Numerical simulation of the Perrin-like experiments

Zygmunt Mazur\textsuperscript{1} and Dariusz Grech\textsuperscript{2}

\textsuperscript{1} Institute of Experimental Physics, Wrocław University, pl. M. Borna 9, 50-405 Wrocław, Poland
\textsuperscript{2} Institute of Theoretical Physics, Wrocław University, pl. M. Borna 9, 50-405 Wrocław, Poland

Received 18 February 2007, in final form 21 October 2007
Published 21 November 2007
Online at stacks.iop.org/EJP/29/91

Abstract
A simple model of the random Brownian walk of a spherical mesoscopic particle in viscous liquids is proposed. The model can be solved analytically and simulated numerically. The analytic solution gives the known Einstein–Smoluchowski diffusion law $\langle r^2 \rangle = 2Dt$, where the diffusion constant $D$ is expressed by the mass and geometry of a particle, the viscosity of a liquid and the average effective time between consecutive collisions of the tracked particle with liquid molecules. The latter allows us to make a simulation of the Perrin experiment and to verify in the detailed study the influence of the statistics on the expected theoretical results. To avoid the problem of small statistics causing departures from the diffusion law we introduce in the second part of the paper the idea of the so-called artificially increased statistics (AIS), and we prove that, within this method of experimental data analysis, one can confirm the diffusion law and get a good prediction for the diffusion constant even if trajectories of just a few particles immersed in a liquid are considered.

1. Introduction

Recently, much progress has been made in application of digital technique in experimental physics what allows us to perform milestone physics experiments even in student laboratories. A good example is the Perrin experiment [1] considered as the first one directly proving the atomic structure of matter. However, its verification at university laboratories [2–5] due to small statistics one takes, may meet some difficulties (see, e.g. [2, 4, 5]). The linear dependence between the average square displacement $\langle r^2 \rangle$ of the particle in media due to its Brownian motion and the observation time $t$ as required by the Einstein–Smoluchowski diffusion law often becomes very problematic.

It is essential therefore to examine the minimal statistics (number of tracked particles) one should take into account in the limited observation time to reveal the major feature of the
diffusion law. We propose the analytical model which can also be easily simulated numerically. The aim of this model is to investigate how the results of \( \langle r^2 \rangle \) versus \( t \) depend on the statistics and what the scaling range of the expected linear relationship is. This study should be helpful to set up the experiment properly as well as to analyse the obtained results more correctly.

In the following section, we present the model directly reflecting the physics standing behind the Perrin experiment. In section 3, the results of the numerical simulation of this model are described and the main features of the diffusion relation with its scaling range are revealed for various numbers of particles to be observed. To avoid the problem of small statistics causing departures from the strict power-law behaviour, we introduce and discuss the idea of artificially increased statistics (AIS) in section 4. This method is then applied both to the results of the numerical simulation and to some experimental data. We argue the method may significantly decrease the level of statistical noise in data, leading to much better agreement with the linear dependence in diffusion theory. In the last section, a summary of obtained results is given.

2. Description of the model

The first derivation of the diffusion law based exclusively on the molecular–kinetic theory in media with viscosity comes from Einstein [6]. Soon after his publication, Langevin [7] and Smoluchowski [8] presented their own approaches to this issue.

Einstein considered diffusion as the random walk of diffused particles. He was able to get the diffusion equation

\[
\frac{\partial f(x,t)}{\partial t} = D \frac{\partial^2 f(x,t)}{\partial x^2} \tag{1}
\]

just from the random motion assumption, where \( f(x,t) \) is the local density of diffused particles in the space \((x)\) and time \((t)\) coordinates, and \( D \) is the diffusion constant. The only assumption Einstein made was that particles make the random walk described by some probability density \( \phi(\Delta) \) for any step of length \( \Delta \) (although he did not name \( \phi(\Delta) \) this way in his original paper [6]).

The positive solution of equation (1) satisfying the boundary condition

\[
\int_{-\infty}^{+\infty} f(x,t) \, dx = n, \tag{2}
\]

where \( n \) is the total number of diffused particles, reads

\[
f(x,t) = \frac{n}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{3}
\]

It tells us that one will find a diffused particle at the distance \( x \) away from the source \((x = 0)\) after some time \( t \) with probability density described by the normal distribution.

Therefore, the mean-squared displacement of diffused particles will be given as

\[
\langle x^2 \rangle_n = \int_{-\infty}^{+\infty} x^2 f(x,t) \, dx = 2Dt. \tag{4}
\]

If \( d \)-dimensional motion is considered, the rhs of equation (4) is obviously replaced by the \( 2Dt \) term with \( D = d\tilde{D} \).

Formula (4) was a milestone step in atomic physics. Einstein was the first person who suggested that from the linear dependence between \( \langle x^2 \rangle_n \) (if measured) and the time \( t \) one can find out the value of diffusion constant \( D \). His derivation also assumed that a particle in suspension behaved exactly as do particles dissolved in a liquid. This assumption does not
Numerical simulation of the Perrin-like experiments

follow from the classical thermodynamics point of view but it can be clarified in the framework of molecular–kinetic theory. As a result, Einstein showed that the diffusion constant \( \tilde{D} \) from equation (1) can be independently expressed for particles in suspension in one-dimensional motion by statistical physics constants (\( R \) is the gas constant, \( N_A \) is the Avogadro number and \( T \) is the absolute temperature) and the viscosity of the liquid \( \eta \) as

\[
\tilde{D} = \frac{RT}{N_\sigma} \frac{1}{6\pi \varrho \eta},
\]

where \( \varrho \) is the radius of diffused spherical particles.

This way, the self-consistency of the molecular–kinetic theory with the statistical physics and the random walk assumption can be proved. The latter was the aim of the Perrin experiment [1] performed a few years after the Einstein paper [6] had been published.

Here we propose another approach to the diffusion problem based entirely on the time-series analysis. The analysis is founded on the average time \( \tau \) between consecutive collisions of the tracked mesoscopic particles with other particles in a liquid (i.e., \( \tau \) has the meaning of the average time between collisions changing significantly the motion of the tracked object). Such an approach seems to be closer to the spirit of the original Perrin experiment [1].

Let the trajectory of the observed particle of mass \( m \) moving in \( d \)-dimensional space is \( x^\alpha(t) \), where \( \alpha = 1, 2, \ldots, d \). We assume \( x^\alpha(t) \) to be discrete \( d \)-dimensional time series with constant spacing \( \tau \) in time \( (t = 0, \tau, 2\tau, \ldots, N\tau) \). The obvious notation

\[
x^\alpha(k\tau) = x^\alpha_k, \quad k = 1, 2, \ldots, N
\]

and

\[
\Delta x^\alpha_k = x^\alpha_{k+1} - x^\alpha_k
\]

will be applied, where \( \Delta x^\alpha_k \) is the instantaneous displacement of the particle at \( t = k\tau \).

For the stationary, integer Brownian motion (no autocorrelations in displacements) with no drift one has for large \( n \)

\[
\langle \Delta x^\alpha \rangle_n = 0
\]

and

\[
\langle \Delta x^\alpha_i \Delta x^\alpha_j \rangle_n = \delta_{ij} \sigma^2
\]

where \( \langle \cdot \rangle_n \) is the average taken over the ensemble of \( n \) tracked particles and \( \sigma^2 \) is the variance of the displacements, i.e.

\[
\langle (\Delta x^\alpha)^2 \rangle_n = \sigma^2.
\]

The total mean-squared displacement \( \langle r^2 \rangle_n \) of the particles from their initial positions after \( N = t/\tau \) collisions can be easily calculated with the help of equation (10):

\[
\langle (\Delta r)^2 \rangle_n = \left( \sum_{\alpha=1}^d \left( \sum_{i}^N \Delta x_i^\alpha \right)^2 \right) = \frac{d\sigma^2}{\tau - 1}.
\]

In order to calculate \( \sigma^2 \) let us note that

\[
\Delta x_i^\alpha = \sqrt{\langle v_i^\alpha \rangle}\tau,
\]

where \( \sqrt{\langle v_i^\alpha \rangle} \) is the standard deviation of the velocity fluctuations.
with $\langle v_{\alpha i}^2 \rangle_\tau$ being the average velocity of the $i$th particle between collisions. Hence, from equations (10) and (12)

$$\sigma^2 = \tau^2 \langle (v_{\alpha i}^2) \rangle_\tau. \tag{13}$$

The equipartition theorem establishes the connection of microscopic quantities with the absolute temperature $T$ and the Boltzmann constant $k$:

$$\frac{1}{2} m \langle (v_{\alpha i}^2) \rangle_\tau = \frac{1}{2} kT. \tag{14}$$

Therefore, equation (11) reads

$$\langle \Delta r^2 \rangle_n = \left( \frac{d k T}{m \tau} \right) t. \tag{15}$$

The above formula is the standard diffusion law with the diffusion constant

$$D = \frac{d k T}{2 m \tau} \tag{16}$$

expressed in terms of $\tau$.

Usually, one writes $D$ in terms of liquid viscosity $\eta$ (see equation (5)) as

$$D = \frac{d k T}{\alpha}, \tag{17}$$

where $\alpha = 6\pi \varrho \eta$ (the Stokes law) and $\varrho$ is the radius of the considered mesoscopic particles.

Hence one gets the simple relation between the parameter $\tau$ in the model and the macroscopic quantities $m$, $\alpha$:

$$\tau = \frac{2 m}{\alpha}. \tag{18}$$

Thus, the model reproduces the known diffusion law and also estimates the average time $\tau$ lapsing between consecutive collisions in the system as the simple function of macroscopically measured quantities. The time $\tau$ can be taken as the input parameter in the numerical study of the Perrin experiment. It will be done in the following section.

### 3. Numerical simulation of the Perrin experiment

The solution in equation (15) can be checked by the numerical simulation of the Brownian motion in viscous media. In fact such simulation is the only way one can find the sufficient statistics, i.e. the number of tracked particles in the ensemble one should observe in real experiment to obtain results confirming the linear relation. If sufficient statistics requirement is not satisfied, one observes significant departures from the linear behaviour $\langle r^2 \rangle_n \sim t$ (see, e.g. [2, 4]).

We simulated all time series $\{x_{\alpha i}^a\}$ in $d = 2$ dimensions usually discussed by experimentalists. The time series were built in the well-known iterative way:

$$x_{\alpha i+1}^a = x_{\alpha i}^a + \Delta x_{\alpha i}^a, \tag{19}$$

$$r_{\alpha i}^2 = (x_{\alpha i}^1)^2 + (x_{\alpha i}^2)^2, \tag{20}$$

where displacements have been generated as the random Gaussian numbers $N(0, \sigma)$ with the standard deviation $\sigma = \tau (kT/m)^{1/2}$ obtained from equations (13) and (14). All simulations were performed for the case of diffusion in pure water ($\eta = 1.00 \times 10^{-3}$ Pa s), the room temperature $T = 293$ K, $m = 4.28 \times 10^{-16}$ kg and $\varrho = 425$ nm which roughly corresponds to the real Perrin experiment parameters.
The essential task to be done just in the beginning was to determine the maximum time up to which \( \langle r^2 \rangle \sim t \) is well observed. This time, called the scaling range \( \lambda \) of the discussed linear dependence, is a function of the number of tracked particles \( n \). We estimated \( \lambda \) for the bunch of simulated trajectories for the number of observed particles changing in the range \( n = 10–500 \). The bunch of 20 trajectories was investigated for any \( n \) in the above range. The examples of just five runs in each bunch (for the clarity of the figure we do not show all the runs) are shown in figures 1(a)–(d). Hence we have found the scaling range relation revealed in figure 2. The best fit gives

\[
\lambda \sim n^\beta,
\]

where \( \beta = 0.51 \pm 0.04 \) and the uncertainty comes from the statistics.

Let us note that if the number of observed particles does not exceed 10 the linear dependence \( \langle r^2 \rangle_n \sim t \) can be confirmed only for the observation time \( t < 3 \) s! Thus, the analysis taking into account longer observation times (as authors of [4] did) is simply incorrect.

Having the scaling range determined we may proceed to calculate the diffusion constant value and its expected standard deviation from the mean. Such analysis was done by us for the simulated trajectories mentioned above. Some chosen cases (again for the clarity of the graph we do not show all of them) with maximal and minimal values of slopes \( (2D) \) for every \( n \) are shown in figures 3(a)–(d). All results of the mean \( D \) values and their standard deviation as the function of \( n \) are presented in figure 4.

Hence we see that the final result within 10% of the expected theoretical value can be found only if one considers the ensemble of \( n \geq 50 \) particles.
4. Analysis of results with artificially increased statistics

The results of the previous section seem to suggest that to get a reasonable agreement with the diffusion law predictions one should take into account in the real experiment data from at least \( n \sim 50 \) particles. In many less professional labs (e.g. student labs), such a requirement...
Numerical simulation of the Perrin-like experiments

Figure 4. The mean $D$ values as the function of $n$ obtained via numerical simulation. The vertical lines represent the standard deviation. The dotted horizontal line points the expected theoretical value $D_{th} = 1.01 \mu m^2 s^{-1}$.

Figure 5. The bunch of $\langle r^2 \rangle \sim t$ for $n = 10$ (a) and $n = 50$ (b) particles worked out with AIS. The enlarged scaling range regime is shown separately.

is virtually impossible to be satisfied—mainly because of the limited time duration of the data collection—particularly if no sophisticated computerized apparatus is used. Below we give the idea that helps us to overcome such a difficulty. We call it artificially increased statistics (AIS).

The main idea of AIS is to build the statistics of consecutive displacements from the very small number of available trajectories, counting all the displacements not from the initial starting point $(x_{10}^1, x_{20}^1) = (0, 0)$ but varying it along the whole one-particle trajectory. Thus, any momentary position of the particle, say $(x_{1k}^1, x_{2k}^1), k = 1, 2, \ldots, N$, is the starting point to collect statistics of all displacements afterwards, i.e. $(\Delta x_{1l-k}^1, \Delta x_{2l-k}^1), l > k$, where $\Delta x_{\alpha l-k}^\alpha = x_{\alpha l}^\alpha - x_{\alpha k}^\alpha$ is the $\alpha$th part of the $(l-k)$ step displacement. This way, for the time series of length $N$, one has $N - m$ data for $m$-steps displacements instead of just one displacement usually taken into account. Then the statistics are averaged in the usual way over the all considered (observed) particles. Therefore, even if $n$ is small, the overall number of data entering the statistics is large enough to fulfill the linear law expectation.
Figure 6. The result of the AIS procedure applied to the real experimental data of \( n = 5 \) latex spherical particles diffusing in pure water (data taken from [4]): (a) represents the ‘naked’ results, while (b) shows the results lifted by AIS.

Let us now look at the results of the application of AIS to the simulated Brownian motion and to the pure experimental data from the real experiments.

In figures 5(a) and (b), we present the bunch of squared displacements in time taken for the statistics of \( n = 10 \) (a) and \( n = 50 \) (b) particles worked out with the AIS procedure. The comparison with the ‘naked’ data from figures 3(a) and (b) shows the tremendous difference. Although the scaling range after AIS does not seem to change a lot, the linear dependence \( \langle r^2 \rangle \sim t \) is now much more convincing. In fact the comparison of the diffusion constants \( D \) obtained from the ‘naked’ analysis and from the data improved by AIS shows about seven times smaller uncertainty in \( D \) evaluation in the case of \( n = 10 \) statistics (see table 1). The corresponding result for the \( n = 50 \) case is improved about three times.

We have also calculated the mean absolute error (MAE) defined as

\[
\delta_{\text{MAE}} = \frac{1}{N} \sum_{k=1}^{N} |D_k - D_{\text{th}}|, \tag{22}
\]

where \( D_{\text{th}} \) is the theoretical value of the diffusion coefficient (\( D_{\text{th}} = 1.01 \, \mu m^2 s^{-1} \) for the considered diffusion process) and the sum is taken over all the simulated runs.

For the sample of ten runs with (and without) AIS for \( n = 10 \) particles, one obtains \( \delta_{\text{MAE}}(n = 10) = 0.40 \, \mu m^2 s^{-1} \) decreasing to \( \delta_{\text{MAE}}^{\text{AIS}}(n = 10) = 0.06 \, \mu m^2 s^{-1} \) when
Table 1. The comparison analysis of diffusion constant values ($2D$ values shown) found as the best fit before and after the AIS procedure. Each item is taken from different simulation of the $n = 10$ or $n = 50$ particles run in $d = 2$ dimensions. The scaling range is fixed according to figures 1(a) and (b) with the sampling time interval $\Delta t = 10^5 \tau \sim 0.01$ s.

| Run no | $n = 10$ particles | $n = 50$ particles |
|--------|--------------------|--------------------|
|        | Before AIS ('naked' data) | AIS data | Before AIS ('naked' data) | AIS data |
| 1      | 1.04               | 1.83              | 1.92               | 1.92              |
| 2      | 2.07               | 1.76              | 1.90               | 1.95              |
| 3      | 1.28               | 1.90              | 1.88               | 1.95              |
| 4      | 1.65               | 2.15              | 2.29               | 1.83              |
| 5      | 3.23               | 2.00              | 2.24               | 2.16              |
| 6      | 0.96               | 1.92              | 2.00               | 2.03              |
| 7      | 2.73               | 1.95              | 1.73               | 2.05              |
| 8      | 4.01               | 2.23              | 1.74               | 1.86              |
| 9      | 1.70               | 1.82              | 2.64               | 2.00              |
| 10     | 1.46               | 1.99              | 1.99               | 1.89              |
| $(2D)$ | 2.00               | 1.96              | 2.03               | 1.96              |
| $\sigma_{2D}$ | 1.0               | 0.15              | 0.28               | 0.10              |

AIS is switched on. The corresponding results for $n = 50$ particles are (in $\mu m^2 s^{-1}$) $\delta_{MAE}(n = 50) = 0.10$ and $\delta_{AIS}(n = 50) = 0.04$, respectively.

The positive feature of the AIS procedure can also be seen directly for the pure experimental data. We show in figure 6(a) the data taken in [4] for the diffusion of $n = 5$ latex spherical particles in pure water. One gets much better correspondence with the linear dependence when the AIS procedure is applied to these experimental points, as it is clearly revealed in figure 6(b). The obtained best fit for the diffusion constant corresponds now closely to the expected theoretical value $D_{th}$ which is not the case of the fit obtained by authors of [4].

5. Conclusions

The proper determination of the scaling range for the linear dependence $\langle r^2 \rangle \sim t$ is the crucial point in the data analysis. We argued that this scaling range behaves like $\lambda \sim n^\beta$, where the constant $\beta$ was determined as $\beta \sim 0.5$. The numerical simulation shows that for the case of mesoscopic particles diffusing in water the scaling range for $n \sim 10$ particles is as short as $\lambda \leqslant 3$ s. For $n < 10$, this scaling range is difficult to determine at all. In many papers, this fact is ignored which gives misleading results.

However, even if one remains in the scaling range regime, the results of simulated runs are not always statistically repeatable if too small statistics are considered. The minimal number of tracked particles to reveal the diffusion law is $n \geqslant 50$. One may nevertheless find the reasonable correspondence between theoretical predictions and experimental results even for the smaller number of tracked particles if the idea of AIS is applied. In this paper, we have described this idea and have shown how it works for simulated data as well as for data taken from the real experiments. It turns out that with AIS analysis one may get results within
10% of the expected theoretical value of the diffusion constant tracking just a few mesoscopic objects. The corresponding input data without AIS gives much bigger uncertainty of the order of 50% (see figure 4). The same applies when MAE is calculated. To decrease the uncertainty to the former level of 10%, one has to track roughly ten times more objects!

We have checked that for \( n = 50 \) tracked particles the AIS procedure decreases the statistical uncertainty in \( D \) from about 15% (the ‘naked’ data case) to \( \sim 5\% \). Simultaneously, the \( \delta_{\text{MAE}} \) drops down about twice (from 0.10 \( \mu\text{m}^2 \text{s}^{-1} \) to 0.04 \( \mu\text{m}^2 \text{s}^{-1} \)). The AIS procedure is here less impressive than for the \( n = 10 \) case but it still shows the significant improvement in data results.

This way it is quite possible to collect data giving a very good prediction for the diffusion constant even in less professional labs where one is not able to measure simultaneously the signals coming from a bigger number of objects. Hence, other important physical constants (like, e.g. Boltzmann constant \( k \) or the Avogadro number \( N_A \)) can be deduced with high accuracy what is often the crucial point in such experiments.

We have also made the simulations for liquids with other viscosities. The same final conclusions as for the case of water can be formulated. Because of very similar results we do not show them explicitly in this paper, but we believe they should be studied in the way of the numerical simulation in any case before the actual experiment is planned.

After this paper had been prepared for publication, we received the paper by Jia et al [9] with the discussion of the one-particle Brownian motion experiment. Because this experimental problem is strongly related to the content of our paper, we would like to use this opportunity to make a few comments on the results obtained in [9].

Jia et al use the time average value of the squared displacement \( \langle r^2 \rangle_t \) instead of the stochastic average \( \langle r^2 \rangle \), i.e. the average taken over many independent stochastic processes. In fact this approach is incorrect. The linear dependence on time in Brownian motion (diffusion law) is proven theoretically in a very strict manner as a result of the convolution theorem applied to the independent stochastic steps with the conditional probability given by the Gaussian distribution. The architecture of such mathematical considerations is based on the stochastic processes and it proves that unconditional variance \( \sigma^2 \) calculated over different stochastic processes has the property \( \sigma^2 = \langle r^2 \rangle \sim t \) (see, e.g. [10]). The use of the term ‘time average’ for the experimental data collection is here very much misleading and it is the fact that should be clarified in [9].

Moreover, the linear fits (see, e.g. figure 2 in [9]) are done without the boundary condition \( \langle r^2 \rangle = 0 \) at \( t = 0 \). The latter requirement is a must because it is not only the theoretical interception point (as Jia et al do confirm) but also the experimental one! (The particle starts at \( t = 0 \) from \( r = 0 \). If one uses this initial condition, the scaling range for the linear dependence will be visible as we have described in our paper. We are sure that the numerical values of all the slopes and their uncertainties in [9] would have been changed significantly if the zero interception point was taken into account. Unfortunately, the authors avoided the quantitative discussion of the obtained data so this comparison is difficult to be analysed.

References

[1] Perrin J B 1909 Ann. Chim. Phys. 19 5–114
[2] Salmon R, Robbins C and Forinash K 2002 Eur. J. Phys. 23 249–53
[3] Nakroshis P, Amoroso M, Legere J and Smith Ch 2003 Am. J. Phys. 71 568–73
[4] Groczylo T and Debowska E 2005 Eur. J. Phys. 26 827–33
[5] Newburgh R, Peidle J and Rueckner W 2006 Am. J. Phys. 74 478–81
[6] Einstein A 1905 Ann. Phys. 17 549–60
See also Einstein A 1956 Investigations on the Theory of Brownian Movement transl. by A D Cowper (New York: Dover)
[7] Langevin P 1908 C. R. Acad. Sci., Paris 146 530–3
[8] Smoluchowski M 1906 Ann. Phys. 21 756–80
[9] Jia D, Hamilton J, Zaman L M and Goonewardene A 2007 Am. J. Phys. 75 111
[10] Tsallis C, Levy S V F, Souza A M C and Maynard R 1995 Phys. Rev. Lett. 13B 3589