Algebraic Statistics of Poincaré Recurrences in a DNA Molecule

Alexey K. Mazur  
UPR9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, Paris, 75005, France

D. L. Shepelyansky  
Laboratoire de Physique Théorique du CNRS (IRSAMC), Université de Toulouse, UPS, F-31062 Toulouse, France

(Received 9 August 2015; published 30 October 2015)

The statistics of Poincaré recurrences is studied for the base-pair breathing dynamics of an all-atom DNA molecule in a realistic aqueous environment with thousands of degrees of freedom. It is found that at least over five decades in time the decay of recurrences is described by an algebraic law with the Poincaré exponent close to \( \beta = 1.2 \). This value is directly related to the correlation decay exponent \( \nu = \beta - 1 \), which is close to \( \nu \approx 0.15 \) observed in the time resolved Stokes shift experiments. By applying the virial theorem we analyze the chaotic dynamics in polynomial potentials and demonstrate analytically that an exponent \( \beta = 1.2 \) is obtained assuming the dominance of dipole-dipole interactions in the relevant DNA dynamics. Molecular dynamics simulations also reveal the presence of strong low frequency noise with the exponent \( \eta = 1.6 \). We trace parallels with the chaotic dynamics of symplectic maps with a few degrees of freedom characterized by the Poincaré exponent \( \beta \sim 1.5 \).

PACS numbers: 87.15.ap, 05.45.Ac, 05.45.Jn, 87.14.gk

DOI: 10.1103/PhysRevLett.115.188104

The celebrated Poincaré recurrence theorem of 1890 [1] guarantees that a dynamical trajectory with a fixed energy and bounded phase space will always return in a close vicinity of the initial state. For dynamical systems with hard chaos the statistics of Poincaré recurrences, and the related probability to stay in a bounded phase space region, behave similarly to coin flipping and drop exponentially with the return time \( \tau \) [2,3]. However, in the generic case of chaos with divided phase space, when islands of integrable motion are embedded in a chaotic sea [4,5], it was established that the probability distribution of recurrences \( P(\tau) \) is described by an algebraic decay

\[
P(\tau) \propto 1/\tau^\beta,
\]

with the Poincaré exponent \( \beta \sim 1.5 \) [6–11]. This slow decay originates from sticking of dynamical trajectories in the vicinity of stability islands and results in a slow decay of the corresponding autocorrelation function \( C(\tau) \propto \tau P(\tau) \) [10]. Most of the studies of the algebraic statistics of Poincaré recurrences (ASPR) considered 2D symplectic maps, notably, the Chirikov standard map, with recurrence times changing by more than 10 orders of magnitude [12]. A few recent studies of Hamiltonian systems with a larger number of degrees of freedom also revealed an algebraic decay of recurrences with similar values of the Poincaré exponent \( \beta \sim 1.3–1.5 \) [13–16].

The generic nature of the ASPR phenomenon is well established. It is known to occur on a huge range of physical scales from electron trajectories for microwave ionization of Rydberg atoms [17,18] to comet orbits in the Solar System [19]. One should expect that it is also inherent in conformational dynamics of macromolecules. These systems are characterized by complex energy landscapes, with numerous barriers and saddle points crossed during thermal motion. It is tacitly assumed that such dynamics results in a developed chaos with multiexponential relaxation decay. Recent experimental evidence suggests, however, that the ASPR may play an important role, notably, in the behavior of the double helical DNA. The power-law relaxation in the B-DNA double helix was discovered [20,21] and carefully studied during the last decade by using the time resolved Stokes shift (TRSS) experiments [22–27]. In the last years these results were analyzed with different theoretical approaches [28–34], but, nevertheless, the possible underlying molecular mechanism remains elusive. By analogy with the long known power-law kinetics in proteins [35], this effect in DNA is interpreted in terms of models developed earlier for glassy systems, with multiple substates, hierarchical relaxation, mode coupling, etc. However, unlike proteins, the B-DNA molecule has only a few well-studied conformational substates with relatively fast and spatially localized dynamics that does not resemble those in glasses. With hydration water and surrounding ions included, the system becomes more complex, but, in spite of all efforts, there is no agreement even on whether the power-law relaxation in the submicrosecond time range is due to DNA itself or the hydration water, or both [27,32,36].

We asked, what if, instead of the spin-glass–like effect of multiple degrees of freedom, the power-law relaxation in DNA represents a manifestation of the ASPR phenomenon. According to experiment, the decay of the TRSS signal in...
DNA is described by time autocorrelation function $C(\tau) \propto 1/\tau^{\nu}$ with $\nu \approx 0.15$ [21], that is, rather close to ASPR in systems with a few degrees of freedom. This decay is observed over six decades in time from $10^{-13}$ to $10^{-7}$ sec; therefore, one can hope to detect it by means of all-atom molecular dynamics (MD) simulations. The TRSS signal is obtained by substituting a polarity sensitive dye coumarin for one of the stacked bases, and the measured effect can be due to any motion that changes the dye local electric field [20]. Since in MD different motions are coupled we started from a trial search of relevant parameters. The statistics of Poincaré recurrences for kinetic energies of selected atoms revealed only exponential decay. A similar behavior, except for minor details, was found for conformational transitions of backbone torsion angles. After a few unsuccessful trials, however, the ASPR has been revealed in the base-pair breathing motion.

The base-pair breathing occurs due to temporal breaking of one or more hydrogen bonds (H bonds) in a Watson-Crick pair. The statistics of Poincaré recurrences was studied for the model system shown in Fig. 1. On the left, a GC pair is displayed with three H bonds. The right panel shows a tetramer duplex formed by two strands with identical GC-alternating sequences. This is a minimal symmetrical structure with two external and two internal base pairs. The duplex is placed in a small water box of 489 water molecules with periodic boundaries. Six sodium ions are added for neutralization. All-atom MD simulations were carried out under normal temperature using internal coordinate Hamiltonian equations [37] and a symplectic integrator [38,39] with the time step of 0.01 ps [40,41]. The recent version of the AMBER force field [42] was used with SPC/E water [45]. The system had 3226 degrees of freedom for 1705 atoms.

The base-pair breathing was followed by measuring the distances ($R$) between the H-bond forming atoms shown in Fig. 1 at every time step. The stopwatch was started when a given distance exceeded a certain threshold ($R_{th}$) and stopped once the boundary was crossed in the opposite direction. These events are called Poincaré recurrences. The integrated probability distribution $P(\tau)$ is obtained by counting the number of recurrences with duration larger than $\tau$ and normalizing it by the total number of events, that is, $P(0) = 1$ by construction. Function $P(\tau)$ is a very powerful instrument of analysis because it is positive definite and, due to statistical averaging over a large number of crossings, stable with respect to fluctuations (see, e.g., discussion in [10]). Importantly, these computations are trivially parallelizable, that is, the $P(\tau)$ statistics can be accumulated in a large number of independent MD trajectories. Most of the results discussed below were obtained by using parallel computations on 65 cores.

Representative results obtained for the three H bonds in terminal base pairs are shown in Fig. 2. The left panel displays double-logarithmic plots of $P(\tau)$ distributions obtained with $R_{th} = 3.15$ Å. This threshold is close to the equilibrium H-bond lengths; therefore, a large fraction of recurrences result from oscillations within the bonded ground state. These motions give for $\tau < 0.3$ ps a characteristic fall of $P(\tau)$ typical for exponential decays. The right panel of Fig. 2 shows the plots of average distances for recurrences of different duration. For returns shorter than 0.3 ps these values remain around 3.5 Å, that is, the H bonds are not broken. With $\tau > 0.3$ ps the average distances grow with $\tau$ and the $P(\tau)$ decay becomes algebraic. The short time boundary of the power-law relaxation is close to that in TRSS experiments [21]. Moreover, the decay exponent $\beta \approx 1$ for all three H bonds is not far from experimental $\beta = 1 = \nu \approx 0.15$. To refine these results we decided to concentrate upon the O6N4H bond in terminal base pairs. This H bond is broken easier than others; therefore, it is an adequate indicator of partial base-pair openings. The internal base pairs were also checked, but they are opened rarely and only traces of algebraic decay were detectable in the tails of $P(\tau)$ distributions.

![FIG. 1](color online). The left panel shows the Watson-Crick base pair formed by guanine and cytosine. The three H bonds are shown by thick dashed lines. Their lengths are about 3 Å. The right panel shows a snapshot of the model system taken at the end of one of the 65 MD trajectories involved in statistical analysis. A tetramer fragment of a B-DNA double helix was built from two identical strands with the self-complementary base pair sequence GCCG. The six sodium ions necessary for charge neutralization are shown as spheres.

![FIG. 2](color online). The left panel shows statistics of Poincaré recurrences $P(\tau)$ for the three Watson-Crick H bonds shown in Fig. 1. The results for distances O6N4, N1N3, and N2O2 are displayed by solid black, dashed red, and dotted blue curves, respectively. The threshold distance is $R_{th} = 3.15$ Å in all three cases. The right panel shows the corresponding dependencies of average bond distances ($R$) obtained for recurrences of different duration. Here and in other figures the logarithms are decimal.
To this end, the threshold of quick returns that do not result in the H-bond opening. As shown in Fig. 3, we find that the short time hump is in excellent agreement with the experimental data [21]. We must note that the short time hump for thresholds below 3.15 Å, thus pushing the system to better sampling of long-time recurrences. In both panels the dash-dotted straight line shows the power law decay with the Poincaré exponent $\beta = 1.15$.

To get a better estimate of the short-time boundary of the algebraic decay in $P(\tau)$ we need to remove the contribution of quick returns that do not result in the H-bond opening. To this end, the threshold $R_{th}$ was gradually increased up to 4.55 Å. The results of these computations are displayed in the left panel of Fig. 3. It is seen that the short time hump is essentially removed already with $R_{th} = 3.55$ and that for the true base-pair breathing the algebraic decay starts from very short recurrences of only 0.1 ps. This refined boundary is in excellent agreement with the experimental data [21].

These computations were also used to refine the estimate of the exponent $\beta$. Linear fits of the plots in the left panel of Fig. 3 were carried out in the range $-0.5 \leq \log(\tau/\mu s) \leq 2.5$, which gives exponent values $\beta = 1.27$, 1.20, 1.13, and 1.05 for thresholds $R_{th} = 3.15, 3.55, 4.15, \text{and} 4.55$ Å, respectively. It is seen that there is a moderate influence of $R_{th}$ upon the exponent, which can be related to the fact that longer recurrences correspond to larger atom-atom separations (Fig. 2). Note that the $\beta$ values obtained give the exponent of correlation decay $\nu = \beta - 1$ very close to the experimental TRSS value ($\nu \approx 0.15$) [21]. For visual comparison, the $P(\tau)$ decay predicted from experiment is shown in Fig. 3 by the dash-dotted lines.

The experimental long-time boundary occurs at $\approx 100$ ns and it corresponds to the maximal resolution of the TRSS method [20,21]. The long-time boundary in our computations is limited by sampling. We tried to push it somewhat further by taking into account that H bonds in our simulations spend almost all the time oscillating around the ground energy minimum. To reduce this nonproductive time a flat-bottom restraint was added that prevented the O6N4 distances to go below 3.15 Å. It is understood that this simple ad hoc trick perturbs realistic dynamics, but the information it provides may be useful. The results of these computations shown in the right panel of Fig. 3 confirm that the sampling is indeed improved, with the time range of the approximate power law extended by about an order of magnitude. For shorter duration the $P(\tau)$ distribution reproduces the curve obtained without the restraint. Therefore, one can reasonably expect that the long-time boundary of the power-law decay of $P(\tau)$ occurs at least at $\approx 10$ ns.

Encouraged by the foregoing findings, notably, the surprisingly good agreement with experiment obtained with no adjustable parameters, we decided to study the power spectrum of the fluctuations of the O6N4 distance. The results shown in Fig. 4 reveal a strong power-law growth of density $S$ for low frequencies. However, the exponent $\eta = 1.63$ estimated by linear fitting does not seem to be related to either the experimental data or the Poincaré recurrences studied above. This paradoxical observation is discussed further below.

To shed light on the possible origin of the foregoing results we use the virial theorem [54] for analysis of chaotic dynamics in systems with representative polynomial potentials. The corresponding Hamiltonian reads

$$H(p,r) = p^2/2 - a/r^{m} = E,$$

where $(p,r)$ is the pair of conjugated variables, with momentum $p$ and radial coordinate $r$. The energy $E < 0$ corresponds to a bounded motion; $a \sim 1$ and $m > 0$ are numerical coefficients (the mass is taken as unity). We assume that in addition to radial dynamics there is a chaotic motion in angle degrees of freedom. For this system the action can be estimated as $J \sim pr \sim (bE)^{(m-2)/(2m)}$ with a certain numerical constant $b < 0$, which follows from relationships $p^2 \sim 1/r^{m}$ and $E \sim p^2$. Thus we have $bE|E| \sim J^{(m+2)/(m-2)}$, and $\omega = dH/dJ \sim J^{(m+2)/(m-2)} \sim 1/\tau$, where $\tau$ is the correlation time.
where $\omega$ is a frequency of motion and $\tau$ is the related characteristic time scale.

As a result, the measure $\mu$ related to the sticking time scale $\tau$ is obtained as $\mu \sim J \sim 1/|E|^{m-2}/(m+2)$, which follows from the fact that in Hamiltonian systems the measure is proportional to the phase volume, that is, $\mu \sim \int J d\theta \sim J \times 2\pi$, where $\theta$ is the angle variable conjugated to action $J$. According to [10] we have $P(\tau) \sim d\mu/d\tau \sim \mu/\tau$, where $\mu(\tau)$ is the measure of a region where a trajectory is stuck for the time $\tau$. This follows from the ergodicity relation according to which the measure of a region is proportional to the time spent by the trajectory in this region $\mu(\tau) \sim \tau P(\tau)/(\tau)$, where $\langle \tau \rangle = \int_0^\infty P(\tau) d\tau$ is the average time of recurrences [2,3]. From these relations one obtains the following expression for the Poincaré exponent:

$$\beta = 2m/(m + 2).$$

Here, $P(\tau)$ can be considered as an integrated probability of Poincaré recurrences or as a survival probability in a given region for time periods longer than $\tau$ since both are proportional to each other [10,12].

Consider some earlier studied potentials with different $m$. With $m = 1$ we get the Kepler problem. This case appears in the microwave ionization of Rydberg atoms [17,18], and also in the comet [19] or dark matter [55] dynamics in the Solar System affected by Jupiter. In both cases the energy change occurs when the particle passes near the perihelion (near the nuclei or near the Sun). This energy change produces chaotic dynamics in the system. In this case the measure $\mu \sim J \sim 1/|E|^{1/2}$ is diverging at $|E| \to 0$ and from (3) we have $\beta = 2/3 < 1$, which agrees with the analytical and numerical results reported in [56]. For $m = 2$ we have a period independent of action, which gives $\beta = 1$ without decay of correlations, that is, $C \sim \tau P(\tau) \propto \text{const}$ with $\nu = \beta - 1 = 0$.

For the most relevant case of dipole-dipole interactions that are dominant in H bonds and, more generally, in neutral polar systems like B-DNA with ions in water, we have $m = 3$ and $\beta = 1.2$. The last value is close to that obtained in our numerical simulations as well as that corresponding to exponent $\nu = \beta - 1$ found in the TRSS experiments. We believe that the above estimates correctly capture the main physical effects in the dynamics of this complex system and are at the origin of the observed slow algebraic decay of Poincaré recurrences. It is understood that in systems with thousands of atoms like that studied here other factors can interfere. It is possible that Coulomb forces ($m = 1$) of locally uncompensated charges are responsible for a certain reduction of $\beta$ to a slightly smaller value compared to the above theoretical estimate. Also, for the van der Waals potential we have $m = 6$ with the corresponding $\beta = 3/2$. We make a conjecture that on small scales in the vicinity of stability regions of dynamical maps the sticking trajectories that generate ASPR [11–16], experience forces similar to those of polynomial potentials (2) thus leading to similar $\beta$ values.

Finally, consider the result shown in Fig. 4. The measure of sticking regions is $\mu \sim J \sim 1/|E|^{1/2} \sim 1/\sqrt{\tau} \sim 1/\tau^{1/5}$. It decreases with large $\tau$, but the typical atom-atom separations are growing as $R_b \sim \tau^{2/5}$. The correlation function $C(\tau) = \langle r(t + \tau) r(t) \rangle$, is commonly defined for a bounded variable $r(t)$, so that the average square variation is $\langle (r(t)^2) \rangle \sim \tau \int C(\tau) d\tau \sim \tau^\nu$, where $\nu = \beta - 1 < 1$ is the correlation decay exponent related to the superdiffusive growth. In the present case, we have a variable that grows with $\tau$, which gives an additional contribution to the average square variation $\langle r(t)^2 \rangle \sim R_b^2 \tau C(\tau) \sim R_b^2 \tau^{3-\beta} \sim \tau^\kappa$ with $\kappa = 3\beta - 1 = 2.6$. Using the Wiener-Khintchine relation (see, e.g., [57]) between the square variation of a time dependent variable $r(t)$ and its spectral density $S(f)$ we obtain

$$S(f) \propto 1/f^n, \quad \eta = \kappa - 1 = 3\beta - 2 = 1.6$$

in good agreement with Fig. 4. Thus, we have a very strong low frequency noise, with the $\eta$ exponent larger than usual [58], that results from only the inherent chaotic dynamics of the system, with no external noise involved.

In summary, using all-atom MD simulations, we uncovered the existence of algebraic decay of Poincaré recurrences in the base-pair breathing motion of the B-DNA molecule, with the decay exponent $\beta \approx 1.2$ and a strong divergence of spectral density of vibration motion with the exponent $\eta \approx 1.6$ over at least five decades in time from $10^{-13}$ to $10^{-8}$ sec. These results are well described by the proposed theory of Poincaré recurrences for chaotic dynamics in polynomial potentials, assuming a dominant contribution of dipole-dipole interactions. The theory correctly captures the qualitative origin of this effect in MD in spite of the difference in the number of degrees of freedom. The exponent $\beta \approx 1.2$ predicts a power-law relaxation of the corresponding correlations with the exponent $\nu = \beta - 1 \approx 0.2$. This prediction as well as the time scale of the algebraic decay in MD are in striking similarity with TRSS experiments, suggesting that this effect is due to the base-breathing motion. Further studies should help to clarify the relationship between the inherent dynamical chaos in DNA and these experimental data.

The computational resources used in this study were supported by the Initiative d’Excellence program from the French State Grant No. ANR-11-LABX-0011-01 (DYNAMO).