P1: Measuring Boltzmann's constant with a microscope

Project 1, all students. Report due Sun 7 March, defense Tue 9 March

The question

In 1908 Jean Perrin used Brownian motion to measure Avogadro's number, and with it Boltzmann's constant. You will do the same with beads in water, a microscope and a camera. Your question is:

> How well can you determine k_B from the Brownian motion of beads, and what limits the accuracy?

The physics is in the notes on active and passive motion: random walks, the diffusion equation, Stokes drag, the Langevin equation and the Einstein relation $\xi D = k_B T$. The data come from the Brownian motion lab.

What you deliver

1. **A report** (one PDF) with your result, how you got it, and why you believe it. LLM help with writing is allowed and expected. Keep it short: the figures and the argument matter, not the prose.
2. **An LLM log** as an appendix (one page is enough): what you used an LLM for, where it was wrong or misleading, and how you found out.
3. **An oral defense** on Tue 9 March: 5 min presentation of your core result, then about 10 min of questions (see below).

Milestones

	When	What you show
M1	Tue 9 Feb (W6, lab)	Your own image sequence, and particles found in one frame with `trackpy.locate`
M2	Thu 25 Feb (W8)	**Validated core figure:** MSD against lag time with a fitted $D \pm \delta D$, the same pipeline on a synthetic movie recovering the known D, and a first value of k_B
M3	Sun 7 Mar	Report and LLM log submitted on Canvas
M4	Tue 9 Mar (W10)	Oral defense

What the project should contain

1. Measurement

- Track the beads in your 5 min movie: locate, link, filter out short tracks and stuck particles, and subtract drift.
- Plot the ensemble MSD $\langle \Delta x^2 + \Delta y^2 \rangle$ against lag time on **linear axes** and fit a straight line. Explain what the slope and the intercept mean.
- From the slope get D. Then use the Einstein relation with Stokes drag, $D = k_B T/(6\pi\eta r)$, to get k_B, with an uncertainty.

2. Three checks that the result is right

Getting trackpy to give *a* number is easy, with or without an LLM. Showing that the number is *right* is the project. Include all three checks:

1. **Synthetic ground truth.** Make a synthetic movie with known D and the same pixel size, frame rate, particle size and contrast as your data (use `synthetic_brownian.py`, or write your own and explain it). Run *exactly* your pipeline on it. Do you get back the D you put in? If not, your pipeline is wrong on the real data too.
2. **Physical expectation.** Before you trust a fit, calculate the D you expect from the nominal bead size, the viscosity of water at the lab temperature and the known k_B. Is your measured value close? If not, why?
3. **Sensitivity.** Show how D changes when you vary the key parameters: `minmass`, `search_range`, the minimum track length and the range of lag times in the fit. Argue why your chosen values are defensible.

3. Model comparison

Compare with a random-walk simulation (for example `randomwalk_exercise.py`). What step length and time step must the random walk have to match your experiment? Why does the mass of the bead not appear in the result?

4. Physics and biology

- Show that the beads are at low Reynolds number and that inertia is irrelevant on the time scale of your frames.
- Discuss one biological consequence of diffusion, for example how long it takes a protein to cross a cell, or the oxygen supply limit for a bacterium.

Working with an LLM on this project

Many students have struggled to get reasonable results from trackpy. You are now expected to use an LLM to help you through that: to set up Python, to write and debug the pipeline, and to understand what the parameters do. Read Working with LLMs first. Three habits make the difference:

- **Give it the physics.** Tell the LLM the pixel size, frame rate, bead size and expected D. Then it can help you choose `diameter` and `search_range` sensibly, instead of guessing.
- **Ask it to explain, not only to write.** For every function call in your pipeline you should be able to say what it does and why that parameter value. You will be asked in the defense.
- **Test, don't trust.** The synthetic movie is the test. An LLM can produce code that runs without errors and still gives a D that is wrong by a factor of two.

The oral defense

About 15 min per student, in groups of 3–4 so that you hear each other's questions. Bring your report on a laptop. Typical questions:

- Derive $\langle x^2 \rangle = 2Dt$ from the random walk, or from the diffusion equation.
- Where does $\xi D = k_B T$ come from? What does the Langevin equation assume?
- What does `search_range` do? What happens to D if it is too small? Too large?
- Your MSD fit has an intercept. What does it mean physically?
- Which check convinced you most that your result is right? What did the LLM get wrong?
- If you repeated the experiment with beads twice as large in glycerol–water, what would change?

The defense is passed when you can explain your own result and the physics behind it. "The LLM wrote that part" is not an answer to "what does this line do?". If the defense is not passed, you get one new attempt.