Viewpoint

Statistics and the single molecule

By Igor M. Sokolov
Institut für Physik, Humboldt-Universität zu Berlin, Newtonstrasse 15, D-12489 Berlin, Germany

Current technology permits tracking single molecules with exquisite precision, but the results need to be interpreted with care. Long-duration measurement of the motion of a single particle yields information that is different and complementary to that obtained from an ensemble average of many particles.

Subject Areas: **Statistical Mechanics, Biological Physics**

One hundred years ago, the atomic-molecular theory of matter was having a hard time, and many physicists considered it merely a kind of convenient shorthand rather than a real description of nature; after all, nobody had really seen a molecule, let alone an atom. Today, developments in micromanipulation and in single-molecule tracking have not only made individual molecules visible, but have led to real breakthroughs in understanding of the molecular basis of life. This ability to follow and to manipulate single molecules has opened new perspectives in nanoscience and nanotechnology. Experts in single-molecule tracking often say that observation of individual trajectories gives more information about the system than only looking at ensemble averages, which is the approach taken in statistical thermodynamics. The idea is that the closer one looks, the more information one can get.

In a paper published in *Physical Review Letters* however, Yong He, Stanislav Burov, Ralf Metzler, and Eli Barkai (at Bar Ilan University in Israel and the Technical University of Munich) show that the information obtained in such single-particle experiments is *different* from that given by the ensemble-averaged cases, so one has to be careful about interpreting the results [1]. This is especially the case when the measured motion exhibits subdiffusion (a process that is slower than normal Fick's law diffusion) that might be nonergodic (the time and ensemble averages give different answers). This situation is often encountered in both nonliving physical systems such as disordered semiconductors and groundwater motion in geophysical formations, and in the crowded interiors of living cells.

He *et al.* base their theoretical analysis and numerical simulations on the so-called continuous-time random walk (CTRW) model, first introduced by Montroll and Weiss in 1965 [2]. CTRW was developed to handle a variety of complex diffusion processes by considering the motion of particles on lattices (Fig. 1). The importance of the model became clear after Scher and Montroll [3] successfully used it in 1975 to explain dispersive charge carrier transport in strongly disordered semiconductors (the ubiquitous working media of copy machines and laser printers). In the CTRW model, a particle hardly moves most of the time, and only occasionally gets an opportunity to jump to a new location. The motion is therefore described as a sequence of jumps into different directions interrupted by periods during which the particle is just waiting for the next jump.

Simple random walks were first discussed by Rayleigh [4] who concentrated on the dependence of the quantities of interest on the number of jumps. The theory of continuous-time random walks instead concentrates on the temporal aspect of the problem. If there exists a well-defined mean waiting time, the overall displacement follows the normal diffusion, in which both the mean squared displacement in the absence of the external force $\langle x^2(t) \rangle \cong Dt$ (where D is the diffusion coefficient), and the mean displacement $\langle x_F(t) \rangle \cong \mu F t$ (where μ is the mobility) under the action of the constant external force F are proportional to each other and both grow as the first power of the time t. A venerable example is the one that captured Einstein's attention: colloidal particles undergoing diffusive Brownian motion, while at the same time falling downward due to gravity.

This proportionality has deep roots in the behavior of physical systems close to thermodynamic equilibrium; the mobility μ and the diffusion coefficient D are not independent, but are connected to each other by Einstein's relation $D = k_B T \mu$. In normal diffusion the "average" can be understood either as an ensemble average $\langle x^2(t) \rangle_{\text{ens}}$ over a large ensemble of moving particles, or as a temporal moving average $\langle x^2(t) \rangle_{t_{\text{avg}}}$ over a very
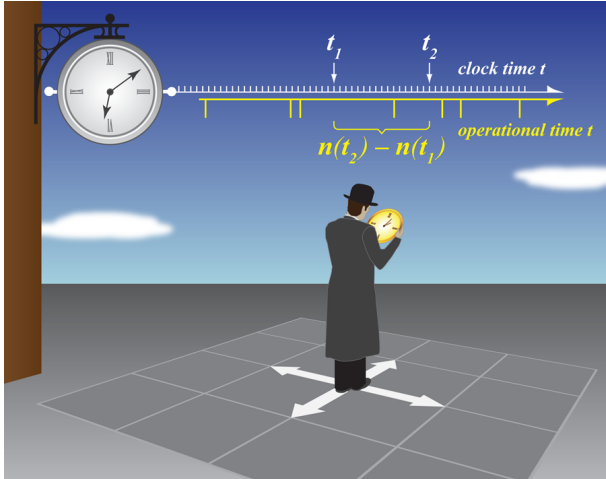
FIG. 1: In continuous-time random walks, the walker's position is governed only by the number of preceding steps. This number of steps $n(t)$ constitutes the operational time of the problem as recorded by the walker's own clock, which ticks once each time n is incremented. Since $n(t)$ grows slower than linearly with the physical (clock) time, this watch is always behind, leading to the overall subdiffusive behavior, as compared with the otherwise normal random walk. He *et al.* have used this model to study time-averaged single-molecule behavior in comparison with ensemble averages of many molecules. (Illustration: Alan Stonebraker/`stonebrakerdesignworks.com`)

long trajectory of motion of duration t_{avg} for a single particle. Normal Fick's law diffusion is an ergodic process (that is, both averages give the same result).

Strange things happen when the calculated mean waiting time diverges, as was the case with the carrier transport investigated by Scher and Montroll where the probability density followed a power law proportional to $t^{-1-\alpha}$. When $0 < \alpha < 1$, the system is said to exhibit subdiffusion, characterized by $\langle x^2(t) \rangle_{ens} \propto \langle x_F(t) \rangle_{ens} \propto t^\alpha$ in the ensemble average. Apart from disordered semiconductors, the CTRW model with power-law waiting-time distribution adequately describes such different phenomena as the spread of pollutants in underground water (where the particles can be caught in stagnant regions of the flow), and many biological situations in the interior of living cells, where the motion is strongly hindered by a bulky cytoskeleton and by the existence of other huge molecules around the molecule we are interested in.

Because nothing happens between the jumps in the CTRW model, it is the number of jumps that is the appropriate internal time variable describing the process, its so-called *operational time*. If a well-defined mean waiting time τ exists, the diffusion is normal, since both $\langle x^2(t) \rangle$ and $\langle x_F(t) \rangle$ are proportional to the mean number of steps n, which in turn goes as t/τ. In the case of anomalous diffusion, the mean squared displacement and the mean displacement under a constant force are

still proportional to each other, but the number of steps shows a different time dependence going as $n \propto t^\alpha$.

In the case of disordered semiconductors, the ensemble average makes sense owing to the multiparticle nature of the physical quantity of interest, namely, the electric current in the form of simultaneous motion of many charge carriers. On the other hand, in single molecule experiments the time average is often used. In their paper, He *et al.* show that in some cases the results of experiments on mRNA molecules and lipid granules are well described by the CTRW model and thus an ensemble average will differ from the single-particle time average. Contrary to what might be expected, one observes in the time-averaged picture not anomalous diffusion, but normal diffusion, albeit with strongly fluctuating diffusion coefficient. The result is easy to grasp. The mean squared displacement during the time interval t between the two instants t_1 and $t_2 = t_1 + t$ is governed by the number of steps that occur in between. This grows on the average as $n(t) = n(t_2) - n(t_1) \propto (t_1 + t)^\alpha - t_1^\alpha$, i.e., approximately as $t_1^{\alpha - 1}$ for $t \ll t_1$. This proportionality to t also survives after temporal integration assumed by the moving time average, giving rise to the overall seemingly normal diffusion behavior $\delta^2(t) = \langle x^2(t) \rangle_{t_{avg}} \propto t$ (where $\delta^2(t)$ is the measured mean squared displacement) as opposed to the ensemble-averaged $\langle x^2(t) \rangle_{ens} \propto t^\alpha$. So, we can be fooled by a single-molecule measurement into thinking that the entire ensemble is undergoing normal diffusion.

My colleagues and I found this basic result recently [5], [6] and used it to show that the standard models of potential landscapes are unable to describe equilibrium fluctuations in peptides. However, He *et al.* have gone much further in their discussion. In particular they generalized Einstein's relation for a given specific situation, and moreover discuss in detail the *distribution* of measured mean squared displacement $\delta^2(t)$, which can be considered as a proxy for the distribution of the diffusion coefficients measured in experiment.

I would like to stress the aspect of *universal fluctuations* connected with this distribution. Typically, in ergodic systems, the longer the averaging time, the narrower is the distribution of the result. For example, the mean squared displacement measured as the moving time average for given time-lag t in the normal diffusion approaches a deterministic value $\bar{\delta}^2 = \langle x^2(t) \rangle_{ens}$ when the averaging time grows, $t_{avg} \to \infty$. The width of the distribution of the relative result $\delta^2/\bar{\delta}^2$ tends to zero. In the case of subdiffusion considered above, the width of the distribution of $\delta^2/\bar{\delta}^2$ stagnates, and increasing measurement time does not improve the result. The overall distribution of $\delta^2/\bar{\delta}^2$ tends to a universal form depending only on the exponent α (which contains the specific details of the system).

Being unaware of this nonergodicity, one could come to the wrong conclusion that the system under investigation is inhomogeneous, i.e., that each of different ran-

dom walkers tracked is physically different, or, mathematically speaking, their motions correspond to realizations of different random processes (normal diffusion with different diffusion coefficients). However, in reality what we see is different realizations of the same random process corresponding to subdiffusion. The overall behavior strongly resembles what has been found in biological experiments (e.g., when following the motion of single viruses in the cell [7]), although one has always to be extremely cautious when comparing the results of theories based on one mechanism or cause, when in fact experiments are influenced by many different factors.

There is another important and interesting result reported by He *et al.*[1]. Up to now, we have discussed the situation in an infinite system, when the walkers' motion is not restricted by any boundaries. The cells, on the contrary, are not only finite but relatively small. As we have seen, the time moving average in the infinite system exhibits normal diffusion (although the underlying process is anomalous). In a finite system, we still see hints of the underlying anomalies, as the authors show by direct numerical simulation of the time-averaged CTRW on a relatively small one-dimensional lattice. The results of these simulations resemble strongly the observations of Golding and Cox [8] on the motion of mRNA molecules inside bacterial cells and can probably explain these findings (although here one has to be cautious, too).

Statistical thermodynamics typically deals with systems that rapidly relax to equilibrium or to a stationary state, implying the system is ergodic. Systems far from equilibrium or showing very slow relaxation may be nonergodic, and subdiffusion as modeled by CTRW may be one of the simplest theoretical examples. One has to be cautious when applying our intuition gained for the close-to-equilibrium cases to such processes: the information contained in the time-averaged and ensemble-averaged results is different and is pertinent to different aspects of the system's behavior. Understanding this fact is necessary when interpreting the results of existing experiments and when planning future studies.
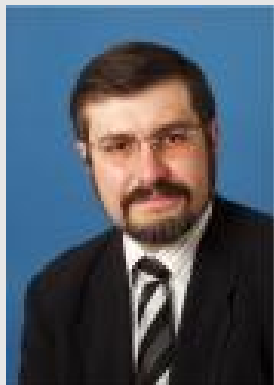
References

[1] Y. He, S. Burov, R. Metzler and E. Barkai, *Phys. Rev. Lett.* **101**, 058101 (2008).
[2] E. W. Montroll and G. H. Weiss, *J. Math. Phys.* **6**, 167 (1965).
[3] H. Scher and E. W. Montroll, *Phys. Rev. B* **12**, 2455 (1975).
[4] Lord Rayleigh, *Nature* **72**, 318 (1905).
[5] A. Lubelski, I. M. Sokolov, and J. Klafter, *Phys. Rev. Lett.* **100**, 250602 (2008).
[6] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, *Phys. Rev. Lett.* **100**, 188103 (2008).
[7] G. Seisenberger, M. U. Reid, T. Endreß, H. Büning, M. Hallek, and C. Bräuchle, *Science* **294**, 1929 (2001).
[8] I. Golding and E. C. Cox, *Phys. Rev. Lett.* **96**, 098102 (2006).

About the Author

Igor M. Sokolov

Igor M. Sokolov received his Ph.D. from the Moscow State University in 1984 and worked until 1990 at the department of theoretical physics at Lebedev Physical Institute in Moscow. In 1990 he moved to the University of Bayreuth, Germany, on a fellowship of the Alexander von Humboldt Foundation. From 1991 to 2001 he held a position at the University of Freiburg, Germany, and left for Berlin in 2001. Since 2005 he has been a full professor of physics at Humboldt University in Berlin. His main scientific interests include nonequilibrium thermodynamics, transport phenomena in flows and in disordered systems, chemical kinetics, networks, and polymer dynamics. In 2008 Igor M. Sokolov was recognized as one of the Outstanding Referees of the American Physical Society.