PHY324: Determining the Boltzmann Constant Using the Motion of Beads due to Thermal Diffusion

By: Gurmanjot Singh and Aran Wade

Due: 16 February 2023

1 Abstract

We gather displacement-time data for 21 beads constrained to an approximate 2 dimensional plane. With this, we have computed the Boltzmann constant in three ways: (i) fitting the mean squared displacement over time to $4Dt$, (ii) fitting the histogram of step size distance between time intervals to a Rayleigh distribution with $\sigma^2 = 2D\Delta t$, and (iii) computing D through the maximum likelihood estimate of the Rayleigh distribution of the step size.

The first two methods seem to generate very similar results up to uncertainties and fit the data more accurately, but do not match the actual value of the Boltzmann constant computed through $k = DT/\gamma$. On the other hand, method (iii) provides a value for Boltzmann's constant that coincides with the actual value, but does not exactly fit the data. This lets us characterize the comparison between curve fitting methods and MLE methods for analyzing data, particularly in the case of Brownian motion.

2 Introduction

The behaviour of particles under thermal forces is considered to be largely random. This is particularly due to Newton's equation for a particle with a thermal force X in one dimension can be written as:

$$m\ddot{x} = -\gamma\dot{x} + X$$

which, after some manipulations and taking a time average (with $\langle xX \rangle = 0$), gives us an expression for $\langle x^2 \rangle$ as:

$$\partial_t \langle x^2 \rangle = 2kT/\gamma + Ce^{-\gamma/mt}$$

for Stokes' drag coefficient γ of the particle and the temperature T of the medium. Since the term γ/mt is small for the times we will be considering in our experiment, we will neglect the exponential.

This gives us the fact that $\langle x^2 \rangle = 2Dt$ if we set the origin to be at the initial point of x, and write the diffusion coefficient $D = kT/\gamma$.

In n dimensions, since each coordinate is independent, this is just $\langle r^2 \rangle = \langle x_1^2 \rangle + \cdots + \langle x_n^2 \rangle = 2nDt$. Particularly, $\langle r^2 \rangle = 4Dt$ in two dimensions and $6Dt$ in three dimensions.

Alternatively, since we know that the particle's behaviour is Brownian, we know that its step size every constant time increment Δt is governed by the polar integrated Gaussian distribution:

$$p(r) = \frac{1}{2D\Delta t} e^{-r^2/(4D\Delta t)}$$

which is a Rayleigh distribution with $\sigma^2 = 2Dt$ as its parameter.

In this experiment, we will be observing this form of motion for beads in two dimensions, with a Stokes drag $\gamma = 6\pi\mu r$ for water viscosity μ and radius of the beads r. With this information, we can fit its mean squared displacement to $\langle r^2 \rangle = 4Dt$ over time.

Alternatively, we could also fit all the step size data–given by the magnitude of displacement for every time increment r_i–to a Rayleigh distribution from a histogram. One of these fitting methods would be to use the well-known maximum likelihood estimate of a Rayleigh distributed data set of size N, as [1]:

$$\sigma^2 = 2D\Delta t = \frac{1}{2N} \sum_{i=1}^{N} r_i^2$$

Note that this is very similar to fitting $\langle r^2 \rangle = 4Dt$, except observing that we are replacing the time parameter with a time increment, and instead of averaging the total displacement magnitude squared at each time, we are averaging over all incremental displacement magnitudes from every time.

We will then be computing $k = D\gamma/T$ with each method, which in comparison with our fits will give us a way to judge the strength of that fit.

3 Methods

To execute this experiment, we are suspending a solution of fluorescent luminescent beads on a glass slide. Particularly, we will place an approximate 50μl on a square of grease on the glass, and then place another glass slide above it to isolate it. Note this placement must be completely flat to prevent any focusing errors when using the microscope and to reduce the third degree of freedom in its motion.

To perform the measurement, we place the slide under a microscope as mentioned, and search for beads in the solutions to observe with a fluorescent light illuminating the slide. Once a bead has been found, we use some camera-microscope software to take a video of the bead moving over time after centering it on the screen, with 120 samples taken over 60 seconds, or a $\Delta t = 0.5 \pm 0.03$ s increment, where the uncertainty is standard for the software (see 1).

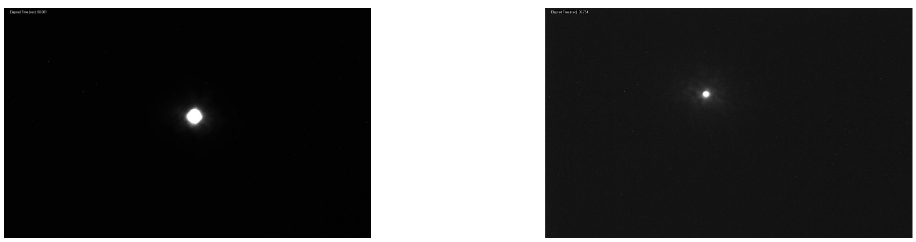

Figure 1: Examples of beads observed and tracked using camera.

Then, we use a physics tracking application to give us the position-time data of the bead in pixels. Using the conversion of $1\text{px} = 0.12408 \pm 0.003\mu\text{m}$ gives us the actual positions in SI units, given the zoom of the microscope. We repeat this for 21 different beads in the solution.

During this acquisition, we noticed several details about the beads. Mainly, the beads came in a variety of different shapes, sizes, and possibly masses. This changed the way some of the beads behaved, as some did not move at all (see 2).

Figure 2: Examples of non-moving beads, hence being omitted due to likely not fitting the assumptions.

Since we are assuming, for the sake of uniform calculations, that the bead are all spherical and have a diameter $d = 1.9 \pm 0.1\mu$m–the diameter at which the motion is visible–we will be neglecting larger non-moving beads as they do not satisfy that assumption.

Moreover, we note that vibrations and movement from the external environment perturbed the motion of the bead solution, which will add to the uncertainty. The tracking software also was not always perfectly centered at the same point it started at on the bead, which gives rise to some additional fluctuations in the data.

Together, this uncertainty amounts to one that is of larger order than just ± 0.003 μm in the position measurements, which we approximate as ± 0.01 μm.

We also measure the other important parameters of the system, particularly the temperature $T = 298 \pm 0.5$ K during the experiment. Given that the viscosity of water at 293 K is 1 ± 0.05 centipoise, or $\mu_0 = 0.001 \pm 5 \times 10^{-5}$ Pas, and decreases by 2% for every degree increase in temperature, we get that $\mu = 9.03 \times 10^{-4} \pm 5 \times 10^{-5}$ Pas.

Using this and the fact that $r = 8.5 \times 10^{-7} \pm 5.0 \times 10^{-8}$ m as the radius of the beads that we assume, we can compute the Stokes drag $\gamma = 3\pi\mu r$ or $\gamma = 1.62 \times 10^{-8} \pm 1.24 \times 10^{-9}$ Ns/m, which will let us directly compute $k = D\gamma/T$ later in this paper.

As a side note, although we will be modelling this system as two dimensional, we found that some beads sometimes went up and down relative to the focal axis, corresponding to phasing in and out of focus. To remedy this during the video tracking, we just made sure to change the focus of the microscope to match the motion of the bead so we could keep tracking it.

The importance of this is that not all of our data is the result of perfect two dimensional motion, but rather is more of a projection of mostly two dimensional motion with some third axis deviation. This means our two dimensional models of the system might not fit the parameters to the actual values of D and k, but instead to the versions of those parameters if universe's physics were two dimensional.

Overall, to compute uncertainties, we will be using the standard quadrature computations for multiplication and the results of curve fitting. When plotting the histograms, they will be scaled so their integral is of unity. The uncertainty of the histogram frequencies will also be the standard square root of the magnitude, scaled with the normalization. This is to properly fit the Rayleigh density to the histogram.

We can hence proceed to the analysis of the results.

4 Results

To first understand the behaviour of each bead, we provide an example of a bead trajectory that we measured, given by 3.

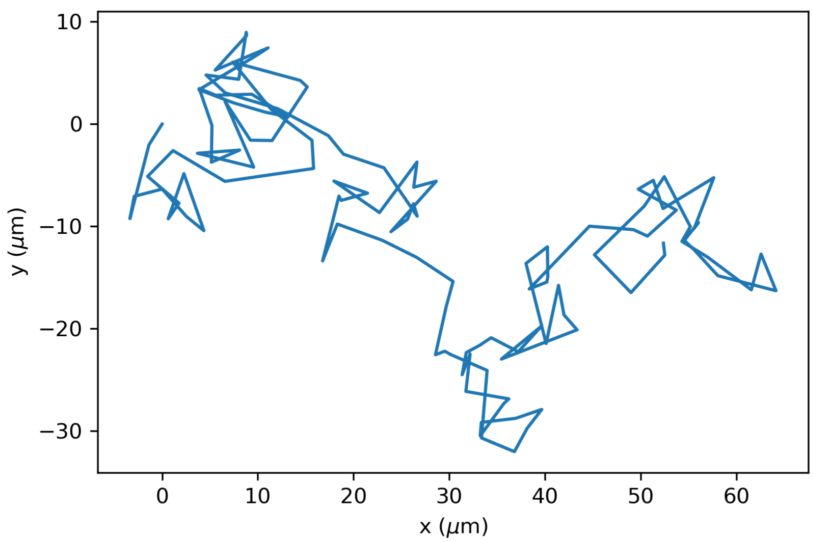

Figure 3: Trajectory of first bead measured over 60 seconds with 0.5 second time increment.

This indeed resembles some form of a random walk, which is also the case with the other beads. Note that we have set the initial position to be the origin, so when we take the time average over the square distance from the origin, it is consistent over all the other beads.

We do not expect the bead's distance squared to be linearly correlated to time individually, but when we take a time average over a large sample of beads, it should start to more closely reflect this relation. When we do exactly this, our result can then be fitted as in 4 (next page).

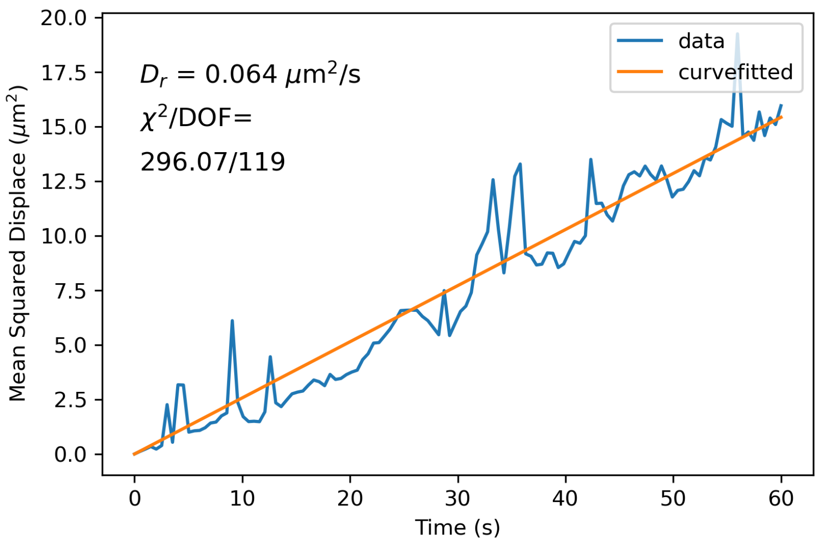

Figure 4: The mean of r^2 over the beads at each time fitted with a linear trend in time.

The uncertainty on this value of diffusion coefficient–$D_r = 0.0643$ μm^2/s– is computed as $\Delta D_r = 8.37 \times 10^{-4}$ μm^2/s. We note that the uncertainties on the r^2 terms were calculated by the uncertainty power laws and quadrature. Even then, most of them were extremely low, not accurately reflecting the actual errors involved. Because of this, we decided to take the maximum of these uncertainties and assigned it to each r^2 value at each time.

Even with this liberal estimate, the chi squared coefficient of the fit was extremely high, as deviations from the actual result were much higher than the uncertainties in the computation, as can be seen by the massive spikes from the general line trend.

These inconsistencies are just an artifact of the actual motion being random–we would expect these spikes to vanish if we had a much larger pool of beads measurements from which we took these measurements–and hence even though we have non-coincidence with the actual values, the fit qualitatively still looks good.

This is strengthened by the fact that, when we instead consider fitting the step size distances of all the beads at once in a single histogram, we get a very similar diffusion coefficient consistent up to uncertainties, $D_c = 0.0691 \pm 0.0046$ μm^2/s (see 5).

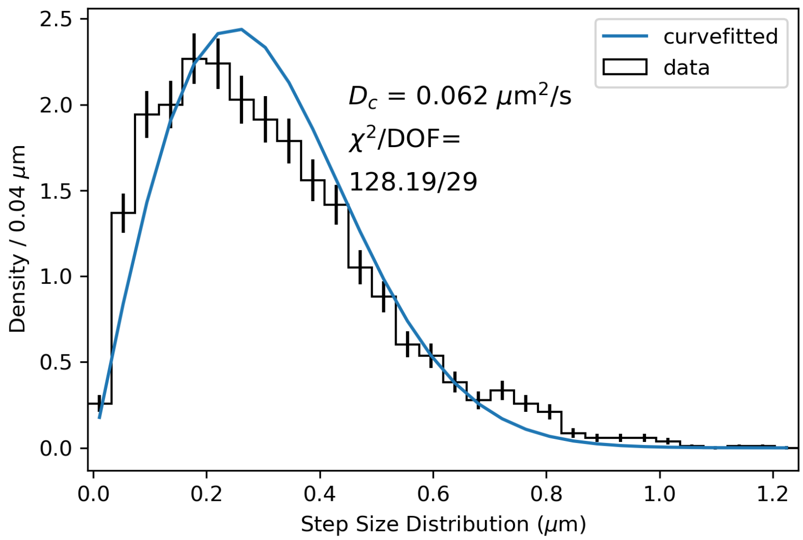

Figure 5: Plot of distributions of step sizes of beads in histogram with 30 bins over $(0, 1.2)$ µm range fitted to Rayleigh distribution.

where we have chosen the bins by trial and error to minimize the chi-squared. Again, we have a very high chi squared value, which is due to the fact that there is some deviation from the actual distribution at around 0.8 μm and near the peak from the randomness of the system. Visually, as in the first case, it still resembles a good fit, on account of the uncertainties not being large enough.

If we were to fit this instead with the maximum likelihood estimate of D, given by:

$$D_m = \frac{1}{4N\Delta t} \sum_i r_i^2$$

with the corresponding uncertainty given by quadrature, we get a value of $D = D_m = 0.298 \pm 0.0434$ μm^2/s, which is on an order of around 4 magnitudes different from the D_r and D_c values. This, when superposed over the actual distribution and the curve fit, shows the tremendous discrepancy (see 6).

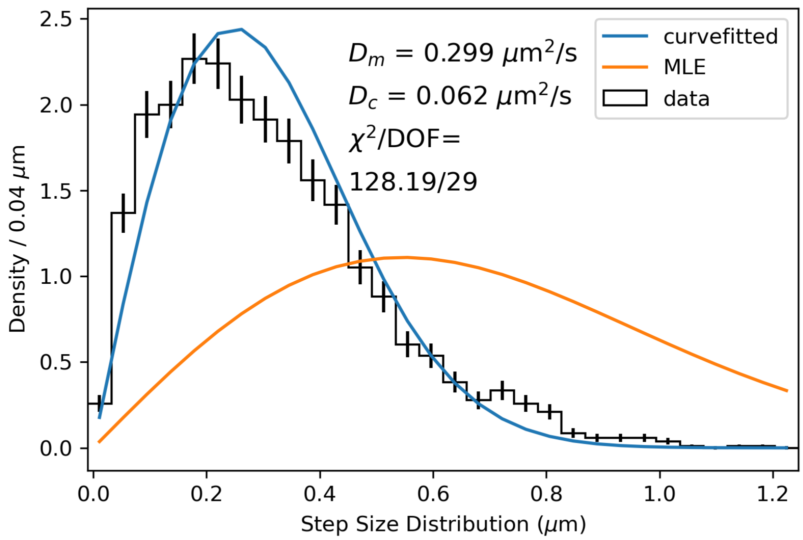

Figure 6: Same plot as 5 but with the maximum likelihood fit.

There is no need to calculate the chi squared as we already expect it to be very large, regardless of the number of bins. There are several possible reasons for this inconsistency to occur. All of these reasons stem from the fact that MLE searches for the parameter that maximizes the likelihood of our dataset occurring if it was Rayleigh distributed.

Primarily, we observe that the MLE computation of D has an equal weight summed contribution from each r_i^2. We know that squaring a number between 0 and 1 decreases its magnitude by an amount proportional to its proximity to 0 (so smaller numbers squared get much smaller and larger numbers get affected less).

Because of this, and the fact that there are some fluctuations in the real distribution, with the smaller values being much closer to the left of the peak and a small bump of r values occuring at the tail near 1 μm, the smaller r_i values contribute much less to D while the few additional values that are larger contribute on average more.

In addition, the actual peak is not very symmetric, with many more values larger than it than smaller due to it being left shifted slightly more than usual–which can be seen just by comparing it to the curve fitted distribution–and so the values in the middle (near the peak of the MLE fitted distribution) contribute the most.

Note that this shows the sensitivity of the maximum likelihood estimate, as even small changes in the data set, as long as they are sufficiently chosen, should be able to make the estimated parameter that maximizes likelihood several factors different from the one that fits the curve the best.

On the other hand, this slightly distorted shape may also be due to the fact that this is data of restricted three dimensional particle trajectories projected to two dimensions as well, rather than just being a result of randomness. In this case, the maximum likelihood would be weighted such that it would always estimate a parameter value near D_m as we computed.

Regardless of what caused this difference, since D_m does maximize the likelihood of our dataset occurring from a Rayleigh distributed random variable, regardless of how well it fits, it still is the distribution that is most likely to generate that set of results, so it should not be discounted.

This is reinforced by the results of computing the corresponding Boltzmann constants for each case, which gives us the following table of values, after performing sufficient unit conversions to SI and the corresponding uncertainty calculations for $k = D\gamma/T$: We note that both k_r and k_c are consistent with each

$D \mapsto k$	k (J/K)	Δk (J/K)
$D_r \mapsto k_r$	3.49×10^{-24}	2.71×10^{-25}
$D_c \mapsto k_c$	3.34×10^{-24}	3.33×10^{-25}
$D_m \mapsto k_m$	1.62×10^{-23}	2.66×10^{-24}

other, while k_m is on a scale of approximately 10 above them. On the other hand, the accepted value for the Boltzmann constant is $k_b = 1.38 \times 10^{-23}$ J/K, which k_m clearly coincides with, while the other two are far away from even coinciding with k_b. The percentage differences from k_b with the percentage uncertainties are given below to outline the discrepancies: So although the fitted

Percentage Differences	Percentage Uncertainties
74.7%	7.75%
75.8%	10.0%
17.4%	16.4%

values are more precise, they are much less accurate than the maximum likelihood estimate, which is consistent in its percentage error and uncertainty from k_b.

5 Analysis

To understand the results of our modelling, we analyze the possible reasons behind the discrepancies. Mainly, this suggests that fitting to the model exactly, especially with step size distance data that does not account for changes in the focal axis, may not be the best method for determining the D value as we do not account for that degree of freedom.

If we accounted for this, all the r_i values would be larger due to the z contributions to the magnitude, shifting the distribution to a Maxwell-Boltzmann distribution, which by itself when fitted to a Rayleigh distribution would have a larger predicted value for D and a better value for k.

All the while, the maximum likelihood seems to accidentally predict the actual value of D and k_b, possibly due to some mathematical property of the likelihood estimate that remains invariant under the projection of the Maxwell-Boltzmann distribution to two dimensions. Particularly, the maximum likelihood estimate for a Maxwell-Boltzmann distribution predicts that:

$$2D_b\Delta t = \frac{1}{3N}\sum_i R_i^2$$

given that $R_i^2 = r_i^2 + z_i^2$ for the unaccounted z_i step displacement. This means that $D_b = \frac{2}{3}D_m + D_z$, where [2]:

$$D_z = \frac{1}{6\Delta N}\sum_i z_i^2$$

Since we expect, by symmetry, the mean squared displacement in the z direction to be a third of the mean squared displacement in every direction, we expect that $D_z \approx \frac{1}{3}D_b$ up to some fluctuations due to randomness that vanish in the data limit.

This means that, after simplification, we expect $D_b \approx 2/3D_m + 1/3D_b$ or $2/3D_b \approx 2/3D_m$, which gives us the key result that $D_b \approx D_m$, with true equality $D_b = D_m$ in the limit.

Hence, the maximum likelihood estimate we calculate for D in the case where we are fitting a Rayleigh distribution to our projected step size displacements r_i is equivalent to the computation for D in the actual three dimensional random walks of the beads.

Since we know that the resulting D_b value in the three dimensional case must give us the real Boltzmann constant from $k_b = \gamma D/T$, we now understand why D_m gave us the same result, even though it came from our insufficient two dimensional model. This also explains why the value for D_m was so far off from the other fitted values, verifying our speculation that it may be the model instead of the data that is causing this.

To fix this the model itself, it might suffice to consider modified R_i data to behave as though it is from the three dimensional model, via $R_i = \sqrt{3/2r_i^2}$, and then fit the distribution of the R_i to a Maxwell-Boltzmann distribution. This would give us much better estimates of D values directly from the curve fitting.

6 Conclusion

In brief, through the measurement of bead motion and performing the aforementioned three methods of extracting D and k, we have observed that the fitting methods coincide with each in uncertainty, but are non-trivially different from the accepted value of $k = k_b$ and the corresponding coincident value from maximum likelihood estimates.

Although initially this would presume some form of error in uncertainty estimates or the data acquisition, the confidence of k_b and that of the maximum likelihood result k_m suggested that data randomness most probably did not cause this, since it would then be very unlikely for these two to coincide so well.

Alternatively, from a modelling standpoint, we found that the maximum likelihood estimate of D in the Rayleigh distribution, when the data is a two dimensional projection of three dimensional displacements, is equivalent to the maximum likelihood estimate of D through the original distances obeying a Maxwell-Boltzmann distribution. Because of this, k_m is expected to coincide with k_b, since the real diffusion coefficient is extracted.

Moreover, this shows how much of a difference it can make to model something that has three dimensional symmetry with a two dimensional model of that symmetry–the behaviour of particles constrained to a plane is very different to that of particles freely moving in space. This detail is important when performing other experiments in physics and is good to keep in mind to avoid mischaracterizations of real world phenomenon.

References

[1] Fitri Ardianti et al, IOP Conference Series: Materials Science and Engineering (2023), *Estimating parameter of Rayleigh distribution by using Maximum Likelihood method and Bayes method*. Last accessed 16 February 2023, `https://iopscience.iop.org/article/10.1088/1757-899X/300/1/012036/pdf`. DOI: `10.1088/1757-899X/300/1/012036`.

[2] Iden Alkanani et al, Baghdad Science Journal (2017), *Bayes and Non-Bayes Estimation Methods for the parameter of Maxwell-Boltzmann Distribution*. Last accessed 17 February 2023, `https://www.iasj.net/iasj/download/72de89afad8b7e9c`. DOI: `http://dx.doi.org/10.21123/bsj.2017.14.4.0808`.