RESEARCH ARTICLE | JUNE 03 2015

Some new surprises in chaos

Leonid A. Bunimovich; Luz V. Vela-Arevalo

Chaos 25, 097614 (2015)
https://doi.org/10.1063/1.4916330

Some new surprises in chaos

Leonid A. Bunimovich[1,2,a)] and Luz V. Vela-Arevalo[2,b)]
[1]*ABC Program, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*
[2]*School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

(Received 30 December 2014; accepted 16 March 2015; published online 3 June 2015)

"Chaos is found in greatest abundance wherever order is being sought.
It always defeats order, because it is better organized"
Terry Pratchett

A brief review is presented of some recent findings in the theory of chaotic dynamics. We also prove a statement that could be naturally considered as a dual one to the Poincaré theorem on recurrences. Numerical results demonstrate that some parts of the phase space of chaotic systems are more likely to be visited earlier than other parts. A new class of chaotic focusing billiards is discussed that clearly violates the main condition considered to be necessary for chaos in focusing billiards. © *2015 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4916330]

Studies of chaotic dynamics continue to bring surprising results by finding new regular features in the dynamics of the most chaotic systems as well as in finding chaos where it should not be according to the existing Chaos theory and intuition. Such findings demonstrate that chaos is even more abundant than previously thought and bring new tools to study chaotic dynamics. We prove a dual statement to the celebrated Poincaré recurrence theorem. It allows us to find new features in chaotic dynamics that lead to finite time predictions of interesting events related to transport in phase space of chaotic systems. Finally, we present a new class of chaotic billiards and explain why it forces us to reconsider the intuition on Hamiltonian dynamical systems that has been building over the last forty years.

I. INTRODUCTION

There is no general consensus about when modern Chaos theory was launched. Some (especially physicists) think that it was marked by the publication of Lorenz's celebrated paper fifty years ago. We do not agree with them because Lorenz's paper remained virtually unknown to the scientific community in the first thirteen years after its publication. In our opinion, Chaos theory started when H. Poincaré made his celebrated mistake (perhaps the most important mistake in the history of Science) that led him to the discovery of what became known as a homoclinic orbit, an unimaginable object so complicated that it was beyond any intuition that existed before its discovery. The history of Chaos theory contains several revolutions that completely changed the understanding of complex dynamics and the ways of thinking about dynamical phenomena. Essentially, the peak of fascination and appreciation of Chaos theory happened at the beginning of the 1990. That is, when this ideology became known to scientists in various areas and to the general public. The ideas and approaches of Chaos theory expanded from physics and mathematics to virtually everywhere. However, already by the mid-1990, this "chaotic" euphoria was essentially over and the general opinion had shifted to the idea that (although mathematicians still could not prove some things) we already knew what to expect. That is, a right intuition on chaotic phenomena was essentially already built and all that was left was to demonstrate that the developed Chaos theory was broadly applicable. So an essential shift towards applications had happened. The purpose of this paper is to demonstrate that chaotic dynamics still brings real surprises. We concentrate on just two topics that could be considered as illustrations of what was said by Pratchett about chaos. The first one demonstrates that the most chaotic dynamics is, in a sense, more ordered than the most regular or integrable dynamics. The second topic demonstrates an enormous abundance of surprising chaotic dynamics that was found there where everything very convincingly appeared to be completely understood after forty years of discovery and numerous theoretical and experimental studies.

II. A COMPLEMENTARY STATEMENT TO THE POINCARÉ RECURRENCE THEOREM

One of the first facts that people learn in nonlinear dynamics is the celebrated Poincaré theorem on recurrences. It states that if a dynamical system has an invariant measure μ of the phase space M, then almost all orbits return infinitely many times to any subset A of M with positive measure, $\mu(A) > 0$.

Suppose that the invariant measure is absolutely continuous (i.e., it has density) with respect to the phase space volume. Then it follows from this theorem that almost all orbits are recurrent in the phase space, that is, almost all orbits return infinitely many times to any arbitrarily small

[a)]Electronic mail: bunimovh@math.gatech.edu
[b)]Electronic mail: luzvela@math.gatech.edu

neighborhood of the point or state at which the orbit started. In other words, almost all orbits are quasi-periodic.

Poincaré theorem thus gives us a clear picture of transport in phase space of dynamical systems with an invariant measure. Those systems can be decomposed into ergodic components. Then, the dynamics could be analyzed on each ergodic component separately because any orbit belongs to only one ergodic component. On each ergodic component, almost all orbits are everywhere dense in the sense that, with probability one, an orbit will visit any set with arbitrarily small but positive measure infinitely many times.

A famous theorem by M. Kac[1] establishes that on each ergodic component B, the average recurrence time, with respect to the invariant measure μ, to any subset A of B equals $\mu(B)/\mu(A)$. Therefore, Poincaré theorem on recurrences together with Kac's theorem seemingly demonstrates the uniformity of transport in each ergodic component in the phase space of a dynamical system. Of course, transport could be quite different on different ergodic components as, for instance, on KAM-tori (as in the Theory of Kolmogorov-Arnold-Moser) and in chaotic sea(s) of a Hamiltonian system. However, in fact, it is not the case. To see that transport is in fact nonuniform, consider another process that also characterizes transport in phase space; namely, instead of recurrences to a subset of positive measure A of the phase space, we will be interested in the moment of the first passage through A, or in other words, the moment of first hitting A.

Let us now be more precise. Suppose that T is a map of the phase space M that preserves a measure μ. For any subset A of M with $\mu(A) > 0$, introduce the following two subsets of M:

$$H_n(A) = \{x : T^n x \in A, T^k x \notin A \text{ for } 0 < k < n\},$$

and

$$S_n(A) = \{x : T^k x \notin A, \text{ for } 0 < k \leq n\}.$$

The first subset $H_n(A)$ consists of all points whose orbits hit or enter the set A for the first time exactly at the moment n. The second subset $S_n(A)$ consists of all points whose orbits hit the set A for the first time after the moment n.

Therefore, $\mu(H_n(A))$ is the probability that the first hitting of the set A will be exactly at time n; while $\mu(S_n(A))$ is the survival probability at the time n, that is, the probability that up to the time n, an orbit will never enter A. Of course, all these probabilities depend on the probability distribution on the phase space. We will consider the most important and natural situation when this probability distribution is a physical invariant measure μ on the phase space, as it is the case of the phase space volume in Hamiltonian systems, Sinai-Bowen-Ruelle (SRB) measures, etc.

The first passage (or the escape) rate $\rho(A)$ through a subset ("hole") A in the phase space is defined via the following relation:[2–4]

$$\rho(A) = \lim_{n\to\infty} -\frac{1}{n} \ln \mu(S_n(A)).$$

It is well known[2] that in strongly chaotic dynamical systems, the survival probability decays exponentially with time, $\mu(S_n(A)) = \exp(-\rho(A)n)$. Therefore, $\rho(A)$ is naturally called, both in physics and mathematics literature, an escape rate for (through) the set A. This term comes from another equivalent interpretation of the first passage problem. One may think about the set A as a "hole" in the phase space. Then, the orbits after coming to A immediately escape from the phase space M. In the interpretation as a first hitting (passage) of A, the orbits that hit A then stop there (at the hitting point) forever. Clearly, these two interpretations (dynamics) are equivalent. The one with holes is used in the theory of open dynamical systems, while the second one in the theory of closed systems.

Theorem (*Dual statement to the Poincaré recurrence theorem*). Let T be an ergodic map of a phase space M with an absolutely continuous invariant measure μ, $\mu(M) = 1$. Suppose that the escape rate is continuous at least at one point. Then for any arbitrarily small positive numbers δ and ε there exists a subset A of M such that $\mu(A) > 1 - \varepsilon$ and $\rho(A) < \delta$, where $\rho(A)$ is the escape (first passage) rate through A.

Proof. Let x be such point of the phase space where the escape rate is continuous. Therefore, there exists a neighborhood $B(x)$ of x such that $\rho(B(x)) < \delta$. It follows from the definition of the escape rate that for any positive integer n, we have $\rho(\cup_{k=0}^{n}(T^{-k}B(x))) = \rho(B(x))$ (see Ref. 5.)

Because T is an ergodic transformation, one can choose n so that the measure $\mu(\cup_{k=0}^{n}(T^{-k}B(x))) > 1 - \varepsilon$, which concludes the proof of the theorem.

Clearly, the proof of this theorem is based on the simple observation that the escape rate through the set $A \cup T^{-1}A$ is the same as the escape rate through the set A.[5] This observation demonstrates certain deficiency of any dynamical characteristic that involves taking the limit when time goes to infinity. Such characteristics could be of great use, but one should remember that by using them we immediately give up any attempt to analyze finite time dynamics.

Remark 1. At first sight, it may seem that the condition of the above theorem is extremely mild and thus is "practically" always satisfied. This view, however, is completely erroneous. In fact, an ergodic dynamics on non-resonant KAM tori is the main example of a dynamical system where the first hitting (escape) rate through any arbitrarily small ball in the phase space is infinite. Therefore, the escape rate is discontinuous at any point of the phase space because, obviously, the escape rate through any point on a KAM torus with Liouvillean (phase volume) invariant measure is zero. This fact is a special case of the theorem that states that for ergodic rotations of a compact metric group the escape rate through any hole that contains an open ball is infinite.[5]

Thus, this dual theorem to the Poincaré recurrence theorem demonstrates that transport in chaotic systems is extremely non-homogeneous. Indeed, for a huge subset of the phase space (that is, a subset with measure arbitrarily close to the measure of the entire phase space,), its escape rate can be arbitrarily small. Of course, these subsets with almost full measure have a complex structure and are, in fact, fractal sets. So, say, in an experimental setup, one cannot really "drill" such hole in order to build an open system with a huge hole but very small escape rate.

To the contrary, transport in systems with very regular (but ergodic) dynamics is much faster, although these systems could even be integrable. Really, the rate of the first visit (passage) of any arbitrarily small ball in the phase spaces of such systems is infinite, while in strongly chaotic systems it is of the order of the measure of the ball (and as such is a very small number). Actually, it is proved[5,6] that in the "most" chaotic systems, the escape rate through an arbitrarily small ball is proportional to the volume (measure) of the ball.

Therefore, in a sense, chaotic dynamics is more organized and predictable than regular dynamics. Indeed, we will explain in Sec. III that for a partition of the phase space of strongly chaotic (or even a random) systems, there is a hierarchy of the first visit probabilities for the elements of the partition. In systems with regular dynamics, all first hitting probabilities are equal for subsets of the same measure. Hence, for chaotic systems, one can predict which one of any two subsets of the phase space will be more likely to be visited first, while for dynamical systems with regular dynamics it is impossible.

Moreover, these predictions seem to be counterintuitive most of the time. For instance, the escape rate through a small hole can be faster than through a bigger one.[5,7] The reason why our intuition turned out to be wrong is that traditionally in the theory of chaos only dynamical characteristics that are asymptotic in time were used and analyzed. However, predictions for finite time evolution of chaotic systems turn out to be sometimes opposite to predictions for infinite time (long time) evolution.[8] We will discuss this issue in Sec. III.

III. HIERARCHY OF THE FIRST PASSAGE PROBABILITIES FOR CHAOTIC AND RANDOM SYSTEMS

Traditionally, Chaos theory deals with long time dynamics as properties of the orbits are basically only considered asymptotic in time (when time goes to infinity). Also, all the classical fundamental characteristics of dynamics involve either a limiting value (when time goes to infinity) or an averaging over an infinite time interval of some observables. A long list of examples includes Lyapunov exponents, Kolmogorov-Sinai entropy, rate of decay of correlations, etc.[9,10]

The first passage (escape) rate through subsets of closed (open) systems also belongs to these infinite time characteristics of dynamics. Indeed, exactly this fact was a cornerstone for the proof of the dual statement to the Poincaré recurrence theorem in Sec. II.

Naturally, the general opinion is that for chaotic dynamics it is impossible to make any type of sensible or interesting finite time prediction. This opinion is based on the fact that chaotic dynamics is, in many respects, similar to the evolution of stochastic (purely random) systems.

However, even for basic classical stochastic systems such as finite Markov chains, some of the problems involving finite time evolution seem to have never been addressed before.[11] Therefore, such natural questions as which state a Markov chain will more likely be in at certain moment of time (if it started, for instance, with a stationary distribution) remained unanswered, again because of the general perception that only studies involving properties that are asymptotic in time may bring sensible and clear answers.

We begin with a discussion of finite time dynamics of arguably the simplest chaotic system: the famous tent map T: $[0, 1] \to [0, 1]$, $Tx = 2x$ if $0 < x \leq 1/2$ and $Tx = 2 - 2x$ if $1/2 < x < 1$. An invariant measure μ for this map is just the length (Lebesgue measure) on [0, 1]. Consider a partition ξ of the phase space into 2^n equal intervals of length 2^{-n}. It was shown in Ref. 5 that in order to determine how the survival probabilities behave for two intervals (elements of the partition) A_1 and A_2, one should just find the periodic orbits with minimal periods Per_1 and Per_2 that intersect these intervals. If the minimal period for interval A_1 is less than for interval A_2, then all survival probabilities after the moment of time $n = Per_1$ are less for an interval that contains a point with larger period.

If instead of survival probabilities we consider first passage probabilities, then for any two intervals of the partition ξ the respective curves of the first hitting probabilities intersect only at one point.[5,7,8] Therefore before the moment in time that corresponds to the point of intersection, it is more likely to first hit an element of the partition for which the curve of the first hitting probabilities was highest. After this moment in time, the forecast about which subset of the phase space will be visited first gets changed into its opposite.[8] To understand this, observe that for any subset A with $\mu(A) > 0$, we have that

$$\sum_{n=1}^{\infty} \mu(H_n(A)) = 1,$$

and therefore the curves of first hitting probability for any two subsets of the phase space of ergodic dynamical or stochastic systems must intersect. If they intersect just at one point (possibly besides the initial interval where they coincide), then short and long term forecast becomes possible in the sense that we can predict which of these two regions of the phase space will be visited first.[8] Also, it ensures the existence of a hierarchy of the first hitting probabilities. In fact, there are two opposite hierarchies, one for intervals of short times and the opposite one for intervals of long times with a relatively short transitional time interval where the short times hierarchy gets transformed into the long times hierarchy.[8]

The most interesting observation here is the possibility of a short term forecast. Such forecast becomes possible because we are looking at the first hitting probabilities rather than escape rates that are a related limiting value as time goes to infinity. In the simple example of the tent map,[5] it can be shown that the escape rate through the longer interval [0,5/16] is smaller than the escape rate through the shorter interval [1/4,1/2]. So, one should deal with such limiting characteristics of dynamics (as time goes to infinity) with great care.

These results were obtained[5,7] for a rather restrictive class of dynamical systems that are called the fair-dice-like systems (see the exact definition in Ref. 7). This class includes the maps $T : [0,1] \to [0,1]$, $Tx = nx \pmod 1$, where n is an integer, the von Neumann-Ulam map, the baker map, etc. All these maps either do not have a distortion or are conjugated to a map without distortion by a nonsingular map, i.e., one for which the image of a positive measure set also has a positive measure. Recall that distortion is the maximal ratio of the (absolute value of the) Jacobians at any two points of the phase space.

Below, we present numerical simulations for chaotic billiards[12] that demonstrate that this non-homogeneity of transport in the phase space is also present for this typical class of chaotic systems.

We will consider a billiard in an asymmetric diamond Q (Fig. 1). The upper and lower sides of the diamond are formed by arcs of circles with radius R and the left and right sides by circles with radius r.

As usual, we will consider a planar billiard map defined on the set of unit vectors on the boundary that are pointing towards the inside of the billiard table. The phase space of this billiard is topologically a cylinder (or an annulus) with length coordinate s, $0 \le s < 1$, and angle coordinate $\varphi, -\pi/2 \le \varphi \le \pi/2$, this is the angle between the velocity vector and the inward unit vector normal to the boundary of the billiard table. We use $\sin\varphi$ instead of φ, as the billiard map in the $(s, \sin\varphi)$ coordinates preserves the phase space volume.

Our billiard belongs to the class of non-uniformly hyperbolic (chaotic) systems[9,10] and therefore there is essential distortion in its dynamics, i.e., the strength of local instability (hyperbolicity) of dynamics varies significantly over the phase space. Two factors mainly contribute to the first hitting probability distribution. One of them is the closeness to periodic orbits with short periods, while the other one is the strength of the local instability, measured by a local Jacobian.

These two factors compete.[5,7,8,13] Indeed, if a set A contains a periodic orbit of period n, then $\mu(A \cap T^n A)$ is the probability to return to A rather than to hit A for the first time at the moment n. Clearly, $\mu(A) = \mu(H_n(A)) + \mu(R_n(A))$, where $R_n(A) = \{x : x \in A, T^n x \in A\}$. Therefore, the faster the orbits return to A, the less the first hitting probability will be. On the other hand, the greater is the local instability of dynamics in A, the smaller is the measure $\mu(A \cap T^n A)$, the probability of recurrence, and therefore the first hitting probability is larger. All these facts are clearly seen in the results of our numerical calculations described below.

Consider first a partition of the phase space of the diamond billiard map into eight subsets of equal measure. The subsets are formed by eight equally spaced intervals in s (the arc length parameter), and $-1 \le \sin\varphi \le 1$. The inset in Fig. 2 shows the billiard table with squares marking the endpoints of the s intervals. The corresponding first hitting probability curves for three of the subsets are depicted in Fig. 2 as a function of time n. The remaining sets have similar curves of first hitting probability due to symmetry. It is easy to see that these curves form two distinct families. The family consisting of the elements of partition that contain the corners gets visited more rarely than the rest of the phase space for short time intervals (small n). It reflects the fact that the coefficients of divergence near corners are smaller and therefore probabilities of short recurrences are larger. Indeed, the time intervals between two consecutive reflections off the boundary are fairly short in the vicinity of the corners.

The lower first hitting probability for subsets defined around the corners of the diamond billiard is also obtained when we consider only the first quadrant of the billiard table (see Fig. 3, the billiard table is showed as an inset). The phase space was partitioned into 7 subsets of equal measure

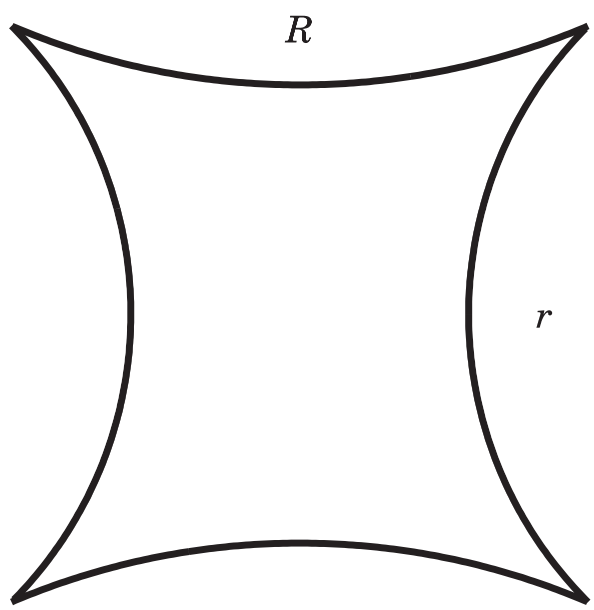

FIG. 1. Diamond billiard. Upper and lower sides correspond to circle arcs of radius R; left and right sides to circle arcs of radius r. In this example, $R = 1.8r$.

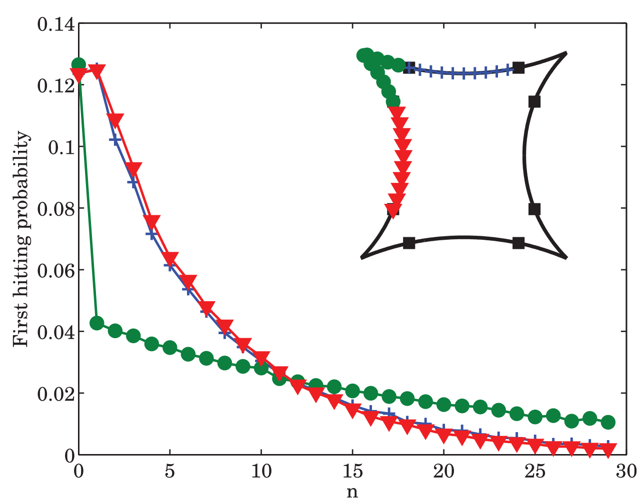

FIG. 2. Curves of first hitting probabilities for the diamond billiard partitioned into 8 subsets. The billiard table is shown in the inset, with markers for three intervals in the arc length parameter s corresponding to three of the subsets, with $-1 \le \sin\varphi \le 1$. For these subsets, the corresponding first hitting probabilities are depicted as a function of time n. The subsets corresponding to the corners have clearly differentiated curves (large dots) from the rest, with lower first hitting probability for short time intervals (small n) and higher for larger times.

with the arc length parameter s of the billiard boundary in equally spaced intervals, and $\sin\varphi$ in $[-1, 1]$. The curve of first hitting probability corresponding to a subset containing the upper right corner (marked with dots in Fig. 3) is lower than the other subsets for short time intervals (for small n).

In Fig. 4, a different partition was used for the first quadrant diamond billiard. We considered 3 subsets of equal measure formed by 3 equally spaced intervals in the boundary arc length parameter s, but we neglected reflections off the symmetry lines in the x and y axes. The angle variable is $\sin\varphi \in [-1, 1]$. In the resulting first hitting probabilities (Fig. 4), the curve corresponding to the subset that contains the corner is again differentiated from the rest.

Further splitting of curves of first hitting probability is obtained when we consider partitions of the phase space that have more specific reflection angles by defining subintervals of both s and φ. The elements of the partition used for Fig. 5 correspond to 8 equally spaced intervals in s, and $0 \leq \sin\varphi < 1/2$ or $1/2 \leq \sin\varphi \leq 1$. In this figure, c_1, c_2, and c_3 represent subsets for the given arc segments in the boundary, and $\sin\varphi \in [0, 1/2)$; while c_4, c_5, and c_6 represent the indicated boundary segments, and $\sin\varphi \in [1/2, 1]$. We notice that the subsets that contain the diamond corners have noticeably lower first hitting probability for short times (segments marked as c_2 and c_5 with large dots and stars), but there is splitting between these curves when the different reflection angles are introduced. Also, the splitting of the curves corresponding to subsets with boundary arc segments c_1 and c_4 (top and bottom of the diamond) demonstrates that symmetry is broken when different reflection angles are considered.

We next calculate first hitting probabilities associated with subsets of the phase space defined around periodic orbits with short periods. In the diamond billiard, the periodic orbits with shortest period are the period-2 horizontal and vertical periodic orbits. The rhombus periodic orbits (one clockwise and one counterclockwise) are also present in the diamond billiard and have period 2, and there are some other period-4 orbits that will be discussed later. We define equally spaced subsets of the phase around the period-2 orbits and one rhombus orbit, and for comparison we consider also subsets that include the corners. Fig. 6(a), shows the subsets in phase space in coordinates s, $\sin\varphi$. In Fig. 6(b), the inset shows the diamond billiard table and the periodic orbits and corner segments defined to define the subsets. The markers for the different sets in Fig. 6(a) are represented in Fig. 6(b): dots for the horizontal period-2 orbit, squares for the vertical period-2 orbit, dotted lines for the rhombus, and triangles for subsets containing the corners. Again, the subsets around corners show lower first hitting probability for shorter times; while the subset containing the horizontal period-2 periodic orbit has higher first hitting probability for shorter times. As usual the situation is reversed for longer times.

In Fig. 7, we compare the first hitting probabilities of diamond billiards with slightly different geometries. We vary

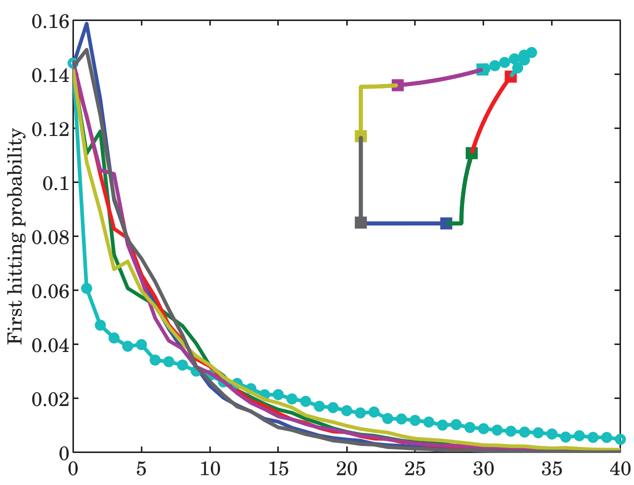

FIG. 3. First quadrant diamond billiard partitioned into 7 subsets of equal measure defined with the arc length parameter s in equally spaced intervals (see inset) and the probabilities of first hitting each subset at time n. The curve corresponding to the subset containing the upper right corner is marked with large dots, and it shows lower first hitting probability for small n and the opposite for large n.

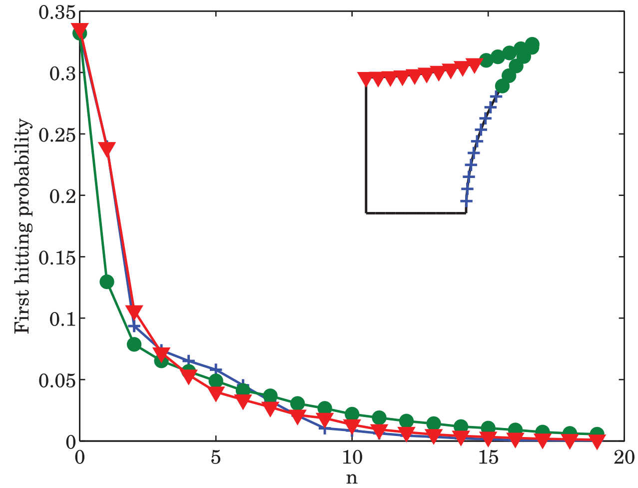

FIG. 4. Probability of first hitting subset i at time n in the first quadrant diamond billiard. The subsets are formed by equally spaced intervals of the boundary arc length parameter s, and reflections off the symmetry lines (the x, y axes) were neglected, $\sin\varphi \in [-1, 1]$.

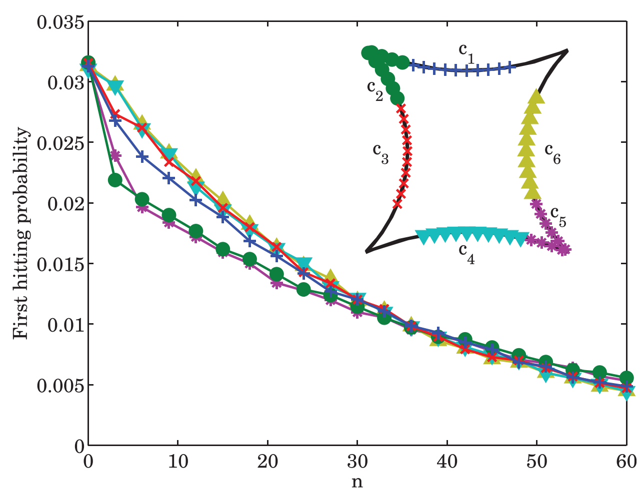

FIG. 5. Probability of first hitting for the diamond billiard for subsets defined with the boundary arc length parameter s in 8 equally spaced intervals; and reflection angle $\sin\varphi$ in [0, 1/2) for the sets with boundary arc segments marked with c_1, c_2, c_3, and $\sin\varphi$ in [1/2, 1] for c_4, c_5, c_6.

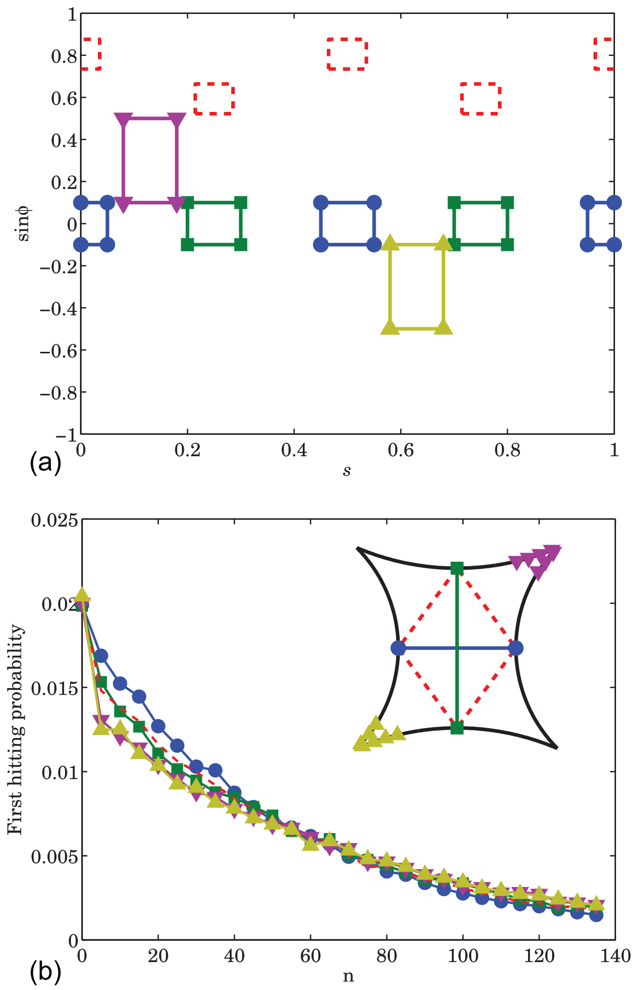

FIG. 6. First hitting probability at time n of subsets defined around periodic orbits and corners. We considered the period-2 horizontal and vertical periodic orbits, and a rhombus period-4 orbit (clockwise). Two more subsets around corners were considered. In (a), the subsets are shown in phase space; in (b), the curves of first hitting probability are shown and the inset has the billiard table and the configuration of the periodic orbits and boundary arc segments considered.

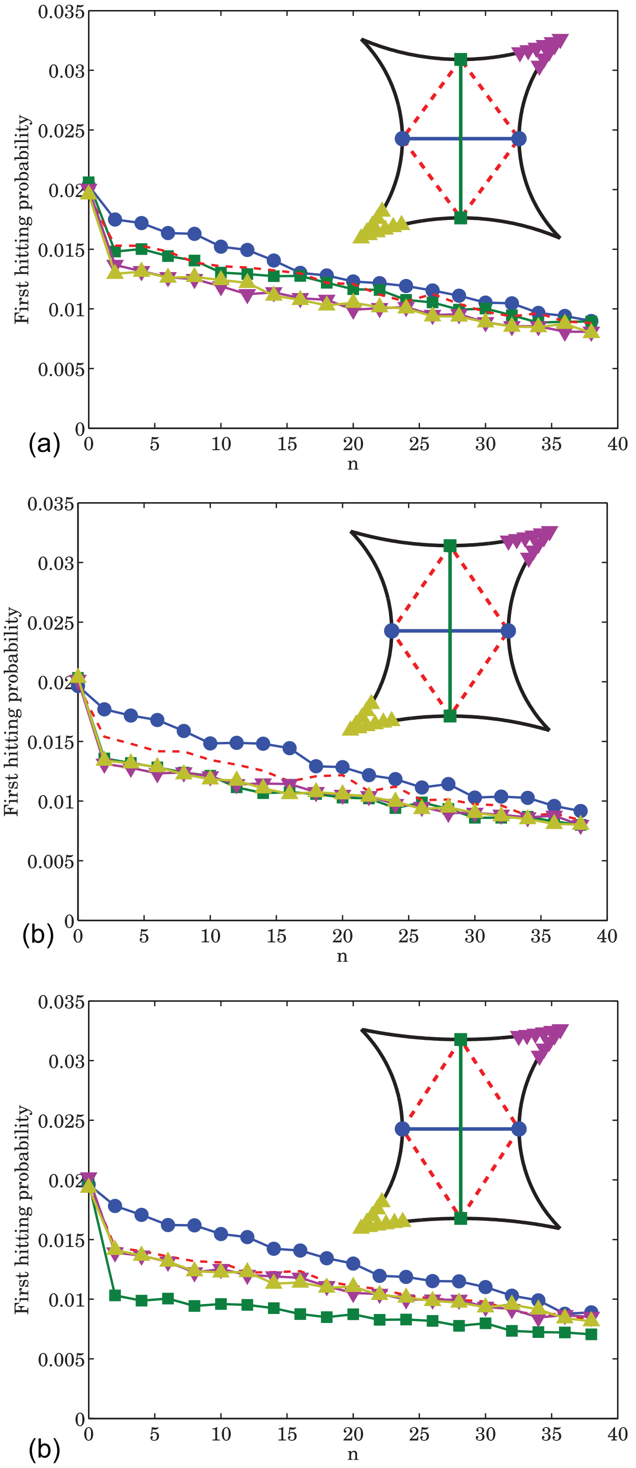

FIG. 7. First hitting probability at time n of subsets defined around periodic orbits and corners in the diamond billiard for three different configurations: (a) $R = 1.8r$, (b) $R = 2.5r$, and (c) $R = 3.5r$. In all cases, the subsets were defined analogously to the ones used in Fig. 6 around the period-2 orbits, the rhombus period-4 orbit, and corners of the diamond.

the ratio r/R of the radius of the circles used to construct the diamond (see Fig. 1), and consider the same type of subsets around periodic orbits and corners as in Fig. 6. The resulting curves of first hitting probability are shown in Fig. 7, together with the billiard tables in the insets, for the cases (a) $R = 1.8r$, (b) $R = 2.5r$, and (c) $R = 3.5r$.

It is interesting to notice in Fig. 7 that changing the geometry of the diamond billiard produces different behaviors of the first hitting probability mainly for the subset around the vertical period-2 periodic orbit (with square markers in the figure.) For $R = 3.5r$, the first hitting probability curve corresponding to the vertical orbit is much lower than the rest of the subsets considered, for small n. As usual, we expect that the first hitting probability curves will cross each other at some point n and the lowest first hitting probability corresponding to the vertical orbit will be the highest for larger times. This is observed in Fig. 8, where the same calculation as for Fig. 7(c) is shown but for longer times.

We next consider the remaining period-4 orbits and define subsets around them. The periodic orbits and the first hitting probabilities for these subsets are shown in Fig. 9. The inset shows the diamond billiard table and the periodic orbits of period 4 that were considered: we called them a "V" orbit (square markers), an "L" orbit (large dots), and one of the rhombus orbit (clockwise, marked in dotted lines.) We defined subsets around these period-4 orbits and also two subsets that include corners of the diamond. For the

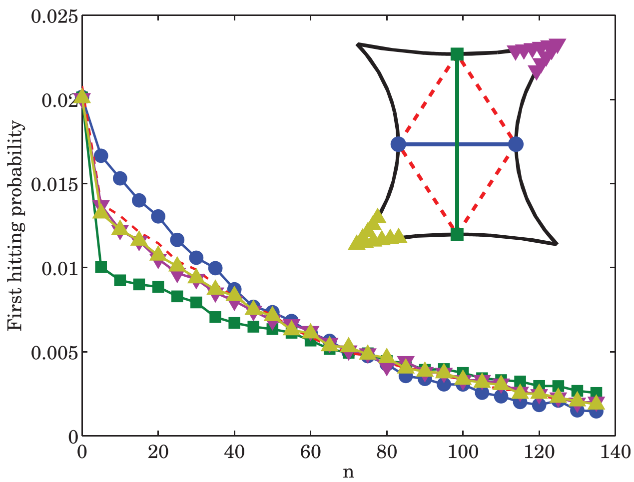

FIG. 8. First hitting probability at time n of subsets defined around periodic orbits and two corners for a diamond billiard with upper and lower circle arc radius of $R = 3.5r$. The parameters and subset definitions are the same as in Fig. 7(c), but this figure shows the point where the first hitting probability curves cross each other and the curve corresponding to the vertical periodic orbit switches from the lowest to the highest.

calculation of the first hitting probabilities, we used rectangles to define the subsets that are smaller than in the calculations previously presented; this produced more noticeable splitting of the curves, but required more computational time. In Fig. 9, part (a) corresponds to a diamond billiard with $R = 1.8r$ and (b) to $R = 3.5r$. For the billiard in (b), one can see that most probabilities are lower than the case (a), except for the first hitting probability corresponding to the "L" orbit that seemed to remain unchanged. Also, notice the relative position of the first hitting probability corresponding to the "V" orbit, it is lower for case (a), and in the middle for case (b); this is for small times n, before the behavior is reversed after the first hitting probability curves cross each other.

To analyze further the splitting of first hitting probability curves for regions around the period-2 orbits, we selected subsets of the phase space around one single reflection point of the periodic orbits. This is different from the previous calculations in that we had considered rectangles around all reflection points of the periodic orbits (see Fig. 6). The first hitting probability curves corresponding to these smaller subsets are shown in Fig. 10, for two different diamond billiard tables: (a) $R = 1.8r$ and (b) $R = 3.5r$. We observe that the first hitting probability for the region around the vertical period-2 orbit is lower in both cases, but it is much lower for case (b); this is for small n before the curves cross and their behavior switches for larger n.

The results could be easily explained if we compare expansion coefficients (strengths of local instability) at these two period-2 orbits. Recall[14] that these coefficients are $1 + k\tau$, where k is the curvature of the boundary and τ is the length of the free path between two consecutive reflections off the boundary. For the billiard tables in Fig. 10, we obtained that the expansion coefficient for the vertical period-2 orbit in case (a) is 1.3214, while for case (b) the expansion coefficient is 1.1814. As was explained above, the faster the orbits return to a subset, the less will be the first

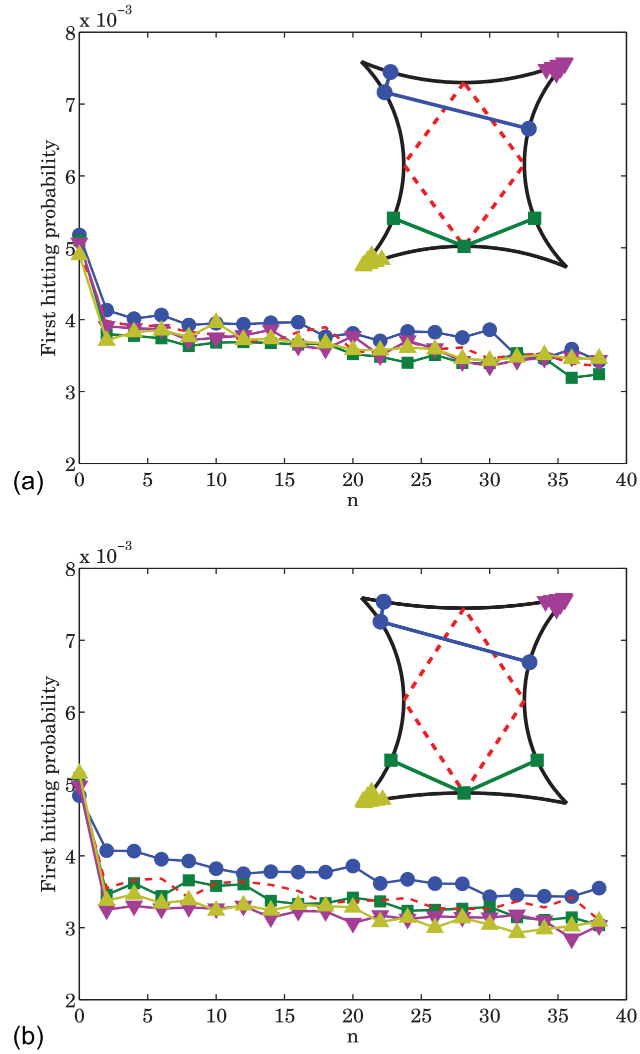

FIG. 9. First hitting probability at time n of subsets defined around period-4 periodic orbits: one "L" orbit (dots), one "V" orbit (squares), and a rhombus orbit (dotted line); we also considered subsets around two corners. Two different diamond billiard tables were used: (a) $R = 1.8r$ and (b) $R = 3.5r$.

hitting probability; clearly, the lower expansion coefficient in (b) results in faster returns around the vertical periodic orbit, and hence smaller first hitting probability.

The numerical results demonstrate that it is possible to attempt short time forecast of transport in phase space of chaotic systems even with the most complex dynamics (highly nonuniform hyperbolicity). For short times, we are able to tell if it is more likely to visit some subset A for the first time than other subset B, and so it is also less likely to return soon to subset A than to B.[8,15] This short time forecast is qualitatively opposite to long time forecast. For the same subsets A and B, for long times, it will be less likely to visit the subset A for the first time than the subset B, and it is more likely to return soon to A than to return to B.

IV. CHAOS SHOWS UP WHERE ORDER HAS ALWAYS BEEN THOUGHT

It is well known and understood that chaotic behavior of deterministic dynamical systems is generated by strong intrinsic instability of dynamics.[9,10] By this intrinsic

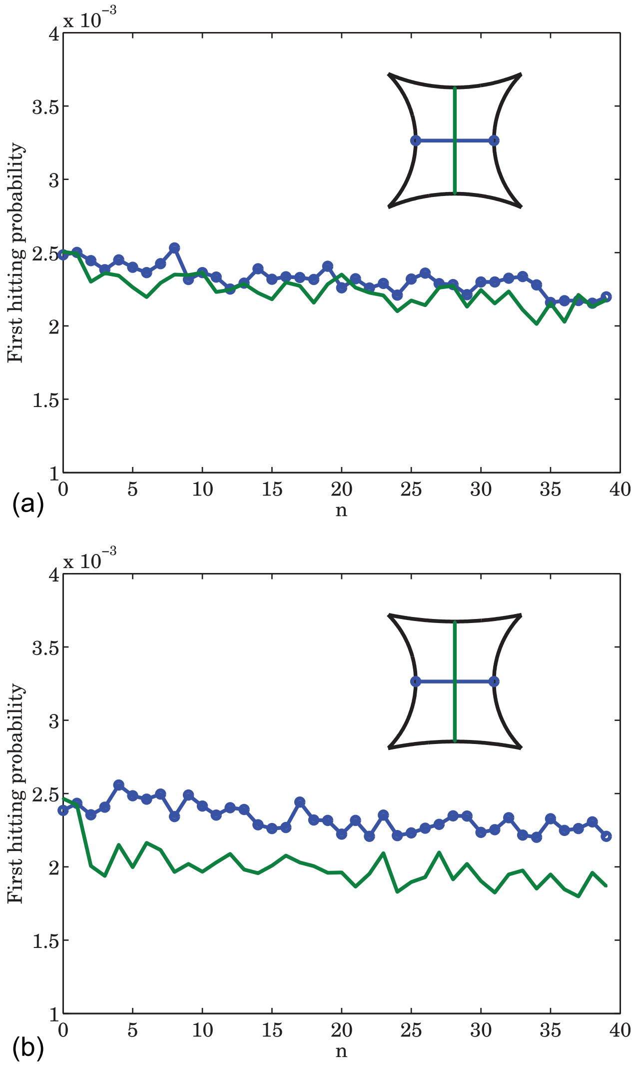

FIG. 10. First hitting probability at time n of subsets defined around a single reflection point of the shortest periodic orbits of the diamond billiard: the horizontal period-2 orbit (dots), and the vertical period-2 orbit (line). (a) $R = 1.8r$ and (b) $R = 3.5r$.

instability, two trajectories that are initially close to each other are forced to diverge in phase space. This divergence must be on average exponential, where "on average" may mean with respect to certain (probability) measure on the phase space, with respect to time, etc. Therefore, Lyapunov exponents in chaotic systems must be nonzero. Systems that are mostly chaotic are called hyperbolic dynamical systems because local behavior of their orbits resembles the behavior of orbits near a saddle fixed point. Clearly, if the phase space is compact (and therefore has finite volume or measure), then nearby orbits can diverge only locally because eventually they must come close to each other again and again.

Historically, chaotic behavior was first detected and proved almost a century ago for a class of Hamiltonian systems formed by the geodesic flows on surfaces of negative curvature;[16,17] such surfaces were subsets of constant energy of closed Hamiltonian systems. Meanwhile, classical examples of integrable geodesic flows on surfaces of positive curvature demonstrated that positive curvature generates extremely regular dynamics; an example of this is Jacobi's proof of integrability of a geodesic flow on ellipsoids.

The seminal paper by Y. G. Sinai[14] on Sinai billiards laid the foundation for rigorous studies of nonuniformly hyperbolic dynamical systems and demonstrated that a dispersing boundary, that is, a billiard table that is convex inward, plays an analogous role in billiards to the one played by negative curvature in the geodesic flows.

Again, there were already known classical examples of integrable billiards in ellipses, i.e., billiard tables with focusing boundary.

From a general point of view, billiards can be seen as geodesic flows on manifolds with a boundary (non-closed manifolds). A negative (positive) curvature for geodesic flows corresponds to dispersing (focusing) boundary in billiards. The mechanism of intrinsic instability of dynamics that generates chaotic behavior in these (as well as in other) classes of dynamical systems is naturally called the mechanism of dispersing. Therefore, a general consensus in physics and mathematics communities was that the presence of negative curvature always generates chaotic behavior in geodesic flows as well as that dispersing boundary always generates chaotic behavior in billiards; meanwhile, positive curvature always tends to generate regular dynamics in geodesic flows, as focusing boundary generates regular dynamics in billiards.

This general philosophy or understanding of the nature of dynamics was strongly confirmed by the results of V. F. Lazutkin,[18] who proved that a billiard in a convex billiard table with sufficiently smooth boundary has an uncountable set of caustics that converge to the boundary of the billiard table. Recall that a curve Γ is a caustic for a billiard if any orbit tangent to Γ between two consecutive reflections off the boundary is also tangent to the same curve Γ between any other pair of reflections. The simplest family of caustics is formed by concentric circles for the billiard in a circle. Observe that if just one caustic exists, the corresponding billiard is already non-ergodic because the set of orbits that intersect this caustic and the complement to this set are both invariant subsets of the phase space with positive measure (phase space volume).

Therefore, it came as a complete surprise when soon after another mechanism of chaos was discovered[19,20] that is different from the one in geodesic flows and in billiards with dispersing boundaries. If orbits in a parallel beam reflect off a dispersing boundary, they become divergent; if a parallel beam of orbits reflects off a focusing boundary, the orbits become convergent, see Figs. 11(a) and 11(b). It seems then that orbits get closer to each other because of focusing, while they get farther away from each other because of dispersing.

Observe, however, that if after a reflection off a focusing boundary there is a free path sufficiently long so that the orbits have enough time to pass their point of intersection, the orbits will diverge (Fig. 11(c).) Therefore, effectively, the distance between these orbits (rays) grows in time and such dynamics is intrinsically unstable. This mechanism of chaos is called the mechanism of defocusing.

Historically, the first example of a focusing billiard with chaotic dynamics is depicted in Fig. 12(a). This billiard table

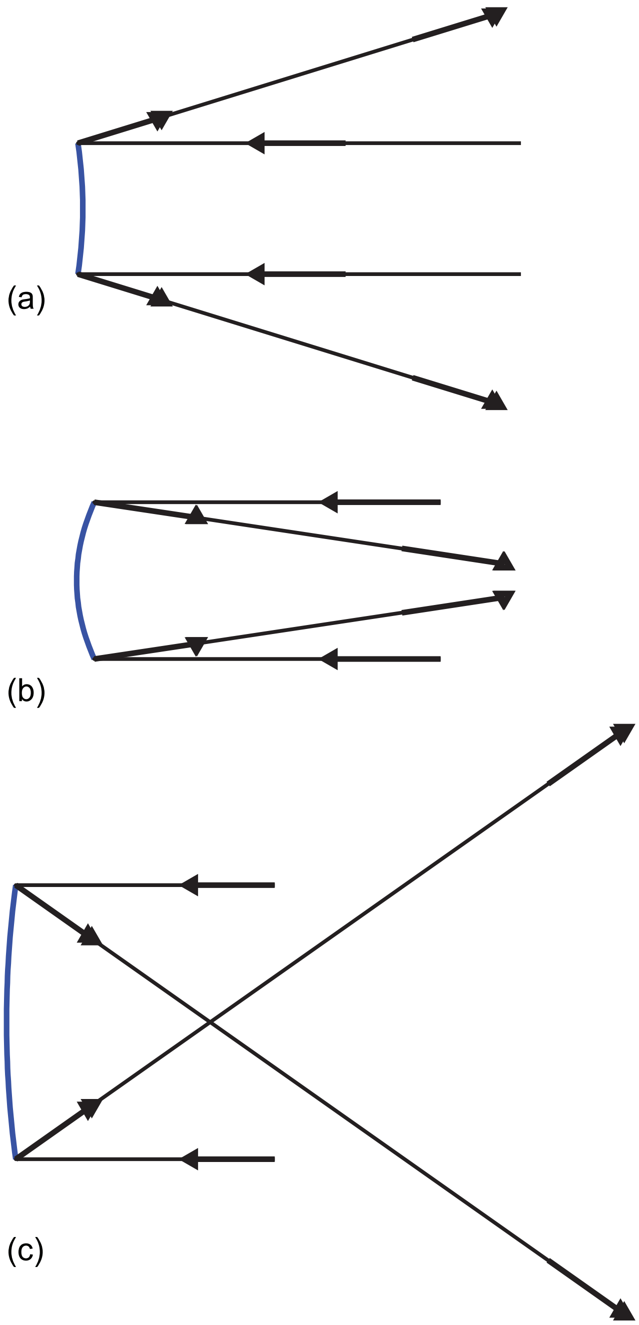

FIG. 11. (a) Dispersing billiard, (b) and (c) focusing billiards.

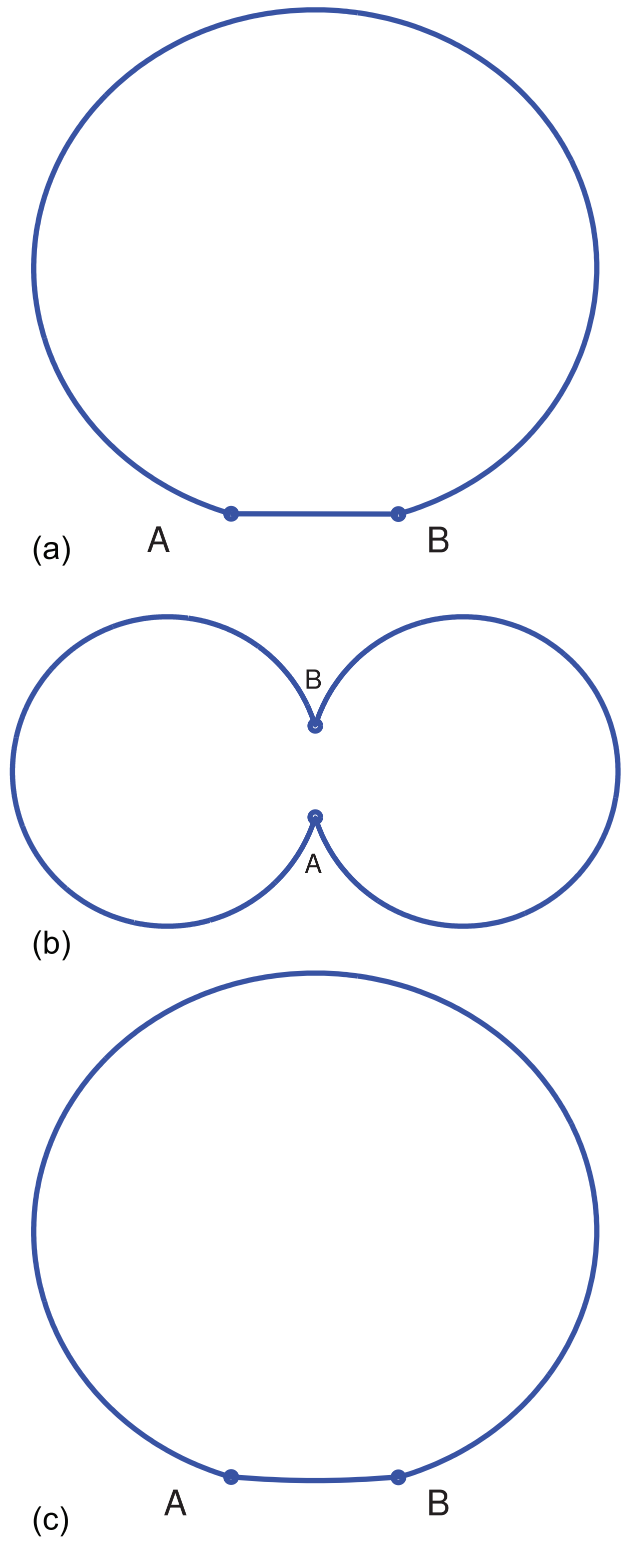

FIG. 12. (a)–(c) Chaotic focusing billiards.

is obtained by cutting a circle in two pieces by a chord AB (which must be shorter than the diameter of the circle) and taking the larger piece as the billiard table.

It is easy to see that the dynamics of the billiard in Fig. 12(a) is completely equivalent to the dynamics the billiard in Fig. 12(b): the billiard table in Fig. 12(b) is obtained by reflecting the table in Fig. 12(a) through the straight segment AB. In geometric optics, this method of reflections is called the method of constructing images. It is clear that the equivalent billiard table in Fig. 12(b) has a set of sufficiently long free paths for the defocusing mechanism to take place, and hence chaotic dynamics is generated.

In the next forty years after the discovery of the defocusing mechanism of chaos, a lot of work has been done to gain better understanding of it, both physically and mathematically. Many new examples of chaotic focusing billiards were built as well as chaotic geodesic flows on manifolds with non-negative curvature, where chaos is generated by the defocusing mechanism. There were claims that the defocusing mechanism would work only in 2D because in higher dimensions astigmatism weakens the defocusing in some planes. It was proved, though, that the mechanism of defocusing produces chaotic billiards in all finite dimensions.[21] A lot of studies were performed on the theory of quantum chaos, and many experimental devices with chaotic focusing billiards were built.

However, it was assumed (consciously or unconsciously) that the following condition was necessary for the defocusing mechanism to generate chaotic dynamics: for a point at a focusing boundary with curvature k, the entire circle with radius $1/k$ tangent to the boundary at that point (the osculating circle) must be contained in the billiard table. This condition was satisfied by the existing examples, but the condition seems to be violated already in the example in Fig. 12(a). However, the condition is meant for "efficient" billiard tables that are obtained after reflecting the billiard table over all neutral (zero curvature) components of the boundary. For the table in Fig. 12(a), the corresponding efficient table (Fig. 12(b)) obviously satisfies the condition that all osculating circles belong to the billiard table.

Indeed, it seems that this condition is at the heart of the mechanism of defocusing because a billiard in a circle is integrable and we need to give rays more time to diverge. It goes along with the intuition developed in the last 40 years about the mechanism of defocusing. However, the following example shows that this condition could be completely violated by convex focusing billiards which may still be chaotic.

Consider the first focusing chaotic billiard depicted in Fig. 12(a). Make now a small perturbation of the boundary of this billiard by substituting the chord AB by an arc of a circle with radius R so big that its center is located outside the billiard table (Fig. 12(c).) It is easy to see that this happens when $r < b < R$, where b is the distance between the centers of the two circles and r is the radius of the smaller circle.

In the billiard table in Fig. 12(c), both circles of radius r and R contain the entire billiard table. Therefore, the main condition of having a sufficiently long free path for the mechanism of defocusing to work is completely violated. However, the corresponding billiard is chaotic (hyperbolic).[22]

Moreover, this class of billiards has another striking feature. A basic fascinating property of chaotic (hyperbolic) dynamical systems is that local exponential instability (hyperbolicity) ensures global chaotic properties of dynamics like positivity of Kolmogorov-Sinai entropy, mixing (decay of correlations) on each ergodic component, etc. This condition of local instability must hold at all (or at almost all) points, i.e., on a set of measure one in the phase space.

Then is a complete surprise that hyperbolicity (dynamical chaos) follows from just the condition that $r < b < R$ for the class of billiards depicted in Fig. 12(c). Formally, this condition ensures that the periodic orbit of period two that goes back and forth between the middle points of the arcs of circles of radius r and R is unstable, and hence there is no KAM island around these two points in the phase space. However, the fact that only this clearly local condition precludes the existence of other KAM islands in the phase space still looks like a miracle, although proven.[22]

Observe that the billiard table in Fig. 12(c) can be considered an asymmetric lemon (Fig. 13). Recall that a

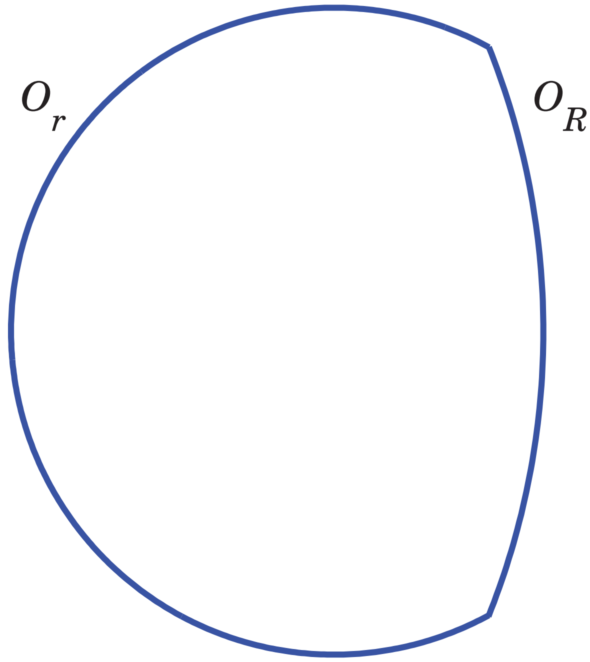

FIG. 13. Asymmetric lemon billiard.

symmetric lemon billiard is obtained by taking the intersection of two circles with the same radius. Lemon billiards form a popular object in quantum chaos.[23] There is little doubt that better understanding of the dynamics in asymmetric lemons and their perturbations will shed new light on our understanding of classical and quantum chaos. The first immediate question is whether this class of chaotic billiards is a very specific one because of some "symmetry" that allows the mechanism of defocusing to produce chaotic dynamics when there is less room for that than in the completely integrable billiard in a circle. Our conjecture is that the answer is "no" and one can perturb the asymmetric lemon's boundary while preserving dynamical chaos.

To conclude, we would just like to mention that Chaos theory is very much alive. It keeps producing surprises that must be better understood, even if the existence of these surprises is rigorously proved. There is no doubt that new surprises are coming, and Chaos theory will continue to be a wonderful rare example of fruitful close collaboration between physicists and mathematicians.

[1]M. Kac, "On the notion of recurrence in discrete stochastic processes," Bul. Am. Math. Soc. **53**, 1002–1010 (1947).
[2]M. Demers and L. S. Young, "Escape rates and conditionally invariant measures," Nonlinearity **19**, 377–397 (2006).
[3]M. Demers, P. Wright, and L. S. Young, "Escape rates and physically relevant measures for billiards with small holes," Commun. Math. Phys. **294**(2), 353–388 (2010).
[4]M. Demers and P. Wright, "Behavior of the escape rate function in hyperbolic dynamical systems," Nonlinearity **25**, 2133–2150 (2012).
[5]L. A. Bunimovich and A. Yurchenko, "Where to place a hole to achieve a maximal escape rate," Isr. J. Math. **182**, 229–252 (2011).
[6]G. Keller and C. Liverani, "Rare events, escape rates and quasistationarity: Some exact formulae," J. Stat. Phys. **135**(3), 519–534 (2009).
[7]L. A. Bunimovich, "Fair dice-like hyperbolic systems," Contemp. Math. **567**, 79–87 (2012).
[8]L. A. Bunimovich, "Short- and long-term forecast for chaotic and random systems," Nonlinearity **27**(9), R51–R59 (2014).

[9]I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, "Ergodic theory," in *Fundamental Principles of Mathematical Sciences* (Springer-Verlag, NY, 1982), Vol. 245.
[10]E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1999).
[11]Y. Bakhtin and L. A. Bunimovich, "The optimal sink and the best source in a Markov chain," J. Stat. Phys. **143**, 943–954 (2011).
[12]N. Chernov and R. Markarian, *Chaotic Billiards*, Mathematical Surveys and Monographs (AMS, 2006), Vol. 127.
[13]G. Cristadoro, G. Knight, and M. Degli Esposti, "Follow the fugitive: An application of the method of images to open systems," J. Phys A **46**, 272001–272008 (2013).
[14]Y. G. Sinai, "Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards," Russ. Math. Surv. **25**, 137–189 (1970).
[15]V. S. Afraimovich and L. A. Bunimovich, "Which hole is leaking the most: A topological approach to study open systems," Nonlinearity **23**, 643–656 (2010).
[16]G. A. Hedlund, "The dynamics of geodesic flows," Bull. Am. Math. Soc. **45**, 241–245 (1939).
[17]E. Hopf, "Statistik der loesungen geodaetischen linien in mannigfaltigkeiten negativer kruemmung," Ber. Verh. Saechs. Akad. Wiss. Leipzig **91**, 261–304 (1939).
[18]V. F. Lazutkin, "Existence of a continuum of closed invariant curves for a convex billiard," Math. USSR Izvestija **7**, 185–214 (1973).
[19]L. A. Bunimovich, "On billiards closed to dispersing," Mat. Sb. **95**, 49–73 (1974).
[20]L. A. Bunimovich, "On ergodic properties of certain billiards," Funct. Anal. Appl. **8**, 254–255 (1974).
[21]H.-J. Stockmann, *Quantum Chaos, An Introduction* (Cambridge University Press, Cambridge, 1999).
[22]L. A. Bunimovich, H. Zhang, and P. Zhang, "On another edge of defocusing: Hyperbolicity of asymmetric lemon billiards," e-print arXiv:1126508.
[23]E. Heller and S. Tomsovic, "Postmodern quantum mechanics," Phys. Today **46**(7), 38–46 (1993).