ARE WE BEATING LIOUVILLE'S THEOREM?

A. G. Ruggiero
Fermi National Accelerator Laboratory

I. Introduction

During the last two years or so, because of the renewed interest in various techniques of beam stacking, such as electron cooling, stochastic cooling, synchrotron radiation, charge exchange, etc., people have been heard to wonder how such techniques could work when Liouville's theorem states that the phase-space area of a beam is preserved. People have made statements like "Liouville's theorem has been beaten," "we went around Liouville," "Liouville's theorem does not apply here," "Liouville's theorem is valid only if you take all the universe into account," and so on. People have even been heard to comment that Liouville's theorem has been proven wrong. But the majority were simply mystified by what they see as a conflict between what Liouville's theorem implies and what is apparent from the beam handling of the various cooling and stacking techniques. Most of the confusion is caused, I believe, by the fact that people make Liouville say things he never meant!

About twenty years ago, the Liouville question was also raised in connection with studies of devices which could produce a damping mechanism for protons similar to the synchrotron radiation for electrons. At that time, effort was devoted to generalization of Liouville's theorem to include dispersive systems and systems of interacting particles.[1] We will not deal here with these relatively more recent findings, but will confine our analysis to the simple form of the Liouville theorem as it was originally formulated. The confusion mentioned above can be removed by simply inspecting how the theorem of Liouville works with the beams of charged particles that we usually accelerate or store.

II. Liouville's Theorem

Liouville published his work[2] in 1837. It is, of course, not easy to find the original paper, but Liouville's theorem is discussed in many books on statistical mechanics. The discussion in the Ehrenfest's book[3] is particularly concise and close to the original.

Let us see what the theorem says. To make things simpler, let us consider only one particle which has motion described by three pairs of canonical variables (q_i, p_i, $i = 1, 2, 3$) and by the Hamiltonian function $H(q_i, p_i, t)$,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}\,. \tag{1}$$

Consider also the six-dimensional phase space of coordinates q_i and p_i ($i = 1, 2, 3$). Assign to the particle some initial conditions, that is a point P of

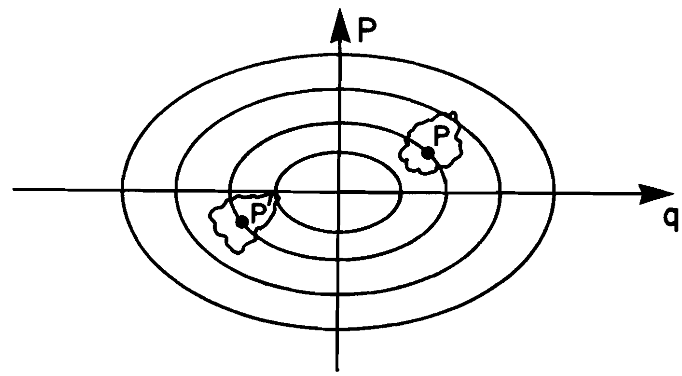

Fig. 1. The phase space

the phase space that it occupies at the initial time t_0. One can solve the equations of motion (1) with the assigned initial conditions and calculate the trajectory of the particle in the phase space. We assume this trajectory is closed and elliptically shaped, as shown in Fig. 1. We can repeat this operation indefinitely for every set of initial conditions, that is, for every point P taken as the starting point. By doing this, we have filled the phase space with an infinitely large number of trajectories which describe the motion of the same particle which assumes different initial conditions.

Consider now a region surrounding a particular point P. This region has volume V and includes an infinitely large quantity of points that we can regard as possible initial conditions of the particle at time t_0. For continuity reasons, all these points will occupy another region surrounding P' at a later time t. This second region can be calculated by solving the equations of motion (1) for each initial condition around P and marking the corresponding particle position at the time t. Obviously, because of the uniqueness of the solution of (1) there is a one-to-one correspondence between the points around P and those around P'. Liouville's theorem states that the volumes of the two regions are the same and equal to V. The proof of the theorem is relatively easy if one reminds oneself that the equations of motion in the Hamiltonian form (1) are equivalent to coordinate transformations with a Jacobian equal to unity. Thus Liouville's theorem can be stated also as follows: The streaming of the image points in the phase space as given by Eq. (1) generates a continuous point transformation, which transforms each six dimensional region into another one of the same volume. This is true at every time t at which we stop our process, and no matter what the initial volume V of the region surrounding P.

III. The μ-Space of a Physical System

Let us consider now a system of N particles like the beam of charged particles we usually deal with in accelerators or storage rings. The motion of each particle is again described by the equations (1). We assume the particles are not interacting with each other so that the Hamiltonian H will depend only on the coordinates of the particle under consideration. At the initial time t_0, the particles will occupy specific locations in a six-dimensional phase space similar to the one we described above and that we call μ-space. The previous space was used to represent the motion of the same particle with different initial conditions, whereas μ-space is used to show the trajectories of several different particles. By solving the equations of motion (1), we then have N trajectories in μ-space, one for each particle, as shown in Fig. 2. One can take a picture of μ-space at

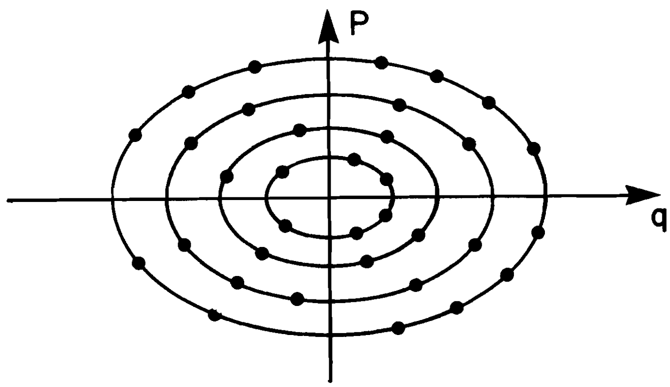

Fig. 2. The μ-space

a given time t and one sees N image points, each describing the location of one particle. Two particles cannot occupy the same location at the same time. There are at most N trajectories; several particles can share the same trajectory.

Even when N is very large but finite, μ-space is practically empty in contrast to the space of Sec. II which is continuously filled with all the possible trajectories of a single particle. It is therefore not obvious how useful the application of Liouville's theorem is to μ-space. One can in principle divide μ-space in six-dimensional cells, each large enough to contain a very large number of particles and yet small enough so that the coordinates do not change appreciably across their volume. With these requirements, one can then define reasonably well the particle density in phase space, which is the number of particles in a particular cell. This is a local average process and is very sensitive to fluctuations from cell to cell. The fluctuations are relevant to the statistical mechanics of a gas, but we will not deal here with them.

If the number of particles N becomes infinite, because they cannot occupy the same location at the same time and because they do not interact with each other, there is no difference noticeable between μ-space and the space we described in Sec. II. Each real particle is represented in the same way as a standard single particle with proper initial conditions. Thus with the above assumptions Liouville's theorem applies also to the μ-space of a real (continuous) system of particles. In particular, the cells could be made infinitesimally small and the density measured by a distribution function $\psi(q,p)$, which when multiplied by the volume element dqdp of the cell gives the number of particles. As defined, ψ is a continuous distribution. It is a consequence of Liouville's theorem that

$$\frac{d\psi}{dt} = 0,$$

which can also be written as

$$\frac{\partial\psi}{\partial t} + \sum_{i=1}^{3} \left(\dot{q}_i \frac{\partial\psi}{\partial q_i} + \dot{p}_i \frac{\partial\psi}{\partial p_i}\right) = 0, \tag{2}$$

a special case of the Vlasov equation for non-interacting particles. It is quite legitimate to make use of Eq. (1) for $\dot{q}_i$ and $\dot{p}_i$ in the left-hand side of (2).

Thus for a system of $N = \infty$ non-interacting particles it is possible to define a density ψ in μ-space and apply Liouville's theorem. The streaming of the image points is governed by the Vlasov equation (2) and statistical-fluctuation considerations do not apply here.

IV. A Beam of a Finite Number of Particles

Suppose that the system is again made of an infinitely large number of particles but they are all confined initially in a finite volume V of the μ-space, so that outside this region $\psi = 0$. Liouville's theorem states that the volume V is preserved during the motion of the system. A real beam of charged particles is always made of a finite number N of particles, but it is quite common to make the approximation of a continuous distribution, which implies $N \to \infty$. With this approximation, it is possible to define a volume V of the phase space constantly occupied by the beam and which is often called the beam emittance. But a closer view of the distribution of the particles of a real beam as shown in Fig. 2 shows that since each particle occupies a zero-volume element of space and there is a finite number of particles, the actual volume occupied by the beam is zero. One can avoid this inconsistency by dividing the μ-space in cells as explained above, and consider only those cells that at a given time are occupied by particles. The sum of the volumes of all these cells can be defined as the beam emittance in the case that N is finite. Similarly, a density function ψ can also be introduced by taking the ratio of the number of particles in a given cell to the volume of the same cell. So defined ψ is a discontinuous function that can be approximated by a smooth one.

If the number of particles in a cell is sufficiently large and uniformly spread, their image points in the μ-space can be thought as representatives of typical possible replicas, at some time t, of the reference particle. All the particles that occupy a particular cell at an initial time t_0 are expected to occupy at a later time t another one with the same volume, apart from statistical fluctuations. Thus one would expect that Liouville's theorem applies also to the case of systems with a finite number N of particles. One would conclude this after having applied local averages as we have described, and, again, apart from statistical fluctuation. These are basically the arguments that make people consider an actual beam of N particles as a Liouvillean system, and so define a distribution function ψ and apply the Vlasov equation (2) to it. Hence, one is encouraged to make the statement that the beam phase-space volume (or area) is preserved.

V. Example of Conservative Systems That Do Not Preserve Phase-Space Area

In the following, we want to give two examples which show that, despite the fact that the motion is conservative and described by a Hamiltonian function, the phase-space area of a beam of a finite number N of particles, defined with the average process described above, is not preserved.

First example. Consider the case of Fig. 3, which shows a debunched beam in the longitudinal phase space of variables ϕ, the phase angle in rf units, and $\Delta p/p$, the relative momentum deviation. Suddenly an rf cavity system is turned on to bunch the beam. The rf voltage creates a stationary bucket whose separatrix is shown in Fig. 3. The motion of the particles changes from a simple drift along the angle-axis to an oscillation around the center O of the bucket. The oscillation frequency is maximum for particles with small amplitude, that is in the neighborhood of O, and decreases moving toward the edge of the bucket; near the separatrix, the phase oscillation frequency becomes very small, practically zero. In Fig. 3, we show the shape of the beam after several phase oscillations. The filamentation is caused by the difference of oscillation frequencies. We have shown with continuous lines the boundary of the beam. The area which is stretched between them would be the area of the beam in the case it is made of an infinitely large number N of particles. This area is invariant, because of Liouville's theorem, and equal to the area of the original strip. In this case, which deals with the beam as a continuous medium, one can calculate the shape of the beam bunch by means of the Vlasov equation (2). As the motion proceeds, the number of fans of the filamentation increases. The beam looks like a long ribbon wrapped on itself in a spiral motion; the ribbon length gets longer and also more and more narrow to preserve the area. This characteristic should always be recognizable for a continuous beam no matter for how long one observes it.

If the beam is instead made of a finite number N of particles, at a particular time the average distance of the particles in one spiral equals the distance between two adjacent spirals. When this happens, as is shown in the last of Fig. 3, the bucket looks as if it is homogeneously filled with particles and any regular structure due to the initial beam ribbon has disappeared. Thus, for practical purposes, after some time (which depends on N) the beam occupies a new area that is larger and that equals the bucket area. One can reach this conclusion by applying the local average process to define the beam area, once at the beginning when the beam is still debunched and then later when the beam has been bunched.

Second example. If we reverse the process described above, we obtain even further increases of the beam phase-space area if the beam is made of a finite number N of particles. To this purpose, consider the pictures of Fig. 4. We start with some number of beam bunches filling up the corresponding accelerating rf buckets. Then suddenly the rf voltage is turned off and the motion of the particles is changed from circulatory around the center of the bucket to rectilinear along the angle-axis. Suppose that particles with larger momentum move faster than those at lower momentum; then the bunches will elongate leaving their center at rest. At a certain instant, the stretching causes overlapping of neighboring bunches and the beam is observed as debunched in the real space. Actually there is still "rf structure" of the beam in the longitudinal phase space after considerable stretching of the initial bunch ellipses. In fact, if the beam is made of an infinitely large number of particles, the "rf structure" will never disappear. In this case, one can apply the Vlasov equation (2) to calculate the beam shape and infer that the beam area is an invariant. The beam bunches, as shown in Fig. 4, get longer but narrower so that their area at any time equals the initial area they had before starting this debunching process.

On the other hand, if the beam is made of a finite number N of particles, at a certain time the rf structure vanishes. This occurs when the average distance between particles equals the distance between the bunch strips. The time required to reach this situation is called "decoherence time" and clearly depends on N. Thus after the decoherence time, the beam is fully debunched not only in time but also in the longitudinal phase space.[4] Application of the averaging process to determine the beam area shows that the final area is larger than the original one when the beam was still bounded.

With the two examples above, we have shown two cases where the beam phase-space area is not preserved. The reason is the finiteness of the number of particles in the system, which is in contradiction with the major requirement to fulfill for Liouville's theorem: the system must be equivalent to a continuous medium. For those systems where N is finite, it is not always possible to make use of the Vlasov equation (2).

VI. The Stacking and Cooling Techniques

At this point the reader should have a reasonably good idea of what a real physical beam of charged particles looks like and what the implications are of Liouville's theorem in this connection. The most important aspect that one should not forget is that the beam is made of a finite number of particles. The beam area is then defined only as a local average process. Indeed, in practice, beam sizes are measured with devices which count the number of particles in one interval or bin, the equivalent of the cell that we described above.

With this in mind, one should then be able to understand how it is possible to reduce the beam size with "cooling" techniques and yet have Liouville theorem's still apply.

If the beam is made of a finite number of particles, there are large empty regions surrounding the image points of the μ-space (see Fig. 2). There is no reason and no limitation in principle why one cannot fill up these empty regions with more particles if one can find a way. The Liouville theorem would certainly not be contradicted. The question is how it is technically possible to add more particles without perturbing the motion of those that are already there. For instance, if a kicker magnet is used to bring more particles into an area of the phase space already occupied by some particles with the same charge, the same magnet would kick the latter particles out. But if the charge of the particles to be kicked in is opposite to that of the particles already in the storage, then one can manage to kick the entire beam including the fresh pulse by the same amount and in the same direction. This is the principle on which the negative-ion injection is based. One does not "go around" or "beat" Liouville's theorem here; it simply does not apply. If the original beam was made of an infinite large number of particles so that no empty regions in the phase space were available, there would be no way to stack more particles, even with the negative-ion injection method.

Similarly, there is no reason and no limitation in principle why one cannot take a particle at the edge of the beam and place it in an empty region in proximity to the beam center. When this is repeated several times and for all the particles, the beam area can be made as small as wanted, in principle zero. Stochastic cooling is based on this principle. But again if the beam is a continuous medium, that is $N \to \infty$, the reduction of the beam size would not be possible. Indeed it is well known that there is no cooling for $N \to \infty$, since no signal would be provided by the beam (no statistical fluctuations!).

The other two techniques, electron cooling and cooling by synchrotron radiation, are based on entirely different principles than negative-ion injection and stochastic cooling. In these cases, particles suffer energy variations that do not depend on the beam intensity and distribution, but on the properties of the medium they travel through. The motion of these particles then cannot be derived from a Hamiltonian and therefore Liouville's theorem does not apply. One can of course write equations of motion which can again be interpreted as a continuous point transformation in a proper phase space, but now the Jacobian of the transformation is not unity, and phase-space area is not preserved under this transformation. This is true for a continuous system whose distribution function must satisfy a different kind of continuity equation than (2), the Fokker-Planck equation.[5]

There is a difference between the effects of synchrotron radiation and the electron cooling. In the former case, all the particles experience a systematic energy loss which depends on their energy, whereas in the latter particles experience an energy variation which changes sign across an equilibrium value of the particle energies. Because of this difference, in the case of radiation, the energy loss has to be compensated with an external rf cavity, whereas in the electron cooling there is no need of energy compensation. Actually it is well known that it is the addition of the rf cavity that gives synchrotron-radiation damping.[6] But in either case the damping time does not depend on the beam intensity, as in stochastic cooling.

It is not clear whether the dynamics of stochastic cooling can be theoretically described by a continuity equation similar either to the Vlasov equation or to the Fokker-Planck equation, which are based on the assumption of a continuous beam, whereas the principle of the cooling is based on the fact that the beam is made of a finite number of particles.

Acknowledgements

I am very much indebted to F. T. Cole for patiently having read, commented and corrected this paper.

References

[1] A. M. Sessler, contribution to this conference.
[2] J. Liouville, J. Math. Pures. Appl. 2, 16 (1837).
[3] Paul and Tatiana Ehrenfest, The Conceptual Foundations of the Statistical Approach in Mechanics, translated by M. J. Moravesik, Cornell University Press, 1959.
[4] S. Pruss and F. Turkot, private communication, Fermilab, 1976. Such effect has indeed been observed in connection of the slow spill extraction out of the Main Ring. Particles are extracted according to their momentum. Also when the rf was turned off, experimentalists complained about a persistent rf structure in their spill. If the beam, though, was let coast in the ring for a time equivalent to the decoherence time, the rf structure was no longer observed in the extracted beam.
[5] S. Chandrasekhar, Rev. Mod. Phys. 15, 1 (1943).
[6] M. Sands, SLAC-121, Stanford, Nov. 1970.

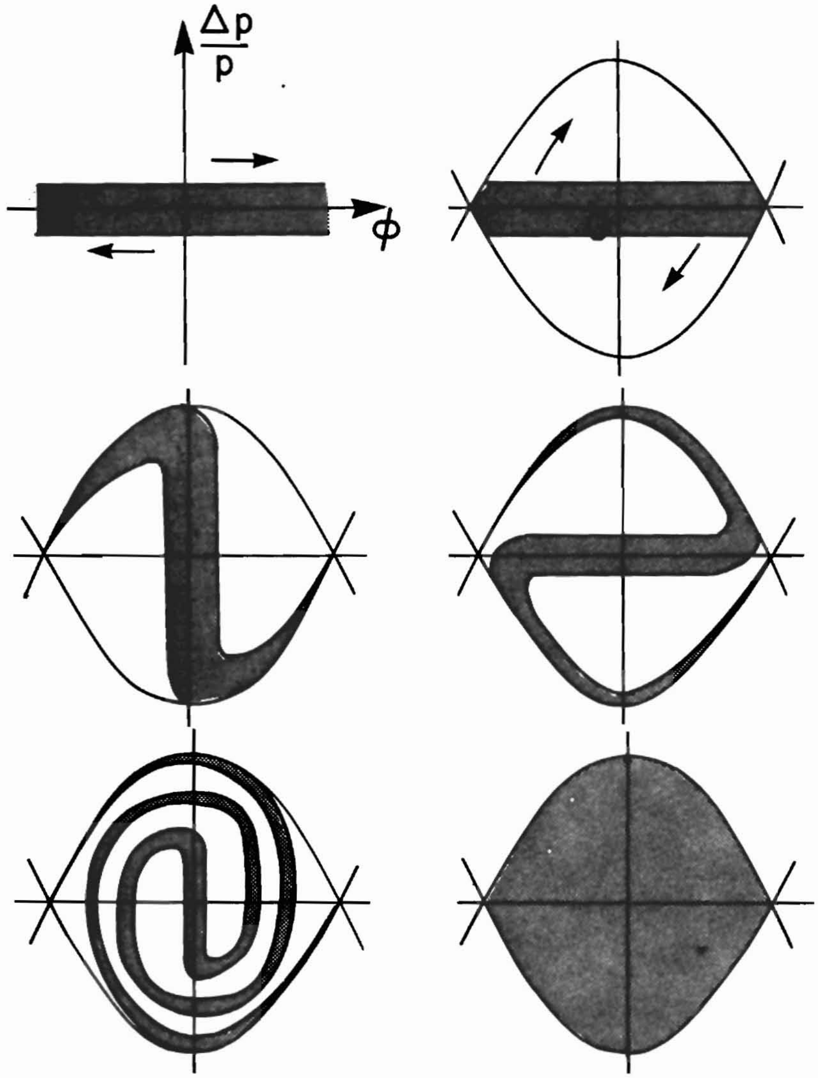

Fig. 3. RF capture.

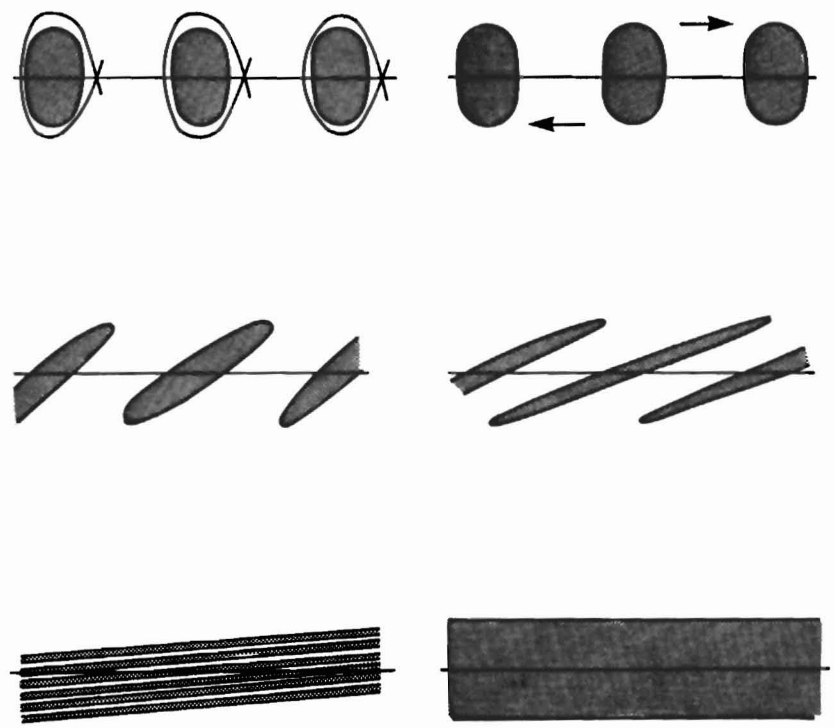

Fig. 4. Beam debunching.