map method. However, there are also several topics we did not have the time or space to discuss which present potentially useful areas of applications.

(1) The estimate of lifetime, i.e., the creation of a map that has an aperture and thus does not conserve particle number.

(2) The study of rare events such as gas scattering. This can be done by computing a map for the rare event alone and ''adding'' it to the ordinary radiative map. There are a lot of tricks involved including the calculations of one-turn maps by taking the nth root of a map. It is a large topic on its own.

(3) The study of injection efficiency which combines some of the tricks needed above.

(4) Nonlinear effects: the map is a function of the distribution. We never tried anything along those lines but it should be possible.

Of course our ability to produce simple maps and perform some transformation on them may be needed as we try to study problems of higher dimensionality. For example, we cannot always evaluate the map at a dispersion free point, but we can do certain canonical transformations on the $6d$ map which will reduce the synchrotron sloshing on the fixed point.

So we conclude on both a note of hope and despair. There are obviously applications for this method, but there are also fundamental limitations.

ACKNOWLEDGMENTS

We would like to express our gratitude to Dr. Kohji Hirata of KEK, Dr. Helmut Mais of DESY, and Dr. Bob Warnock of SLAC. The correct proof in Sec. III B was provided by Bob Warnock. This work was supported by the Japanese Ministry of Education and Culture and by the U.S. Department of Energy. This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division, U.S. Department of Energy under Contract No. DE-AC03-76F 00098.

[1] H. Mais and E. Forest (private communication).

[2] A. Pauluhn, Ph.D. thesis, Universität Hamburg, 1993 (unpublished).

[3] J. Irwin (unpublished).

[4] E. Forest, *Beam Dynamics: A New Attitude and Framework* (Harwood Academic Publishers, Amsterdam, The Netherlands, 1997).

[5] A. Lasota and M. Mackey, *Probabilistic Properties of Deterministic Systems* (Cambridge University Press, Cambridge, UK, 1985).

[6] D. P. Barber, K. Heinemann, H. Mais, and G. Ripken, DESY Technical Report No. DESY M-91-16 (unpublished).

[7] A. Piwinski, CERN Report No. CERN 85-19, 1985 (unpublished).

[8] T. L. Saaty and J. Bram, *Nonlinear Mathematics* (McGraw-Hill, New York, 1964).

[9] This effect was brought to our attention by M. Furman. The exact position and shape of the boundary of the distribution can be obtained through normal form theory.

[10] K. Hirata, KEK Technical Report No. KEK 98-67, 1998 (unpublished).

[11] K. Ohmi, K. Hirata, and K. Oide, Phys. Rev. E **49**, 751 (1994).

[12] E. Forest, M. F. Reusch, D. Bruhwiler, and A. Amiry, Part. Accel. **45**, 66 (1994).