PHYS CS 5: Non-Equilibrium Statistical Mechanics

Sidharth Kannan

Winter 2026

Contents

1 Lecture 1- Stochastic Processes

1.1 Introduction

Statistical mechanics, broadly speaking, is the study of complex systems with many, many degrees of freedom. We want to understand how such systems behave, even though they are so complex that even though we know the microscopic equations of motion, we can't even begin to dream of solving them. Our saving grace is the idea that when these systems are sufficiently complex, they exhibit certain collective behaviors that make it possible to understand how the systems behave *on average*.

In PHYS 119A and PHYS 119B, you study these systems through the lens of systems *at equilibrium*, where the macroscopic quantities like temperature, pressure, and volume do not change over time, *on average*, and the fluctuations around that average are small enough to be physically inconsequential. At first glance, this seems to be highly restrictive. After all, the whole rest of physics is dedicated to understanding dynamical behaviors of various quantities. Why is it that we can suddenly restrict ourselves to assuming that things are constant? Well, it turns out that if you leave *most* systems alone for long enough, they eventually settle down to equilibrium. Thus, we find that many systems in the real world are in equilibrium, and so the machinery of equilibrium statistical mechanics is extremely useful. However, there are a few problems and outstanding questions:

1. I said *most* systems reach equilibrium when left alone for long enough. How do we know that? Are there systems that *don't* ever equilibrate? What controls how quickly systems reach equilibrium?

2. While systems reach equilibrium when left alone for long enough, many natural systems are not being left alone. Things are constantly either interacting with each other, or being poked by pesky experimental physicists. As such, we would like to understand how systems behave when poked.

3. When the systems are extremely large, the central limit theorem helps us justify why we can disregard fluctuations. Unfortunately, physicists also feel the need to study small, medium, and large (but not quite *extremely large*) systems. For example, we study cells and crystals, which can consist of hundreds, thousands, millions, or billions of molecules. These systems are many orders of magnitude less complex than the Avogadro's number of molecules in a bucket of water or container of ideal gas. They fit into that middle, *mesoscopic* regime where things are still far too complex to model exactly with classical and quantum mechanics, but far too simple to completely ignore fluctuations and deviations from equilibrium.

These are the types of problems we will try to answer in this course. Finally, I would like to point out that non-equilibrium statistical mechanics consists of *statistical laws* which are generally not tied to any particular microscopic model. As such, its methods are very broadly applicable. In this course, we will explore applications to biophysics, astrophysics, and fluid mechanics, among other domains, but part of what makes this subject so beautiful (in my opinion) is that its laws apply to almost *any* complex system. As such, its methods have found broad application in economics, computer science, social modeling, and more. I will try to highlight some of these applications on the problem sets, but I encourage you to explore them independently as well!

Our exploration will be divided roughly into three parts. We will start by exploring the role of fluctuations in complex systems, both in and out of equilibrium. We will develop the mathematics of stochastic processes, and employ it to understand diffusive processes and Brownian motion. After this, we will turn our attention to *kinetic theory*, and the microscopic foundations of statistical mechanics. Finally, we will return to our study of fluctuations by turning out attention to how systems respond to various kinds of poking.

1.2 Stochastic Processes

In equilibrium statistical mechanics, the primary mathematical objects are random variables, and their corresponding probability distributions. We have a system that we are studying, S, and it can occupy a set of states, $\Omega = \{X_i\}$. Our goal, then, is to determine a closed form expression for the probability of being in a particular state X_i in equilibrium. The two most important examples of this are the microcanonical and canonical ensembles. In the microcanonical ensemble (an isolated system), the fundamental assumption is that all microstates are equally likely. That is, for any X_i,

$$P(X_i) = \frac{1}{|\Omega|}$$

where the $|\cdot|$ symbol denotes *cardinality*, or the size of the set Ω. In the canonical ensemble (a system in thermal contact with a reservoir), we know that the distribution instead takes the form of the Boltzmann distribution,

$$P(X_i) = \frac{1}{Z}\exp(-\beta E(X_i))$$

We can also determine the probability distributions of various other quantities. The most famous example of this is Maxwell's speed distribution, which gives the distribution of speeds of particles in a ideal gas:

$$P(v) = \left(\sqrt{\frac{m}{2\pi k_B T}}\right)^3 4\pi v^2 \exp\left(-\frac{mv^2}{2k_B T}\right)$$

In non-equilibrium statistical physics, we are primarily concerned in how these systems evolve in time, and so we need to understand probability distributions over trajectories through phase space. Thus, our first order of business will be developing the mathematics of *stochastic processes*.

As a simple example, let's begin by considering a *random walk* on a periodic lattice. The system has 4 states, and at each time step (say, every second), the system can make a transition. The allowed transitions are a clockwise move, counterclockwise move, and remaining where you are. Each can happen with equal probability. Our goal will be to determine the probability of finding the walker at each site after k time steps. The system is illustrated in Fig. 1.

This is a pretty abstract mathematical model, and it may help to have a more concrete example in mind. If this helps you, then you can imagine that our system is a two-spin Ising model, and at each time step, we allow *one* spin to try and flip.

Let's say the system starts in state 1. We can write the probabilities of the system being in each of the states as a vector, called the *probability mass vector*.

$$\vec{p}(t) = \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{pmatrix}$$

Naturally, since the walker must be on one of the four sites, we have the condition $\sum_{i=1}^{4} \vec{p}_i(t) = 1$. The subscript denotes the index into the vector. At $t = 0$, this vector is then

$$\vec{p}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

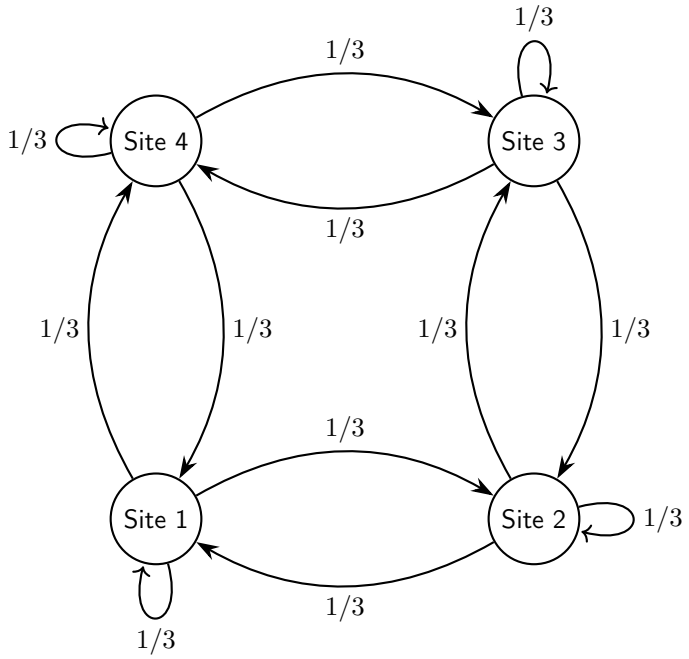


Figure 1: A 4-site lattice. The random walker can move from site to site.

After one time step has passed, the system could either still be in state 1, or it could have transitioned to states 2 and 4. The new probability mass vector is

$$\vec{p}(1) = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \\ \frac{1}{3} \end{pmatrix}$$

Computing the probability mass for the next time step is a little harder. How would we do it? Well, let's start by focusing on a single site. Say we want to know the probability of finding the system in state 3 at time step 2. I'll denote this $\vec{p}_2(t=2)$. We can see that the probability should be given by adding the probabilities of each of the possible ways that the system could get to state 2 at $t = 2$. Let's denote the probability of transitioning from state 1 to state 2 as T_{21}, from state 2 to state 2 as T_{22}, and so on. Then

$$\vec{p}_2(t=2) = T_{21} \cdot \vec{p}_1(t=1) + T_{22} \cdot \vec{p}_2(t=1) + T_{23} \cdot \vec{p}_3(t=1) + T_{24} \cdot \vec{p}_4(t=1) = \frac{2}{9}$$

Each term in this sum is saying that the contribution to $\vec{p}_2(t=2)$ from state j is given by the probability that I was in state j at the previous timestep *times* the probability that I transition from j to 2.

More generally, we can write that for *any* state i and timestep k,

$$\vec{p}_i(t) = \sum_j T_{ij} \cdot \vec{p}_i(t)$$

This notation may be suggestive to you all. We have already introduced a probability mass vector, and now it may be clear that the transition probabilities, T_{ij} can be arranged into a matrix, and then this is just a matrix-vector product. The *transition matrix*, T, has at the i-th row and j-th column is the transition probability T_{ij} of transitioning from state j to state i. For our 4-site lattice, this matrix is:

$$T = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

Note that each column of the transition matrix sums to 1, which must be the case since if the walker is in state j, it must transition to some state, i. The power of this formalism lies in how simply we can describe the *time evolution* of the system. The probability distribution at time $t+1$ is found by applying the transition matrix to the distribution at time t:

$$p_i(t+1) = \sum_j T_{ij} p_j(t)$$

We can check this with our example, where we start with certainty at state 1 at $t = 0$:

$$\vec{p}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

After one time step, the state of the system will be:

$$\vec{p}(1) = T\vec{p}(0) = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 0 \\ 1/3 \end{pmatrix}$$

This result tells us that after one step, there is a 1/3 probability of finding the walker at Site 1 (stay), Site 2 (moved CW), or Site 4 (moved CCW), and zero probability of finding it at the non-adjacent Site 3, which is exactly what we expect. Applying the matrix T repeatedly allows us to predict the probability distribution at any future time step.

Now, let's look at the long-time behavior of the system. That is, what happens when we evolve the system according to the transition matrix for a long period of time? Mathematically, we are interested in the matrix $\lim_{k\to\infty} T^k$. We can gain insight into this matrix by diagonalizing T:

$$T^k = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{pmatrix}^k \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}$$

Now the matrix of eigenvalues is a diagonal matrix, and so raising it to a power is equivalent to raising each of the entries on the diagonal to that power. Then, we easily see that all of the diagonal entries will vanish as k gets very large, except for the first, which will stay as 1. Then,

$$\lim_{k\to\infty} T^k = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \tag{1}$$

If we multiply this matrix on the left by *any* real, non-negative, stochastic vector, we will get $\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}^T$. So, we see that this distribution is the equilibrium distribution of our periodic random walker. Now, this should be fairly intuitive. The random walker has equal probability of walking to each node, so in the long time limit, it is unsurprising that it would exhibit no preference for one site over another. Some comments are due about the properties of transition matrices.

The first important observation is that the elements of $\mathbf{P}$ are real numbers on the interval $[0, 1]$, and its columns sum to 1. This is true by construction, since each element is a probability; the total transition probability from one state to all other states is 1. In our notation,

$$\sum_i T_{ij} = 1 \qquad \text{for each column } j \tag{2}$$

Next, *by construction*, the transition matrix is not necessarily symmetric. A symmetric transition matrix would imply that for every pair of states, the forward transition from state $i \to j$ has the same probability as the reverse transition, $j \to i$. This may be true in some cases, like the random walk above, but not in general. This has an important consequence: the eigenvalues are not necessarily real. Nevertheless, one can introduce eigenvectors which satisfy

$$T\chi_i = \lambda_i \chi_i \tag{3}$$

It can be shown that transition matrices always have the eigenvalue 1, and, under some mild conditions (*irreducibility* and *aperiodicity*, for those who are curious), the right eigenvector with eigenvalue 1 is unique. We call this eigenvector the *stationary distribution*, π, because it no longer changes as time moves forward:

$$T\pi = \pi \tag{4}$$

1.2.1 Basic concepts

The example above was a simple example of a stochastic process. We now turn to a more formal definition of stochastic processes.

There are two complementary views that one can have on a stochastic process. We can view the process as a probability distribution over functions. That is, we can view the stochastic process as assigning a probability to every possible trajectory that the random walker could take, or we can view it as a sequence of random variables, X_t, representing the walker's position at time t, independent of the trajectory it took to get there.

Since we typically cannot write an explicit formula for the trajectories, in this class, we will tend to prefer the latter description. Let's say that some quantity, X is described by our process. Fix a finite collection of times $t_1, t_2, \ldots, t_n$. The random vector

$$\big(X(t_1), X(t_2), \ldots, X(t_n)\big)$$

which represents the values of the quantity, X at n different times has a joint probability mass

$$P_n(x_1, t_1;\ x_2, t_2;\ \ldots;\ x_n, t_n). \tag{5}$$

In this notation, $P_1(x_1, t_1)$ is the probability that X_1 takes the value x_1 at time t_1, and $P_2(x_1, t_1; x_2, t_2)$ is the joint probability that X takes the value x_1 at t_1 and x_2 at t_2. There are certain conditions we need to impose on these distributions in order to make them valid probability masses of some stochastic process:

(i) $P_n(x_1, \ldots, x_n) \geq 0$. I.e. a probability mass cannot be negative.

(ii) Symmetry under relabeling: permuting the pairs (x_i, t_i) and the corresponding arguments leaves P_n invariant up to the same permutation.

(iii) Marginalization (consistency): for any choice of t_{n+1},

$$P_n(x_1, t_1; \ldots; x_n, t_n) = \sum_{x_{n+1}} P_{n+1}(x_1, t_1; \ldots; x_n, t_n; x_{n+1}, t_{n+1}).$$

That is, every n-point mass must be the marginal of an $(n+1)$-point mass.

(iv) Normalization: for each t,

$$\sum_x P_1(x, t)\, dx = 1.$$

By (iii) this implies the proper normalization of all joint densities.

These properties extend to continuous state spaces in the usual way; instead of a discrete probability mass, we define a continuous probability density, ρ, and the sums become integrals.

1.2.2 Completely random processes

Stochastic processes may be classified in many ways. A particularly useful classification is by the degree of *memory* the process retains about its past values.

The simplest class are processes with *no memory* of the past: the random variables at distinct times are mutually independent. If $t_1, t_2, \ldots, t_n$ are distinct times, independence is expressed by the factorization of the joint probability

$$P_n(x_1, t_1;\, x_2, t_2;\, \ldots;\, x_n, t_n) = \prod_{i=1}^{n} P_1(x_i, t_i). \tag{6}$$

An example of such a process would be the following: if I rolled a die over and over again and displayed the latest result to you on some sort of monitor, this would be a completely random process. The value that you see on the monitor would tell you nothing about the next value that you will see. If I instead displayed a running sum, or running mean, that would *not* be a completely random process, as the range of possible next values is clearly dependent on the running tally up to the current moment.

1.2.3 Markov Processes

More generally, processes can exhibit correlations across time. The most important class consists of processes with *short memory*, namely Markov processes. Let $t_1 < t_2 < \cdots < t_n$. A process is said to be *Markov* if the conditional distribution of the future given the entire past depends only on the most recent value. Mathematically, we say

$$P\big(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_1, t_1\big) = P\big(x_n, t_n \mid x_{n-1}, t_{n-1}\big), \tag{7}$$

for every choice of ordered times. The quantity $P(x', t' \mid x, t)$ is called the *transition probability* from (x, t) to (x', t'). Our random walker was an example of a Markov process. If I told you the probability mass at the current time, you could use that to compute the next probability mass vector (and all future probabilities), irregardless of what initial condition I used to get to the vector that I gave you.

Using repeated conditioning, the Markov property implies that the full joint probability factorizes into an initial marginal and a product of transition probabilities. For three times,

$$\begin{aligned} P(x_3, t_3; x_2, t_2; x_1, t_1) &= P\big(x_3, t_3 \mid x_2, t_2; x_1, t_1\big)\, P(x_2, t_2; x_1, t_1) \\ &= P\big(x_3, t_3 \mid x_2, t_2\big)\, P\big(x_2, t_2 \mid x_1, t_1\big)\, P(x_1, t_1). \end{aligned} \tag{8}$$

Applying the same logic again and again extends this result to n times:

$$P(x_n, t_n; \ldots; x_1, t_1) = P(x_1, t_1) \prod_{k=2}^{n} P\big(x_k, t_k \mid x_{k-1}, t_{k-1}\big). \tag{9}$$

Thus a Markov process is fully determined (in distribution) by the initial condition, and the transition kernel. We can show explicitly that the trajectories don't matter by *integrating out* the intermediate states. Conceptually, this is saying that if I start at some point, I can calculating the probability of ending up somewhere else at a later time by summing over all the possible ways I could get there. Recall again that we did exactly this for the random walker in the previous section. All we're doing now is putting it in a bit more formal of a notation. For $t_1 < t_2 < t_3$ we start from (8) and sum over x_2:

$$P(x_3, t_3; x_1, t_1) = \sum_{x_2} P(x_3, t_3 \mid x_2, t_2)\, P(x_2, t_2 \mid x_1, t_1)\, P(x_1, t_1).$$

Dividing both sides by $P(x_1, t_1)$ yields the *Chapman–Kolmogorov equation* for transition probabilities:

$$P(x_3, t_3 \mid x_1, t_1) = \sum_{x_2} P(x_3, t_3 \mid x_2, t_2)\, P(x_2, t_2 \mid x_1, t_1). \tag{10}$$

This expresses the fact that the probability to go from (x_1, t_1) to (x_3, t_3) can be obtained by summing (or integrating) over all intermediate states x_2 at time t_2: first transition from x_1 to x_2 and then from x_2 to x_3. The two successive steps are statistically independent conditioned on the intermediate state because of the Markov property.

1.3 Master Equations

Many systems, for obvious reasons, are better described by permitting time to be a continuous variable. We now turn our attention to these. Let's stick discrete state spaces; the evolution equation for such a probability mass function is called a *master equation*, or more humbly, a rate equation.

Let $P_m(t)$ denote the probability to be in state m at time t, and let $T_{mn}(\Delta)$ be the probability to transition from state n to state m in a time interval Δ. The discrete-time evolution reads

$$P_m(t + \Delta) \;=\; \sum_n P_n(t)\, T_{mn}(\Delta) \tag{11}$$

This equation is just saying that the probability of being in state m at time $t + \Delta$ is just sum over all states of the probability of transitioning to state m from state n in time Δ times the probability of being in state n at time t. Now, obviously if $\Delta = 0$, $T_{mn} = \delta_{mn}$, where δ here is the Kronecker delta. This is because if no time has passed, the state must remain unchanged. Over any non-zero time interval, however, the transition probability may also be nonzero. In this way, $T_{mn}(\Delta)$ can be understood as a function of Δ. In particular, we see it must be a monotonic function of Δ, which at $\Delta = 0$, itself equals zero. Let the time interval, Δ be small, such that we can approximate T_{mn} to first order:

$$T_{mn}(\Delta) \;=\; \begin{cases} \frac{dT_{mn}}{d\Delta}\,\Delta + \mathcal{O}(\Delta^2), & m \neq n, \\ 1 - \sum_{l\neq m} \frac{dT_{lm}}{d\Delta}\,\Delta + \mathcal{O}(\Delta^2), & m = n, \end{cases} \tag{12}$$

The $m = n$ case comes from the fact that the probability of staying in state m is the same as the one minus probability that we transition out of state m. We will denote the derivatives as W_{mn}; they are also called the *transition rates*. The transition rates satisfy

$$W_{mn} \geq 0 \;(m \neq n), \qquad W_{mm} \;=\; -\sum_{l\neq m} W_{lm} \tag{13}$$

The first condition is a result of the fact that T_{mn} is a monotonic function of Δ, and the second condition is a result of the conservation of probability. Now that we have a formalism that treats time continuously, we can go about coming up with an equation of motion for the probability mass.

Insert (12) into (11):

$$\begin{aligned} P_m(t+\Delta) &= \sum_{n\neq m} P_n(t)\left[W_{mn}\Delta + \mathcal{O}(\Delta^2)\right] + P_m(t)\Big[1 - \sum_{l\neq m} W_{lm}\Delta + \mathcal{O}(\Delta^2)\Big] \\ &= P_m(t) \;+\; \Delta\left[\sum_{m\neq n} P_n(t)W_{mn} - P_m(t)\sum_{l\neq m} W_{lm}\right] \;+\; \mathcal{O}(\Delta^2) \end{aligned} \tag{14}$$

Subtract $P_m(t)$ from both sides and divide by Δ:

$$\frac{P_m(t+\Delta) - P_m(t)}{\Delta} = \sum_{n\neq m} P_n(t)W_{mn} \;-\; P_m(t)\sum_{l\neq m} W_{lm} \;+\; \mathcal{O}(\Delta) \tag{15}$$

Taking the limit $\Delta \to 0$ gives the master equation:

$$\boxed{\frac{dP_m(t)}{dt} = \sum_{n\neq m}\Big(P_n(t)W_{mn} - P_m(t)W_{nm}\Big) = \sum_{n}\Big(P_n(t)W_{mn} - P_m(t)W_{nm}\Big)} \tag{16}$$

The second step comes from the fact that when $n = m$, that term in the summation is 0, and so we can include it for free. Now, the problem just reduces to computing the transition rates, W_{mn}. This will vary from system to system, and usually depends on the underlying laws of physics. We will examine a few instructive examples. Note, that this is a system of n coupled differential equations, and so solving the master equation is only really something we can do in a few cases.

1.3.1 Example: Random walk on the infinite 1D lattice

Imagine an infinite, 1D lattice, as depicted in Fig. 2. The walker starts at some position, and moves either left or right, with some rate, k. Now, for site n, there are then two ways for probability to flow in to the site (from the left and the right sites), and two ways for it to flow out (also to the left and the right). Then, we can read off the master equation for the n-th site as

$$\frac{dP_n}{dt} = k(P_{n-1} + P_{n+1} - 2P_n) \tag{17}$$

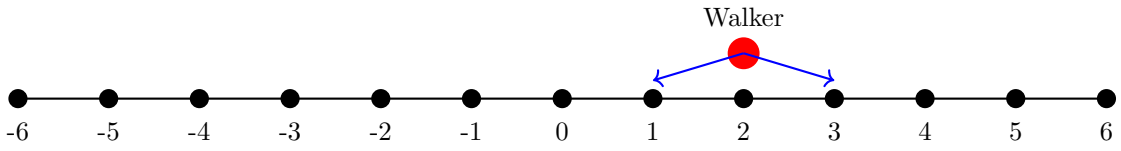


Figure 2: A random walker on an infinite 1D lattice.

We can use the master equation to compute various quantities of physical interest. The first question we might ask is what is the average position of the walker? Define the average site as

$$\bar{n}(t) = \sum_{n=-\infty}^{\infty} n\, P_n(t).$$

Using the master equation, we see that the time derivative of $\bar{n}$ is

$$\frac{d}{dt}\bar{n}(t) = \sum_n n\, \dot{P}_n(t) = \sum_n n\, k\Big(P_{n+1} + P_{n-1} - 2P_n\Big).$$

We can simplify the expression by shifting the indices in the sums, yielding

$$\sum_n nP_{n+1} = \sum_m (m-1)P_m = \sum_m mP_m - \sum_m P_m = \bar{n} - 1,$$

$$\sum_n nP_{n-1} = \sum_m (m+1)P_m = \sum_m mP_m + \sum_m P_m = \bar{n} + 1,$$

and

$$\sum_n 2nP_n = 2\bar{n}.$$

Substituting this back into the master equation yields:

$$\frac{d}{dt}\bar{n}(t) = k\big((\bar{n}-1) + (\bar{n}+1) - 2\bar{n}\big) = 0.$$

Thus, we can see that the average site does not change over time. Wherever the particle starts is its average location. This ought to have been expected, since we know the transition rate to the left and the right are equal. We can come to some more interesting conclusions by computing other quantities. For example, we can compute the average deviation of the particle from its mean position; that is, the variance of its position.

$$\overline{n^2}(t) = \sum_n n^2\, P_n(t).$$

Once again, using the master equation, we can compute its time derivative

$$\frac{d}{dt}\overline{n^2}(t) = \sum_n n^2\, \dot{P}_n(t) = \sum_n n^2\, k\Big(P_{n+1} + P_{n-1} - 2P_n\Big).$$

Shift indices again:

$$\sum_n n^2 P_{n+1} = \sum_m (m-1)^2 P_m = \sum_m (m^2 - 2m + 1)P_m = \overline{n^2} - 2\bar{n} + 1,$$

$$\sum_n n^2 P_{n-1} = \sum_m (m+1)^2 P_m = \sum_m (m^2 + 2m + 1) P_m = \overline{n^2} + 2\bar{n} + 1.$$

Then:

$$\frac{d}{dt}\overline{n^2}(t) = k\Big((\overline{n^2} - 2\bar{n} + 1) + (\overline{n^2} + 2\bar{n} + 1) - 2\overline{n^2}\Big) = 2k$$

So, the mean square displacement grows linearly with time, or put another way, the average deviation from the mean grows with the square root of time.

$$\overline{n^2}(t) = 2kt$$

This demonstrates that the walker spreads out over time, even though the mean position remains the same. This is a characteristic property of diffusive processes, which we will revisit later.

2 Lecture 2 - Transport, Intro to Brownian Motion

We'll pick up where we left off last week, with the master equation for a diffusion process. We kind of rushed through the relation to the continuum limit, so I want to work through that carefully, as it bears some relevance to our discussion of Brownian motion.

Let's revisit the master equation for our random walker. In order to relate this to the continuum, we need to impose some sort of length scale on the problem. Let's call the lattice spacing (the distance between the sites) Δx. Then,

We have then that

$$\frac{dP(x,t)}{dt} = k \cdot [P(x - \Delta x, t) + P(x + \Delta x, t) - 2P(x,t)]$$

Now, in the continuum limit, we can Taylor expand the probability densities around x to get

$$P(x \pm \Delta x, t) = P(x,t) \pm \Delta x \cdot \frac{\partial P(x,t)}{\partial x} + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 P(x,t)}{\partial x^2} + ...$$

We have kept terms to second order for reasons that will soon become apparent. If we plug that into the right hand side of the master equation, we get that the zeroeth and first order terms cancel out, leaving

$$\frac{dP}{dt} = k(\Delta x)^2 \cdot \frac{\partial^2 P}{\partial x^2} = D\frac{\partial^2 P}{\partial x^2} \tag{18}$$

By convention, the macroscopic diffusion constant is given by this quantity: $D = \lim_{\Delta x \to 0} k(\Delta x)^2$. Note that in order for this limit to make any sense, the diffusion constant, k, must grow inversely with the lattice spacing at least quadratically fast.

This is *Fick's second law*, also know as the diffusion equation. It is the continuum equation that describes the diffusion process in a medium. You will note it also looks a lot like a heat equation. This is not by accident. We will explore this connection in more detail later in this course. As a last note about the random walker, I will point out that the diffusion equation with a delta source (all the probability initially starts in one spot, x_0) has solution

$$P_n(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x-x_0)^2}{4Dt}} \tag{19}$$

It is a fun exercise to attempt to show that our discrete random walker obeys the same probability distribution in the long-time limit.

2.1 Transport Coefficients

Last week, we spent our time on two major topics: the mathematics of stochastic processes, and the specific example of how diffusive behavior emerges from random walks. The next question we will try and answer is how these ideas fit with models of real physical processes. That is, we will try and understand why this model of a random walk is a good physical model for things, and what sorts of quantities exhibit this kind of diffusive behavior.

It turns out, for reasons that we will explore later in the course, when we discuss kinetic theory, that one of the most important mechanisms by which diffusion and relaxation processes occur is through collisions. One way to understand this is by observing that particles in free motion have conserved energy and momentum, until they collide with another. Collisions facilitate momentum and energy transfer, and so they can drive a system towards some kind of equilibrium. Again, I promise this will all become more precise when we discuss kinetic theory, but bear with me for now.

Because of this, it seems prescient to try and understand a little bit about collisions and how they happen. To do this, we will return to a canonical example in thermodynamics: a fluid of N particles confined to a volume V.

A basic question we might ask is: How long, on average, does a particle travel before it collides with another? This is called the *mean free path* of a particle, and it is a basic example of a collision related quantity.

Say that each fluid particle has diameter d. Obviously, the particle has cross sectional area $\frac{\pi d^2}{4}$. However, for the purposes of determining the mean free path, a more useful quantity is the "effective cross section", $\sigma = \pi d^2$. Consider a single particle of fluid moving amongst the rest. As the particle moves, if the center of any other particle is within d of its center, then they collide. Therefore, we call the πd^2 the effective cross section.

If the particle moves a distance l, then the volume it traces out is then $\pi d^2 l$. This volume will, on average, contain a particle when

$$\pi d^2 l = \frac{V}{N} = \frac{1}{n}$$

where we have introduced n for the gas density. Solving for l, we get that

$$l = \frac{1}{n\pi d^2}$$

In general, in this course, we will work in the limit where $l \gg d$. That is, the mean free path is very large compared to the actual size of the particles. Now that we know the distance traveled between collisions, the next question we can ask is how frequent are collisions? That is, how long does a particle have, on average, before it collides with another? This quantity is called the relaxation time, $\tau = \frac{l}{\bar{v}_{rel}}$.

The velocity, $\bar{v}_{rel}$ is the average relative speed of the colliding particles.

$$\begin{aligned}
\bar{v}^2 &= \langle (v - v')^2 \rangle \\
&= \int d^3v \int d^3v' (\vec{v} - \vec{v'})^2 f(v) f(v') \\
&= \langle v^2 \rangle + \langle v'^2 \rangle - 2\langle \vec{v} \cdot \vec{v'} \rangle \\
&= 2\langle v^2 \rangle
\end{aligned}$$

In the second line, we are just explicitly writing out the expectation against the Boltzmann distribution for velocities, $f(\vec{v})$. Note, however, that we are multiplying the two distributions, instead of integrating over the joint distribution $f(\vec{v}, \vec{v'})$. This expresses the assumption that the two velocities are independent. Makes intuitive sense, right? This is a wicked little assumption though, which we will spend a great deal of time discussing when we get

to kinetic theory. In the third line to the last line, we use the fact that the two velocities are uncorrelated, and so, on average, they will have no preferential mutual direction. This makes the dot product term vanish, and it lets us argue that $\langle v^2 \rangle = \langle v'^2 \rangle$.

Finally, we can say that

$$\langle v^2 \rangle = \frac{3kT}{m}$$

You can compute this explicitly by doing the integral against the Boltzmann distribution, or you can get to it by appealing to equipartition. All of that put together gives us that

$$\tau = \frac{1}{n\pi d^2}\sqrt{\frac{m}{6kT}}$$

This is the time in between collisions, and as I mentioned, we will generally assume that this is much longer than the time it takes for the actual collision.

2.1.1 Viscosity

We now turn our attention to the problem of "transport". That is, how do macroscopic quantities like temperature, density, or pressure get moved around a fluid. Each of these will be associated with a conserved quantity, and with a so-called transport coefficient.

The first transport coefficient of interest is called the *viscosity* of a fluid. Loosely speaking, this is the quantity which describes a fluid's internal resistance to flow. That is, it describes how much a fluid will resist being set in motion.

To see why this might arise, consider the following setup: two parallel sheets, with some fluid between them. Let's take the plates to be extended in the x direction, with a separation of Δz. If we set one sheet to be in motion with velocity v in the x-direction relative to the other, then friction will cause the fluid near the mobile plate to also move at, or near, the velocity v. Meanwhile, the fluid near the stationary plate will be stationary. This indicates the development of a velocity gradient between the plates. Experimentally, it has been determined that the force per unit area of the plate,

$$\frac{F}{A} = \eta\frac{dv_x}{dz} \approx \eta\frac{v}{\Delta z}$$

The coefficient η is called the viscosity. We can derive this from first principles in the following way: The particles that collide with the moving slab pick up some momentum in the x-direction. They then collide with molecules lower down in the stream, and so on, imparting momentum all the way down. This induces some velocity gradient down the z-direction.

Consider a slice of the fluid at constant z. To figure out the force on this fluid, we need two things: The amount of momentum imparted to it by each collision (from above and below), and the the number of collisions in either direction.

The number of particles going through the slab depends on the flow rate through the slab. That is, it depends on the velocity of particles going through the slab. We know the velocity distribution is given by the Boltzmann distribution, which gives that the net number of particles per unit time per unit area hitting the slab is

$$nv_z P(\vec{v}) d^3 v$$

Each of these imparts some momentum to the slab. If the particle traveled Δz before its last collision, then

$$\Delta p = m\underbrace{(v_x(z+\Delta z) - v_x(z))}_{\text{relative velocity}} \approx m\frac{dv_x}{dz}\Delta z = -m\frac{dv_x}{dz}l\cos\theta$$

where in the last step, θ is the angle between the velocity and the z-axes, and we noticed that the distance traveled, on average, will be the mean free path. For those of you following along in Tong, note that he defines θ differently, leading to a few signs being flipped. Now, recall that we wanted to compute the force per unit area:

$$\frac{F}{A} = -\frac{1}{A}\frac{\Delta p}{\Delta t}$$

The minus sign is due to the fact that $\Delta p/\Delta t$ is the force of the fluid pushing back, but we want the force needed to keep the fluid moving. The change in momentum per unit time per unit area is just the product of the two above expressions:

$$\frac{F}{A} = -n \int d^3v \, \Delta p \, v_z \, P(\vec{v})$$

which, plugging in our expression for Δp and the Maxwell-Boltzmann distribution for $P(\vec{v})$ gives

$$\frac{F}{A} = mn\frac{du_x}{dz} \int d^3v \, v_z l \cos\theta \cdot \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-\frac{m|\vec{v}|^2}{2kT}\right)$$

Note here that we are using the Boltzmann distribution for the velocity *vector*, not the Maxwell speed distribution. All that remains is to do the integral, which happens to be easiest in spherical coordinates:

$$\begin{aligned}
\frac{F}{A} &= mn\frac{du_x}{dz} \int d^3v \, v_z l \cos\theta \cdot \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-\frac{m|\vec{v}|^2}{2kT}\right) \\
&= mn\frac{du_x}{dz} \int dv \, v^2 \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi \, (v\cos\theta) \cdot l\cos\theta \cdot \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-\frac{m|\vec{v}|^2}{2kT}\right) \\
&= 2\pi mn\frac{du_x}{dz} \int dv \, v^2 \int_0^{\pi} \sin\theta d\theta \, (v\cos\theta) \cdot l\cos\theta \cdot \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-\frac{m|\vec{v}|^2}{2kT}\right) \\
&= \frac{4\pi}{3} lmn\frac{du_x}{dz} \int dv \, v^3 \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-\frac{mv^2}{2kT}\right) \\
&= \frac{1}{3} lmn\langle v\rangle \frac{du_x}{dz}
\end{aligned}$$

where in going from the penultimate to the last line, we recognized that the dv integral, up to some constant factors, is simply the integral of the speed times the Maxwell-Boltzmann distribution, so it is the average speed of a particle in a Maxwellian gas. Thus, comparing with our earlier result that

$$\frac{F}{A} = \eta\frac{dv_x}{dz}$$

we get that

$$\eta = \frac{1}{3} lmn\langle v\rangle$$

One very important and somewhat counterintuitive consequence of this formula is that the viscosity does not depend on the density of the fluid! Substituting in our earlier expression for the mean free path, we see that

$$\eta = \frac{m}{3\pi d^2}\langle v\rangle$$

The intuition is that even though a denser fluid has more particles, and thus a higher rate of collisions, the collisions impart much less momentum, since the particles have traveled less distance.

2.1.2 Thermal Conductivity

As a next example of a transport coefficient, we will derive a similar expression for the thermal conductivity of a fluid. The argument is very similar. Empirically, one finds that heat flow (energy per unit time per unit area) is proportional to the gradient of temperature:

$$\vec{q} = -\kappa \nabla T$$

We follow the same setup: place a fluid between two slabs of metal, and set up a temperature gradient between the top and bottom. The number of particles with a particular velocity passing through a slice at a particular height, z, is given by the same expression as before:

$$n v_z P(\vec{v}) d^3 v$$

And, using equipartition, the energy of such a particle is given by

$$E(z) = \frac{3}{2} k T(z)$$

Now, we need to come up with a simple model for how energy transfer occurs when the particle reaches the slab. For the purposes of simplicity, we will assume that particles that are hotter than the particles in the slab deposit all their excess energy, and particles that are colder absorb all the excess energy. Then,

$$\Delta E = E(z + \Delta z) - E(z) = \frac{3}{2} k \frac{dT}{dz} \Delta z = \frac{3}{2} k \frac{dT}{dz} \cdot l \cos\theta$$

Repeating the same steps as the viscosity derivation, we get that

$$|\vec{q}| = n \int d^3 v \Delta E v_z P(v) = -\frac{1}{2} k n l \langle v \rangle \frac{dT}{dz} = \frac{1}{3} c_V l \langle v \rangle \frac{dT}{dz}$$

The proportionality constant is the thermal conductivity, κ, written in terms of the heat capacity of a monatomic ideal gas, $c_V = \frac{3}{2} k$.

2.1.3 Diffusion and Conservation Laws

What are we to make of the similarities in these derivations? We see that particles undergo diffusion, and that momentum and energy are smoothed out via transport. This turns out to be the case simply because these are all locally conserved quantities. Locally conserved quantities obey continuity equations. That is, we can write down an equation that expresses the idea that for any locally conserved quantity, its spatial density can only change due to a flux. For example, for energy:

$$\frac{\partial E}{\partial t} + \nabla \cdot \vec{q} = 0$$

Once again, we know due to equipartition that $E(\vec{x}) = \frac{3}{2} k T(\vec{x})$. Plugging that in to the continuity equation, we get

$$\frac{\partial E}{\partial t} + \nabla \cdot \vec{q} = 0$$
$$\longrightarrow \frac{3}{2} k \frac{\partial T(\vec{x})}{\partial t} = -\nabla \cdot \vec{q} = \kappa \nabla^2 T$$
$$\longrightarrow \frac{\partial T}{\partial t} = \frac{\kappa}{c_V} \nabla^2 T$$

So, we see that temperature, like particle number, obeys a diffusion equation. A similar story holds for viscosity and pressure. The big idea is that in general, conserved quantities diffuse; fluctuations smooth out over time. Of course, we know this is not the whole picture, since the full equations of hydrodynamics contain a much richer spectrum of behavior than just diffusion. We'll get to that when we discuss kinetic theory, but for now, we are just drawing the conceptual link.

3 Brownian Motion and Langevin Dynamics

Take a particle suspended in a fluid at some non-zero temperature. If you look at it under a microscope, you see that it jitters and travels around over time. This phenomenon was first described by the biologist Robert Brown in the 1830s, but it took until the 1900s for a satisfactory understanding of it to be developed by Paul Langevin, Marian Smoluchowski, and most famously by Albert Einstein as part of his *annus mirabilis* papers. It holds a special place in the history of physics as one of the phenomena that provided incontrovertible evidence that atoms exist. Since then, Brownian motion has become a cornerstone tool in statistical physics, financial modeling, and machine learning, among other fields.

3.1 Langevin Dynamics

We will start with one-dimensional motion. Consider the motion of a particle in some fluid. Its equation of motion is given by Newton's Second Law

$$m \frac{dv}{dt} = F_{tot} \tag{20}$$

Neglecting the external force, we know that a particle in a fluid experiences a drag proportional to its velocity:

$$m \frac{dv}{dt} = -\gamma v \longrightarrow v(t) = v(0) e^{-\gamma t/m} \tag{21}$$

Of course, we know this cannot be quite correct, since thermodynamics tells us that at equilibrium, particles have energy dictated by equipartition:

$$\langle E \rangle = \frac{kT}{2} \longrightarrow \langle v^2 \rangle = \frac{kT}{m} \tag{22}$$

It turns out, empirically, that the appropriate modification to make to the equation of motion is to add a *random force*, δF.

$$m \frac{dv}{dt} = -\gamma v + \delta F(t) \tag{23}$$

In this way, we partition the force into a deterministic drag term, and a stochastic fluctuation. This is the random force due to the interaction with the environment. If it's hard for you to think about a force being random, it may

help to have the following picture in mind. Instead of thinking about the continuous time differential equation, think about the first order Euler approximation to it. The idea is just that at every time step, instead of receiving a kick from the potential, we first kick from the potential and then add a random number to the derivative. (Actually, the order in which we do things will make a difference, but that's a problem for later).

We can determine some of the characteristics that our random force should have based purely from physical arguments.

1. The random force has mean 0. $\langle \delta F(t) \rangle = 0$.
2. The random force at different times should not be correlated. That is, $\langle \delta F(t) \delta F(t') \rangle = 2\zeta \delta(t - t')$ for some constant ζ. The constant sets the variance of the force at a particular time.
3. The distribution of the random force is Gaussian.

Let's unpack these. The first assumption is fairly straightforward. If the random force had non-zero mean, it would lead to drift over time. This is not very characteristic of fluctuations. Collisions from a random heat bath have no preferential direction. Anyways, if there were a preferred direction, we could incorporate that into the potential. The second property is a statement about time-scales. We are saying that the thermal fluctuations happen "quickly." That is, so many thermal fluctuations happen within an infinitesimal time scale that any correlations are quickly lost. The last statement is an application of the central limit theorem; the distribution of the sum of a large number of i.i.d. random variables is Gaussian.

Now, how do we go about actually solving an equation like Eq. 23? It is a first order, inhomogenous ODE (*stochastic differential equation*, actually). We can rewrite it in the form

$$\frac{dv}{dt} = -\frac{\gamma}{m} v(t) + \frac{1}{m} \delta F(t) \tag{24}$$

Making the substitution

$$v(t) = e^{-\frac{\gamma}{m} t} s(t) \longrightarrow \frac{dv}{dt} = -\frac{\gamma}{m} e^{-\frac{\gamma}{m} t} s(t) + e^{-\frac{\gamma}{m} t} \frac{ds}{dt} \tag{25}$$

Equating this to Eq. 24, we obtain

$$\frac{ds}{dt} = \frac{1}{m} e^{\frac{\gamma}{m} t} \delta F(t) \rightarrow s(t) = s(0) + \frac{1}{m} \int_0^t dt' e^{\frac{\gamma}{m} t'} \delta F(t') \tag{26}$$

Which in terms of $v(t)$ yields the general solution,

$$v(t) = v_0 e^{-\frac{\gamma}{m} t} + \int_0^t dt' e^{-\frac{\gamma}{m}(t - t')} \frac{\delta F(t')}{m} \tag{27}$$

3.1.1 Correlation Functions and Fluctuation-Dissipation

This equation is still somewhat hard to interpret, since it is not obvious what the latter integral means, seeing as it is a function of a stochastic variable. To gain some more insight, let's compute the *mean* velocity as a function of time.

$$\langle v \rangle = \left\langle v_0 e^{-\frac{\gamma}{m} t} \right\rangle + \left\langle \int_0^t dt' e^{-\frac{\gamma}{m}(t - t')} \frac{\delta F(t')}{m} \right\rangle \tag{28}$$

Now the first term is deterministic. It contains no random variables, so its average is simply itself. To evaluate the second term, we can use linearity of expectation to move the expected value computation inside of the integral.

$$\langle v \rangle = v_0 e^{-\frac{\gamma}{m}t} + \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} \left\langle \frac{\delta F(t')}{m} \right\rangle \tag{29}$$

Since the mean of the fluctuation vanishes, the integral as a whole vanishes, and we get that

$$\langle v \rangle = v_0 e^{-\frac{\gamma}{m}t} \tag{30}$$

which is just the deterministic result from classical mechanics. Evidently, the stochastic term does not affect the average behavior of the system. What about the mean-squared velocity? This is where we will pick up next week.

4 Lecture 3 - Brownian Motion

We left off last week discussing Brownian motion. The setup was that we had some particle, called a *Brownian particle*, suspended in some sort of fluid. Due to collisions with the various fluid elements, it undergoes some kind of random motion. We describe this with the *Langevin equation*.

$$m\frac{dv}{dt} = -\gamma v + \delta F(t)$$

The first term is the usual drag that we know and love from classical mechanics, but the second term is a new creature. It is a *random* force, which causes the jittery, stochastic behavior we want. We reasoned about the characteristics of such a random force, namely that it has zero mean, $\langle \delta F \rangle = 0$ and nonzero variance, $\langle \delta F(t) \delta F(t') \rangle = 2\zeta\delta(t-t')$. The delta function expresses the idea that the correlations decay quickly in time. Obviously we know that this isn't truly the case! As an example, say the particle receives a huge kick in the positive x direction. Then, since in its frame, all the fluid particles are moving in the negative x direction now, we would obviously expect the next kick has some negative bias. In this sense, the kicks should be correlated in time, but we are assuming that the collisions happen frequently enough that such correlations decay quickly. We wrote down a formal solution to the Langevin equation:

$$v(t) = v_0 e^{-\frac{\gamma}{m}t} + \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} \frac{\delta F(t')}{m}$$

which is not terribly insightful, since we don't really know how to interpret the latter term. We got around this by computing so-called *correlation functions*. We asked: what is the average velocity as a function of time, where the average is taken over the ensemble. As expected, this yielded the standard result from classical mechanics:

$$\langle v(t) \rangle = v_0 e^{-\frac{\gamma}{m}t}$$

The plan for today is to first start by computing the higher order correlation functions, and seeing if we get anything new. Then, we will see how the correlation functions are related to transport coefficients. Finally we will break out of correlation function jail and come up with an expression for the true time evolution of the probability of finding our Brownian particle in various locations. Along the way, we'll do some super cool physics and commit some minor math crimes.

4.1 Correlation Functions, Continued.

$$\begin{aligned}\langle v^2\rangle &= \left\langle \left[v_0 e^{-\frac{\gamma}{m}t} + \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')}\frac{\delta F(t')}{m}\right]^2 \right\rangle \\ &= \langle v_0^2 e^{-\frac{2\gamma}{m}t}\rangle + \left\langle 2v_0 e^{-\frac{\gamma}{m}t}\int_0^t dt' e^{-\frac{\gamma}{m}(t-t')}\frac{\delta F(t')}{m}\right\rangle + \left\langle \int_0^t dt'' \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} e^{-\frac{\gamma}{m}(t-t'')}\frac{\delta F(t')\cdot\delta F(t'')}{m^2}\right\rangle\end{aligned}$$

Now, by the same arguments as before, the first term is deterministic. The second term vanishes due to the vanishing of the mean of the fluctuations. The third term, however, is more interesting. As before, we can use linearity of expectation to bring the expectation into the integral:

$$\begin{aligned}\left\langle \int_0^t dt'' \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} e^{-\frac{\gamma}{m}(t-t'')}\frac{\delta F(t')\cdot\delta F(t'')}{m^2}\right\rangle &= \int_0^t dt'' \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} e^{-\frac{\gamma}{m}(t-t'')}\left\langle \frac{\delta F(t')\cdot\delta F(t'')}{m^2}\right\rangle \\ &= \frac{2\zeta}{m^2}\int_0^t dt'' \int_0^t dt' e^{-\frac{\gamma}{m}(t-t')} e^{-\frac{\gamma}{m}(t-t'')}\delta(t'-t'') \\ &= \frac{2\zeta}{m^2}\int_0^t dt' e^{-\frac{2\gamma}{m}(t-t')} \\ &= \frac{\zeta}{\gamma m}\left[1 - e^{-\frac{2\gamma t}{m}}\right]\end{aligned}$$

Which yields the final result

$$\langle v^2\rangle = v_0^2 e^{-\frac{2\gamma}{m}t} + \frac{\zeta}{\gamma m}\left[1 - e^{-\frac{2\gamma t}{m}}\right] \tag{31}$$

In the long time limit, the exponentials drop out, yielding

$$\langle v^2\rangle = \frac{\zeta}{\gamma m} \tag{32}$$

From equipartition, we know that once the system equilibrates, the mean squared velocity should be kT/m, so

$$\frac{kT}{m} = \frac{\zeta}{\gamma m} \rightarrow \boxed{\zeta = \gamma kT} \tag{33}$$

This formula is our first example of the fluctuation-dissipation theorem. It relates the strength of the fluctuating force, ζ, to the strength of the dissipative force, γ. A balance between these two is required to achieve an equilibrium state. If the thermal fluctuations are too large, the velocity and position will grow larger and larger over time. Conversely, if the damping is too strong, the system will just decay until nothing is moving. It is the balance between these quantities that leads to the interesting dynamics we observe.

We could now proceed to compute other correlation functions of interest, like the mean squared *displacement*. That will be on the problem set. The result is

$$\langle(\Delta x)^2\rangle = \frac{2kT}{\gamma}\left[t - \frac{m}{\gamma} + \frac{m}{\gamma}e^{-\frac{\gamma t}{m}}\right] \tag{34}$$

As $t \rightarrow \infty$, the first term dominates, so

$$\langle(\Delta x)^2\rangle = \frac{2kT}{\gamma}t$$

Equating this with the long term behavior of our random walker, we then find that the value for D, the diffusion constant is

$$\frac{2kT}{\gamma}t = 2Dt \rightarrow D = \frac{kT}{\gamma}$$

This is the *Einstein relation* for the diffusion coefficient. It tells us that the diffusion coefficient is proportional to the temperature (hot systems diffuse faster), and inversely proportional to the damping (damped systems diffuse slower).

4.2 Elements of Linear Response

Consider a system at thermal equilibrium. We know that it obeys the Boltzmann distribution, so for any dynamic observable, its average is given by

$$\langle A \rangle = \frac{1}{Z}\int A \cdot \exp(-\beta E) dE$$

However, we know that there will be fluctuations around this mean. What is the typical size of the fluctuation?

Now, consider the following situation. A system is in thermal equilibrium with its surroundings, and then a weak external perturbation is applied. How will the system respond? We expect that given sufficient time, the system will equilibrate to this new Hamiltonian, but there will be some transient response. What does this transient response of the observable A look like?

On its face, these may seem like two entirely separate questions, but it turns out that these two are intimately related. One way to understand this is by noting that there is no way for the system to know that it has been whether it has been kicked away from equilibrium by a perturbation to the field, or by a random thermal fluctuation. We'll start by exploring this relationship in the special case of the famous *Green-Kubo relation* for the diffusion coefficient. Later in the course, we will explore the general framework of linear response theory in more detail.

The above results show that

$$\lim_{t\to\infty} \langle (\Delta x)^2 \rangle = 2Dt$$

It turns out that the diffusion coefficient can be rewritten in terms of a set of time integrals. We can start with the definition of the mean squared displacement:

$$\langle (\Delta x)^2 \rangle = \langle (x(t) - x(0))^2 \rangle$$

We can use the definition of velocity to write the term in parentheses as

$$x(t) - x(0) = \int_0^t dt' v(t')$$

So, plugging that in, we get

$$\langle (\Delta x)^2 \rangle = \int_0^t dt' \int_0^t dt'' \langle v(t') v(t'') \rangle$$

When the system is at equilibrium, we can invoke time-translation symmetry to argue that the velocity correlation function should only depend on the time difference, not the specific times. So, we say that

$$\langle (\Delta x)^2 \rangle = \int_0^t dt' \int_0^t dt'' \langle v(t' - t'') v(0) \rangle$$

Then, we can define $\tau = t' - t''; \alpha(\tau) = \langle v(\tau)v(0)\rangle$. We can change variables, using $dt'' = -d\tau$, and rewrite the above integral as

$$\langle(\Delta x)^2\rangle = \int_0^t dt' \int_{t-t'}^{t'} d\tau \alpha(\tau)$$

The idea of doing this is to make it such that the arguments to the autocorrelation function do not depend on one of the time variables that is being integrated over. We want to switch the order of integration so that the dt' integral is evaluated first, since the autocorrelation function depends on τ. The result is

$$\langle(\Delta x)^2\rangle = \int_{-t}^{0} d\tau\, \alpha(\tau) \int_0^{t+\tau} dt' + \int_0^t d\tau\, \alpha(\tau) \int_\tau^t dt'$$

We can then do the dt' integrals easily:

$$\langle(\Delta x)^2\rangle = \int_{-t}^{0} d\tau\, (t+\tau)\alpha(\tau) + \int_0^t d\tau\, (t-\tau)\alpha(\tau)$$

We can change variables in the first integral to $\tau' = -\tau$

$$\begin{aligned}
\langle(\Delta x)^2\rangle &= \int_{-t}^{0} d\tau\, (t+\tau)\alpha(\tau) + \int_0^t d\tau\, (t-\tau)\alpha(\tau) \\
&= -\int_t^0 d\tau'\, (t-\tau')\alpha(\tau') + \int_0^t d\tau\, (t-\tau)\alpha(\tau) \\
&= \int_0^t d\tau'\, (t-\tau')\alpha(\tau') + \int_0^t d\tau\, (t-\tau)\alpha(\tau) \\
&= 2\int_0^t d\tau\, (t-\tau)\alpha(\tau)
\end{aligned}$$

where in the last line we use that $\alpha(\tau) = \alpha(-\tau) = \alpha(\tau')$ due to time reversal symmetry at equilibrium.

Finally, putting this together with our original equation for D, we get

$$\begin{aligned}
D &= \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, (t-\tau)\alpha(\tau) \\
&= \lim_{t\to\infty} \left[\frac{1}{t}\int_0^t d\tau\, t\alpha(\tau) - \frac{1}{t}\int_0^t d\tau\, \tau\alpha(\tau)\right] \\
&= \int_0^\infty d\tau\, \alpha(\tau) - \lim_{t\to\infty} \frac{1}{t}\int_0^t d\tau\, \tau\alpha(\tau)
\end{aligned}$$

As t goes to infinity, we expect the velocity autocorrelation to go to 0. This is saying that due to the large number of collisions, the particle eventually "forgets" its initial velocity. Thus, the only contribution to the second integral happens at small τ. This means the integral becomes something finite, which means as we take t to infinity, the second term vanishes. The result is the *Green-Kubo* formula for the diffusion coefficient:

$$\boxed{D = \int_0^\infty d\tau\, \alpha(\tau)}$$

Comments:

- There are systems for which the autocorrelation does not decay quickly enough to make the arguments we did. This is called *anomalous diffusion*, and is generally harder to treat analytically. However, this formula holds well for many physical systems.
- Why is this formula so important? Well, as we mentioned, it establishes a link between a decidedly *non-equilibrium* transport quantity, the diffusion coefficient, and a correlation function computed *at equilibrium*.
- Practically speaking, the velocity autocorrelation function is something that can easily be computed numerically (for example, in molecular dynamics simulations). Physicists and chemists use these measured correlation functions to measure the diffusion coefficients of various real materials to understand how, for example, impurities travel through materials.
- Computation in higher dimensions (even 2 and 3) of the diffusion coefficient does not trivially generalize from this computation, though it certainly is possible.

4.3 Fokker-Planck Equations

It is often cumbersome to work with a stochastic differential equation like the Langevin equation, since its solution is a stochastic process. As we discussed previously, stochastic processes have some unusual mathematical properties, which makes it difficult to design rules of calculus for them, which naturally leads to challenges in solving stochastic differential equations (particularly of the non-linear form). In addition, its very tedious to have to compute correlation functions!

Instead, we might attempt to write down an ordinary differential equation which describes how the probability density itself changes over time. Notice that in making this simplification, we are saying that we don't care how the how the particle got to a particular position. We are no longer considering a stochastic process, which gives us a probability distribution over trajectories, but instead just a probability density over positions. We will start by considering a Langevin equation in a very viscous fluid. What this means is that the deterministic part of the force is dominated by the drag term, not the inertial motion, which is equivalent to setting $m = 0$. Then, the Langevin equation becomes

$$\gamma v = \gamma \dot{x} = \delta F(t) - \frac{\partial V}{\partial x} \tag{35}$$

Notice here the addition of an extra potential term, $\frac{\partial V}{\partial x}$. Suppose we drop a particle at some position x_0 at time t_0. If the subsequent evolution is noisy, governed by the Langevin equation, then we cannot predict its precise trajectory. The best we can do is talk about *probabilities*.

We denote by

$$P(x, t|x_0, t_0)$$

the probability that the particle sits at position x at time t, given that it started at x_0 at time t_0. Note that we do *not* care about how the particle got there. If we wanted to determine the probability distribution over all paths, that would be a more complex functional calculation. Instead, we ask the simpler question: what is the probability that the particle is at position x at time t, regardless of the path taken to get there?

Formally, if we fix a given noise realization $\delta F(t)$, the Langevin equation produces a definite trajectory $x_{\delta F}(t)$. For that trajectory, the probability distribution is just a delta function:

$$P(x,t) = \delta\big(x - x_{\delta F}(t)\big).$$

To account for the randomness of the noise, we average over all realizations:

$$P(x,t) = \Big\langle \delta\big(x - x_{\delta F}(t)\big) \Big\rangle.$$

To do this, we return to the Langevin equation in the viscous limit. If the particle starts at location x at time t. We can look at the displacement at some small time Δt later.

$$\Delta x \equiv x(t+\Delta t) - x(t) = \dot{x}(t)\cdot\Delta t = -\frac{1}{\gamma}\frac{\partial V}{\partial x}\cdot\Delta t + \frac{1}{\gamma}\int_t^{t+\Delta t} dt'\,\delta F(t').$$

Note here that the potential is evaluated at the initial time and location, so we are implicitly assuming that the displacement is small enough that this is a good approximation. Conversely, the noise is averaged over the whole time interval. This is the correct mathematical model for this case. As I will say over and over in this class, it is a statement about timescales. We are assuming that collisions happen frequently, and so many of them happen in the time that it takes for the particle to travel enough distance that the potential varies appreciably. However, these assumptions can be rather problematic in the general theory of stochastic processes, since the magnitude of the random force is drawn from a Gaussian, and thus technically is unbounded. An individual kick may be infinitely large, making it challenging to formally define such an integral. For more comments on this, see the appendix on the stochastic calculus.

Now, since $\langle \delta F(t)\rangle = 0$, we get that the average displacement,

$$\langle \Delta x\rangle = -\frac{1}{\gamma}\frac{\partial V}{\partial x}\Delta t$$

We can also compute the variance of the position, $\langle (\Delta x)^2\rangle$.

$$\langle (\Delta x)^2\rangle = \left\langle \frac{1}{\gamma^2}\left(\frac{\partial V}{\partial x}\right)^2 (\Delta t)^2 - \frac{2}{\gamma^2}\left[\int_t^{t+\Delta t} dt'\,\delta F(t')\right]\cdot\frac{\partial V}{\partial x}\cdot\Delta t + \frac{1}{\gamma^2}\left[\int_t^{t+\Delta t} dt'\,\delta F(t')\right]^2 \right\rangle$$

Clearly, we see that the first term is $\mathcal{O}((\Delta t)^2)$. The second term vanishes in the ensemble average, since we can move the ensemble average into the integral and use the fact that $\langle \delta F\rangle = 0$. This leaves the last term

$$\langle (\Delta x)^2\rangle = \frac{1}{\gamma^2}\int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt''\,\langle \delta F(t')\delta F(t'')\rangle = \frac{2\zeta}{\gamma^2}\int_t^{t+\Delta t} dt' = \frac{2\zeta}{\gamma^2}\,\Delta t.$$

Hence, to leading order in Δt,

$$\boxed{\langle \Delta x^2\rangle = \frac{2\zeta}{\gamma^2}\,\Delta t + \mathcal{O}(\Delta t^2)}.$$

We can actually plug in the result from the fluctuation-dissipation relation for ζ to get:

$$\frac{2\zeta}{\gamma^2} = \frac{2\gamma kT}{\gamma^2} = \frac{2kT}{\gamma} = 2D$$

Our strategy now is to construct a probability distribution that reproduces the short-time expectation values

$$\langle \Delta x \rangle = -\frac{1}{\gamma}\frac{\partial V}{\partial x}\,\Delta t, \qquad \langle (\Delta x)^2 \rangle = 2D\,\Delta t,$$

that we derived previously.

We begin by considering the conditional probability

$$P(x, t + \Delta t | x', t),$$

the probability that the particle is at position x at time $t + \Delta t$, given that it was at x' a short moment earlier. From the definition, we can write

$$P(x, t + \Delta t | x', t) = \Big\langle \delta\big(x - (x' + \Delta x)\big) \Big\rangle,$$

where Δx is the random displacement during the short time Δt, so its values are what we take the expectation over.

Next, we will commit a math crime and expand the delta function in powers of Δx:

$$P(x, t + \Delta t | x', t) = \left[1 - \langle \Delta x \rangle \frac{\partial}{\partial x'} + \frac{1}{2}\langle (\Delta x)^2 \rangle \frac{\partial^2}{\partial x'^2} + \dots \right] \delta(x - x'). \tag{36}$$

Doing this is mathematically dubious; it is not apparent that a distribution, such as the delta function, can be expanded in a Taylor series, but you can imagine that we are Taylor expanding some smoothed out version of the delta function, like a sharply peaked Gaussian, and then taking the limit where it goes to zero. Really, what we are doing is okay though, since derivatives of the distribution are like the distribution itself: defined only to be integrated against! We truncated at second order because $\langle \Delta x \rangle = O(\Delta t)$ and $\langle (\Delta x)^2 \rangle = O(\Delta t)$, while all higher derivatives are $O(\Delta t^2)$ and can be neglected.

To proceed, we recall that what we ultimately want is the full probability $P(x, t; x_0, t_0)$ given some initial condition $x(t_0) = x_0$. This probability satisfies the Chapman–Kolmogorov relation, which we discussed briefly in the section on Markov processes (though there, we used the discrete formulation). The idea, if you recall, is that for a Markov process, we can integrate over the probabilities of being in any state at some intermediate time:

$$P(x, t | x_0, t_0) = \int_{-\infty}^{\infty} dx'\, P(x, t; x', t')\, P(x', t'; x_0, t_0), \qquad t_0 < t' < t.$$

In our case, our initial time is t_0, our final time is $t + \Delta t$, and the intermediate time is t. We make these substitutions, (t' with t and t with $t + \Delta t$), and substitute the expansion (36) into this equation. This gives

$$\begin{aligned} P(x, t + \Delta t | x_0, t_0) &= \int_{-\infty}^{\infty} dx'\, P(x, t + \Delta t | x', t)\, P(x', t | x_0, t_0) \\ &= \int_{-\infty}^{\infty} dx' \left[1 - \langle \Delta x \rangle \frac{\partial}{\partial x'} + \frac{1}{2}\langle (\Delta x)^2 \rangle \frac{\partial^2}{\partial x'^2} + \dots \right] \delta(x - x')\, P(x', t | x_0, t_0) \end{aligned}$$

Let's go term by term now: The first term is

$$\int_{-\infty}^{\infty} dx' \delta(x - x')\, P(x', t|x_0, t_0) = P(x, t|x_0, t_0)$$

The second term is

$$\int_{-\infty}^{\infty} dx' \langle \Delta x \rangle \cdot \frac{\partial}{\partial x'} \delta(x - x') \cdot P(x', t|x_0, t_0)$$

We can now integrate by parts, moving the derivative off the delta function and onto $P(x', t; x_0, t_0)$. For those of you who may be rusty on how to do this, we say $\int u\, dv = uv - \int v\, du$, where here $u = \langle \Delta x \rangle \cdot P(x', t|x_0, t_0)$, and $dv = \frac{\partial}{\partial x'} \delta(x - x')$. Then, we get that

$$\begin{aligned}
&\int_{-\infty}^{\infty} dx' \langle \Delta x \rangle \cdot \frac{\partial}{\partial x'} \delta(x - x') \cdot P(x', t|x_0, t_0) \\
&= [P(x', t|x_0, t_0) \delta(x - x') \langle \Delta x \rangle]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} dx' \delta(x - x') \frac{\partial}{\partial x'} P(x', t|x_0, t_0) \langle \Delta x \rangle \\
&= -\frac{\partial}{\partial x} P(x, t|x_0, t_0) \langle \Delta x \rangle
\end{aligned}$$

If we repeat that again for the third term, (again, integrating by parts, though twice this time) we get

$$\begin{aligned}
P(x, t + \Delta t; x_0, t_0) = P(x, t; x_0, t_0) &- \frac{\partial}{\partial x} \Big(\langle \Delta x \rangle P(x, t; x_0, t_0) \Big) \\
&+ \frac{1}{2} \frac{\partial^2}{\partial x \partial x} \Big(\langle (\Delta x)^2 \rangle P(x, t; x_0, t_0) \Big) + \dots
\end{aligned} \tag{37}$$

Almost there! We can now plug in our expressions for the mean and variance of Δx:

$$\langle \Delta x \rangle = -\frac{\partial V}{\partial x}\, \Delta t, \qquad \langle (\Delta x)^2 \rangle = 2D\ \Delta t.$$

Substituting into (37) gives

$$P(x, t + \Delta t|x_0, t_0) = P(x, t|x_0, t_0) + \Delta t \left[\frac{\partial}{\partial x} \left(\frac{\partial V}{\partial x} P(x, t|x_0, t_0) \right) + D \frac{\partial^2}{\partial x^2} P(x, t|x_0, t_0) \right] + \dots \tag{38}$$

On the other hand, we can Taylor expand the probability distribution directly in time, for small Δt:

$$P(x, t + \Delta t|x_0, t_0) = P(x, t|x_0, t_0) + \Delta t\, \frac{\partial}{\partial t} P(x, t|x_0, t_0) + \dots$$

Setting these two equal, and solving for $\frac{\partial P}{\partial t}$ gives the *Fokker–Planck equation*:

$$\boxed{\frac{\partial P}{\partial t} = \frac{1}{\gamma} \frac{\partial}{\partial x} \cdot \Big(P \frac{\partial V}{\partial x} \Big) + D \frac{\partial^2}{\partial x^2} P. \quad \ddot\smile}$$

This equation is also known as the *Smoluchowski equation*, or in probability theory as the *Kolmogorov forward equation* (which yes, implies the existence of the related *Kolmogorov's backwards equation*). It gives the time evolution of a probability density under inviscid Langevin evolution. Now, we made certain assumptions about (for example), viscosity, in this derivation. It happens to be the case that there is a corresponding Fokker-Planck equation for most every (Markovian) Langevin equation. I have a derivation for the general form of the Fokker-Planck equation corresponding to a general class of Markovian SDEs on my website. You may find this interesting.

5 Lecture 4 - Consequences of the Fokker-Planck Equation

Last week, we concluded with a derivation of the Fokker-Planck equation for the inviscid Langevin equation.

$$\boxed{\frac{\partial P}{\partial t} = \frac{1}{\gamma}\frac{\partial}{\partial x}\cdot\left(P\frac{\partial V}{\partial x}\right) + D\frac{\partial^2}{\partial x^2}P. \quad \text{☺}}$$

The derivation was rather complex and technical, but the result is pretty simple! It gives us a differential equation that tells us how the probability of finding our Brownian particle in various locations changes over time. Now, recall that our motivation for even deriving an equation like this was that the Langevin equation was quite unwieldy. Is this Fokker-Planck equation any easier to work with?

I would wager that for most of you, it is, at the very least, a good bit more familiar looking and approachable. I would also guess that it is a lot more intuitive to understand what the different terms are doing, and how they compete against each other. But what about actually solving it? This is what we will turn our attention to today.

5.0.1 Recovering the Boltzmann Distribution

Before we start thinking about a full solution, an interesting question we might ask is whether or not our Fokker-Planck equation converges to the appropriate equilibrium distribution. To answer this question, it is convenient to rewrite the Fokker-Planck equation in the form of a continuity equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial J}{\partial x} \tag{39}$$

where the *probability current*,

$$J = -\frac{P}{\gamma}\frac{\partial V}{\partial x} - D\frac{\partial P}{\partial x}$$

At equilibrium, we expect that $\frac{\partial P}{\partial t} = 0$, so $\frac{\partial J}{\partial x} = 0$. In other words, J is constant. In particular, in a system with an infinite domain, in order to be at steady state, the current has to be 0. Otherwise, the probability would be drifting over time.

$$\begin{aligned}
&\frac{P}{\gamma}\frac{\partial V}{\partial x} + D\frac{\partial P}{\partial x} = 0 \\
\rightarrow &-\frac{1}{\gamma D}\frac{\partial V}{\partial x} = \frac{1}{P}\frac{\partial P}{\partial x} \\
\rightarrow &-\frac{1}{\gamma D}\frac{\partial V}{\partial x} = \frac{\partial}{\partial x}\ln P \\
\rightarrow &-\frac{1}{\gamma D}V(x) + C = \ln P \\
\rightarrow &P(x) = e^C \exp\left(-\frac{V(x)}{\gamma D}\right)
\end{aligned}$$

Using the Einstein relation, $D = \frac{kT}{\gamma}$, we get

$$P(x) = e^C \exp\left(-\frac{V(x)}{kT}\right) \tag{40}$$

Finally, we can identify the constant out front, e^C, to be fixed by normalization, and so we get the Boltzmann distribution:

$$P(x) = \frac{1}{Z} e^{-\frac{V(x)}{kT}} \tag{41}$$

There is no kinetic term in the energy, since we began the derivation by setting mass to 0. As an aside, we show here that the long-time limit of the inviscid Fokker-Planck equation is the Boltzmann distribution. The general question of whether the Fokker-Planck equation for any arbitrary diffusion process converges to an equilibrium is not so easy to solve. Finally, before we move on its worth spending a bit of time on the condition that we used to derive the equilibrium distribution:

$$J = -\frac{P}{\gamma}\frac{\partial V}{\partial x} - D\frac{\partial P}{\partial x} = 0$$

This is a special case of a requirement called *detailed balance*. In words, what it says is that at equilibrium, every transition from a state $x \to x + dx$ is perfectly balanced out by the reverse process, $x + dx \to x$. This allows the distribution to remain stationary over time. In particular, it is the state in which the *drift* of the probability density with time, due to the potential, is perfectly balanced out by the diffusive term, leading to a steady state.

5.0.2 Fokker, Planck, and Schrödinger Walk into an $\hbar$

Stationary distributions are one thing, but can we actually solve the beast? The answer is that, like many PDEs you encounter in physics, yes, but it might be hard. There is a rather magical insight that makes the whole thing a bit easier, though. Let's see how its done.

When presented with a PDE, what is your first instinct? For me, at least, it is separation of variables. If you recall, separation of variables works by taking a PDE, assuming that it has solution $P(x,t) = \psi(x)\phi(t)$, and then seeing if such a solution is separable. In general, this yields some sort of eigenvalue problem for the solutions, $P(x,t)$. However, in order to do this and guarantee that we are going to get all of the possible solutions, we need to understand the structure of the differential operator a bit better. When you do separation of variables, you generally rely on the fact that eventually, you will be able to superpose the various eigenfunctions that you get, in order to build up any arbitrary solution. This relies on a) the linearity of the operator, and b) the completeness of the eigenfunctions. Is the Fokker-Planck operator linear? In this case, yes. Let's denote the Fokker Planck operator as $\mathcal{L}$.

$$\begin{aligned}
\mathcal{L}(\alpha A + \beta B) &= \frac{1}{\gamma}\frac{\partial}{\partial x}\cdot\left((\alpha A + \beta B)\frac{\partial V}{\partial x}\right) + D\frac{\partial^2}{\partial x^2}(\alpha A + \beta B) \\
&= \frac{1}{\gamma}\left[\alpha\left(\frac{\partial A}{\partial x}\frac{\partial V}{\partial x} + A\frac{\partial^2 V}{\partial x^2}\right) + \beta\left(\frac{\partial B}{\partial x}\frac{\partial V}{\partial x} + B\frac{\partial^2 V}{\partial x^2}\right)\right] + D\frac{\partial^2}{\partial x^2}(\alpha A + \beta B) \\
&= \alpha\frac{1}{\gamma}\frac{\partial}{\partial x}\cdot\left(A\frac{\partial V}{\partial x}\right) + D\frac{\partial^2}{\partial x^2}A + \beta\frac{1}{\gamma}\frac{\partial}{\partial x}\cdot\left(B\frac{\partial V}{\partial x}\right) + D\frac{\partial^2}{\partial x^2}B \\
&= \alpha\frac{\partial A}{\partial t} + \beta\frac{\partial B}{\partial t}
\end{aligned}$$

Okay, so that lets us apply superposition. Is it Hermitian? Unfortunately, here we are out of luck. As you will show on the problem set, the Fokker Planck operator is not Hermitian, and so we can't apply the spectral theorem to say that the eigenfunctions will be complete. What to do then? Here's where the magic happens. It turns out to bear a striking resemblance to Schrödinger's equation (in imaginary time).

We make the change of variables $P(x,t) = e^{-\frac{V(x)}{2\gamma D}}\tilde{P}(x,t)$. For the time being, this is going to seem like it comes kind of out of thin air, but bear with me for a moment. After all the algebra, I promise to return to the physics. If we substitute that into our Fokker-Planck equation, we get

$$\begin{aligned}
\frac{\partial}{\partial t}\left(e^{-\frac{V(x)}{2\gamma D}}\tilde{P}(x,t)\right) &= \frac{1}{\gamma}\frac{\partial}{\partial x}\cdot\left(\left(e^{-\frac{V(x)}{2\gamma D}}\tilde{P}(x,t)\right)\frac{\partial V}{\partial x}\right) + D\frac{\partial^2}{\partial x^2}\left(e^{-\frac{V(x)}{2\gamma D}}\tilde{P}(x,t)\right) \\
e^{-\frac{V}{2\gamma D}}\frac{\partial}{\partial t}\tilde{P} &= \frac{1}{\gamma}\left[\frac{\partial}{\partial x}\left(e^{-\frac{V}{2\gamma D}}\tilde{P}\right)\cdot\frac{\partial V}{\partial x} + \left(e^{-\frac{V}{2\gamma D}}\tilde{P}\right)\frac{\partial^2 V}{\partial x^2}\right] + D\frac{\partial^2}{\partial x^2}\left(e^{-\frac{V}{2\gamma D}}\tilde{P}\right) \\
e^{-\frac{V}{2\gamma D}}\frac{\partial\tilde{P}}{\partial t} &= \frac{1}{\gamma}\left[\left(\frac{\partial\tilde{P}}{\partial x}e^{-\frac{V}{2\gamma D}} - \frac{\tilde{P}V}{2\gamma D}\frac{\partial V}{\partial x}e^{-\frac{V}{2\gamma D}}\right)\cdot\frac{\partial V}{\partial x} + e^{-\frac{V}{2\gamma D}}\tilde{P}\frac{\partial^2 V}{\partial x^2}\right] + D\frac{\partial^2}{\partial x^2}\left(e^{-\frac{V}{2\gamma D}}\tilde{P}\right) \\
\frac{\partial\tilde{P}}{\partial t} &= \frac{1}{\gamma}\left[\left(\frac{\partial\tilde{P}}{\partial x} - \frac{\tilde{P}V}{2\gamma D}\frac{\partial V}{\partial x}\right)\cdot\frac{\partial V}{\partial x} + \tilde{P}\frac{\partial^2 V}{\partial x^2}\right] + De^{\frac{V}{2\gamma D}}\frac{\partial^2}{\partial x^2}\left(e^{-\frac{V}{2\gamma D}}\tilde{P}\right)
\end{aligned}$$

Lots of algebra later, you end up getting that

$$\boxed{\frac{\partial\tilde{P}}{\partial t} = \left[D\frac{\partial^2}{\partial x^2} + \underbrace{\frac{1}{2\gamma}\frac{\partial^2}{\partial x^2}V - \frac{1}{4\gamma^2 D}\left(\frac{\partial V}{\partial x}\right)^2}_{U_{\text{eff}}}\right]\tilde{P}} \quad \ddot{\smile}$$

which looks quite *similar* (pun intended) to the Schrödinger equation,

$$i\hbar\frac{\partial\psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U(x)\right]\psi$$

but with an effective potential. To make the connection a bit more explicit, we can also not that the Schrödinger equation is connected to the diffusion equation via what's called a *Wick rotation*. Take the Schrödinger equation for a free particle,

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi$$

and perform the mapping to imaginary time: $t \to -i\tau$. Then, $\frac{\partial}{\partial t} \to i\frac{\partial}{\partial\tau}$. It is a rotation in the sense that we have rotated time in the complex plane, from the real axis to the imaginary axis. Then, the Schrödinger equation becomes

$$\frac{\partial\psi}{\partial\tau} = \frac{\hbar}{2m}\frac{\partial^2}{\partial x^2}\psi$$

which is just a diffusion equation with diffusion constant $D = \frac{\hbar}{2m}$. If we add a potential, and apply the same change of variables as before, then we get the Fokker-Planck equation.

What's really going on here? It makes some sense that the Fokker-Planck equation and the Schrödinger equation are related. After all, they are both describing the time evolution of some sort of probability density. The point I

want to get across is that this is more than just a little bit of mathematical chicanery. There is both some really cool physics going on here, and some very useful math.

Let's start with the math. The Fokker-Planck equation, in general, is quite hard to solve. This is because, unlike Schrödinger's equation, the Fokker-Planck operator is not Hermitian. This means that, unlike the Schrödinger operator, it does not admit a complete set of orthogonal eigenfunctions, and so spectral methods of the sort that you apply in quantum mechanics are not very helpful. However, if we apply this change of variables, which mathematically is a *similarity transformation*, then the Fokker-Planck operator gets mapped to the self-adjoint Schrödinger operator, and we know how to solve Schrödinger's equation. We can then solve Schrödinger's equation for $\tilde{P}(x,t)$, and then invert the transformation to get $P(x,t)$. You'll see an example of this on the problem set.

So why this transformation? Its because what we are doing is essentially pulling out the *equilibrium distribution*. We already showed that at equilibrium, the Fokker Planck equation converges to $P(x) \propto \exp(-V(x)/\gamma D)$. We just pulled out a factor of the square root of this. The idea is that if we divide this out of the Fokker-Planck operator, then instead of describing the flow towards equilibrium, it starts describing the fluctuations around it.

But wait! There's more! Taking the ground state energy to be 0, solutions to the time dependent Schrödinger equation are given by

$$\psi(x,t) = \phi_0 + \sum_{n=1}^{\infty} \exp(-iE_n t/\hbar)\phi_n(x)$$

where ϕ_n is the n-th eigenfunction of the Hamiltonian. What does this look like after our Wick rotation?

$$\psi(x,t) = \phi_0 + \sum_{n} \exp(-E_n \tau/\hbar)\phi_n(x)$$

So, oscillatory phase dynamics in quantum mechanics become exponentially decaying modes in statistical physics, and once again, the spectral gap dictates how fast the system relaxes to equilibrium.

I should be careful here and point out that this is not a general purpose method for solving the Fokker-Planck equation. This trick only works for FP-operators which satisfy *detailed balance*, and even if it does work, solving the corresponding Schrödinger equation may be challenging. But it turns out to be a useful method regardless, as mappings have been found for a number of problems of interest. For example, the harmonic potential can be treated in this way, as you will do on the problem set. Finally, to complete the picture: How does Langevin dynamics factor into all this? Fokker-Planck corresponds to the view of Brownian motion that looks at the probability distribution over time. Recall that we also talked about how stochastic processes can be seen as probability distributions over trajectories. In the classical picture, this view is expressed via the Chapman-Kolmogorov equation. In the quantum mechanical version of this, we get Feynman's famous *path integral formulation of quantum mechanics*. The analogues between statistical mechanics and quantum mechanics are extremely rich, and unfortunately this is all that we will have time to explore. However, I encourage you all to look into this further!

The (h)-bartender says "You guys look similar."

5.0.3 Example: The Kramers Problem

As an example of the applications of the Fokker-Planck equation, we will consider the problem of escape through a potential barrier. Imagine we have a set of particles in a local minimum of the potential. On either side, the potential rises, forming a barrier. This is illustrated in 3. We are interested in computing the average rate at which particles escape from the potential well. This is the Kramers' problem, and has wide ranging applications in chemical kinetics (where reaction rate calculations can be modeled in this way), or in quantum computing, where it is used to studying Josephson junctions. We will look at a very simple example of a Kramers' problem, where the potential well is locally quadratic around the minimum ($V(x) \approx \frac{1}{2}\omega^2(x - x_{\min})^2$).

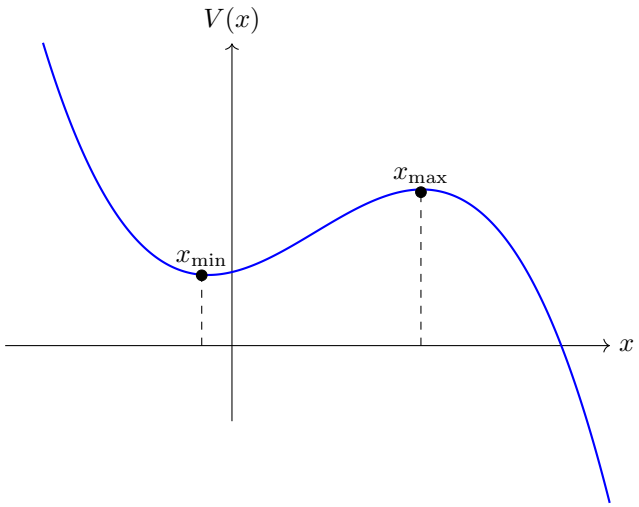


Figure 3: An example metastable potential

One angle of attack is to start from the Smoluchowski equation and calculate the probability current, which gives the rate of probability flow from the local minimum outwards.

$$J = -\frac{P}{\gamma}\frac{\partial V}{\partial x} - D\frac{\partial P}{\partial x} \tag{42}$$

Using the Einstein relation, we say that

$$J = -\frac{P}{\gamma}\frac{\partial V}{\partial x} - \frac{kT}{\gamma}\frac{\partial P}{\partial x} \tag{43}$$

Let's assume that the initial condition is a delta function in the center of the well. That is, all particles start at $x_{\min}$. Now, when the potential well is tall enough ($V(x_{max}) \gg T$), then we can assume that the probability current is small (the rate at which particles escape the well is low), and so the relaxation will occur in two phases: first, the particles will relax into their equilibrium distribution in the well, and then once the distribution of particles in the well reaches steady state, then the dominant effect will become the leakage.

At equilibrium, we know that the particles ought to obey the Boltzmann distribution,

$$P(x) = \frac{1}{Z}\exp\left(-\frac{V(x)}{kT}\right) = \sqrt{\frac{\omega^2}{2\pi kT}}\exp\left(-\frac{\omega^2(x-x_{\min})^2}{2kT}\right)$$

At this point, it is useful to rewrite the probability current in the following way:

$$J = -\frac{kT}{\gamma}\frac{\partial P}{\partial x} - \frac{P}{\gamma}\frac{\partial V}{\partial x} = -\frac{kT}{\gamma}e^{-\frac{V(x)}{kT}}\frac{\partial}{\partial x}\left(Pe^{\frac{V(x)}{kT}}\right)$$

Now, we will invoke the equilibrium assumption; when the system is at equilibrium (or steady state), then the probability current is constant. Now, the argument we are making here is worth spending some time on. Like many of the approximations we will see in the course, it is a statement about time-scales. As mentioned earlier, there

are two stages to the escape dynamics. First, the dynamics are dominated by the relaxation to the Boltzmann distribution, and then the escape process becomes dominant. When we say the probability current is constant, we are saying that the relaxation time is much, much faster than the escape time. As particles escape, the distribution does deviate from the Boltzmann distribution, but we are assuming that the system relaxes back to local equilibrium very quickly.

Then, we can think of the shape of the distribution being Boltzmann at all times; it is just the amplitude decreasing over time. If the shape is not changing in the barrier region, then we know that the net current within the barrier region must be constant (in space, not time), since if the current varied over the barrier region, then probability would build up in certain regions and reduce in others.

$$J_0 e^{\frac{V(x)}{kT}} = -\frac{kT}{\gamma}\frac{\partial}{\partial x}\left(Pe^{\frac{V(x)}{kT}}\right)$$

We want the rate of escape over the barrier, which we can get by integrating this equation over the barrier,

$$J_0\int_{x_{\min}}^{\infty} e^{\frac{V(x)}{kT}}dx = -\frac{kT}{\gamma}\int_{x_{\min}}^{\infty}\frac{\partial}{\partial x}\left(Pe^{\frac{V(x)}{kT}}\right)dx = -\frac{kT}{\gamma}\left(Pe^{\frac{V(x)}{kT}}\right)\Bigg|_{x_{\min}}^{\infty}$$

As we argued before, the probability far away from the potential barrier is vanishingly small, so we get

$$\begin{aligned}-\frac{kT}{\gamma}\left(Pe^{\frac{V(x)}{kT}}\right)\Bigg|_{x_{\min}}^{\infty} &= \frac{kT}{\gamma}\left(P(x_{\min})e^{\frac{V(x_{\min})}{kT}}\right)\\ &= \frac{kT}{\gamma}\sqrt{\frac{\omega^2}{2\pi kT}}\end{aligned}$$

Now we have to deal with the integral on the left hand side. This requires us to come up with an expression for the barrier potential. Let's say that it is also parabolic:

$$V(x) \approx V_{\max} - \frac{1}{2}\tilde{\omega}^2(x - x_{\max})^2 \tag{44}$$

Then, we have

$$\begin{aligned}J_0\int_{x_{\min}}^{\infty} e^{\frac{V(x)}{kT}}dx &= J_0\int_{x_{\min}}^{\infty} e^{\frac{V_{\max}-\frac{1}{2}\tilde{\omega}^2(x-x_{\max})^2}{kT}}dx\\ &= J_0 e^{\frac{V_{\max}}{kT}}\int_{x_{\min}}^{\infty} e^{\frac{-\tilde{\omega}^2(x-x_{\max})^2}{2kT}}dx\\ &\approx J_0 e^{\frac{V_{\max}}{kT}}\int_{-\infty}^{\infty} e^{\frac{-\tilde{\omega}^2(x-x_{\max})^2}{2kT}}dx\\ &= J_0 e^{\frac{V_{\max}}{kT}}\sqrt{\frac{2\pi kT}{\tilde{\omega}^2}}\end{aligned}$$

Equating the expressions for the left and right hand sides, we get

$$\boxed{J_0 = \frac{\omega\tilde{\omega}}{2\pi\gamma}e^{-\frac{V_{\max}}{kT}}} \tag{45}$$

This is the tunneling rate out of the barrier.

5.0.4 Example: Brownian Motors

I want to conclude our discussion of Brownian motion with a really neat example of how Brownian motion is put to use in our bodies. The concept is something known as a Brownian motor. It goes like this.

Imagine with have some asymmetric, periodic potential, like a sawtooth. For example, $V(x) = c\sum_{k=1}^{\infty}\frac{1}{k\pi}\sin(k\pi x)$. Now, imagine the strength of this potential is oscillatory, such that it turns on and off at regular intervals. What happens? When the potential is off, the particles diffuse evenly, due to Brownian motion. Then, when the potential turns on, the particles snap to the bottom of whichever potential well they currently find themselves in. Since the potential wells are asymmetric, the particles will preferentially move in one direction! Note that if the temperature is high enough and the potential is strong enough, this mechanism can even be used to generate motion against some external field.

I have shared a code example on the course website which demonstrates this mechanism visually, and you can see the different stages in Fig. 4.

Notice how both the asymmetric potential and the thermal noise are necessary to generate motion. Without the noise, the particles would never diffuse into the neighboring wells, and without the asymmetry of the potential, the diffusion would not generate a net current.

This type of mechanism, called noise-induced transport, is vital for many physical processes in our body. For example, some motor proteins in our body are able to transport cargo around cells by hydrolyzing a molecule called ATP (creating a sawtooth potential), and then undergoing free Brownian motion when the energy from that reaction dissipates. Nature is truly marvelous.

6 Lecture 5 - Kinetic Theory

6.1 The Liouville Equation

Our approach to non-equilibrium thermodynamics in the previous section has been largely phenomenological. We used our knowledge of the equilibrium distribution, the classical forces, and kind of enforced the necessary irreversibility and late time behavior by inserting the appropriate random force. It turns out that we can come to a deeper understanding by attempting to relate *reversible*, microscopic behavior (think Newton's laws or Schrödinger's equation) to the thermodynamic limit. Our goals in this section will be to a] develop an understanding of how one can relate classical mechanics to statistical mechanics, and when, specifically, irreversibility enters the picture, b] to derive basic examples of transport equations, and c] to derive Navier-Stokes from statistical mechanics. We will start this ambitious project by exploring a fundamental result in Hamiltonian mechanics: *the Liouville equation.*

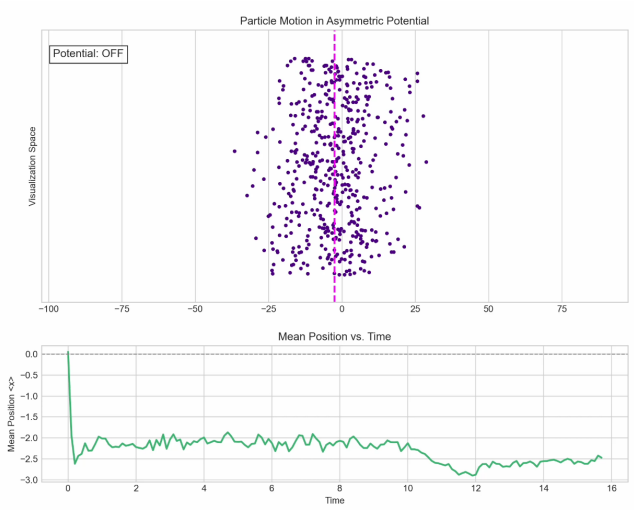
Recall that in Hamiltonian mechanics, we work in *phase space*, a $6N-$dimensional space, where particle coordinates are given by their generalized positions, q_i and momenta, p_i. Hamilton's equations dictate a system's trajectory through phase space,

$$\dot{q_i} = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{46}$$

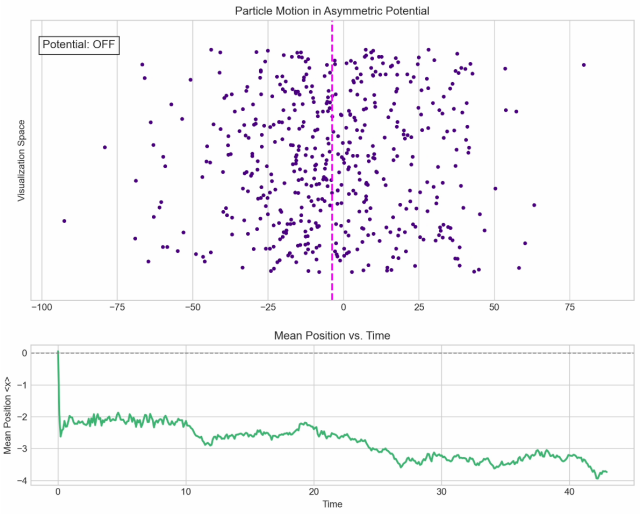
We are interested in a probability density over phase space, $f(\vec{r}_i, \vec{p}_i, t)$, which is both normalized and locally conserved. Like any conserved quantity, the probability density obeys a continuity equation:

$$\frac{\partial f}{\partial t} + \nabla \cdot (f\dot{\mathbf{X}}) = 0 \tag{47}$$

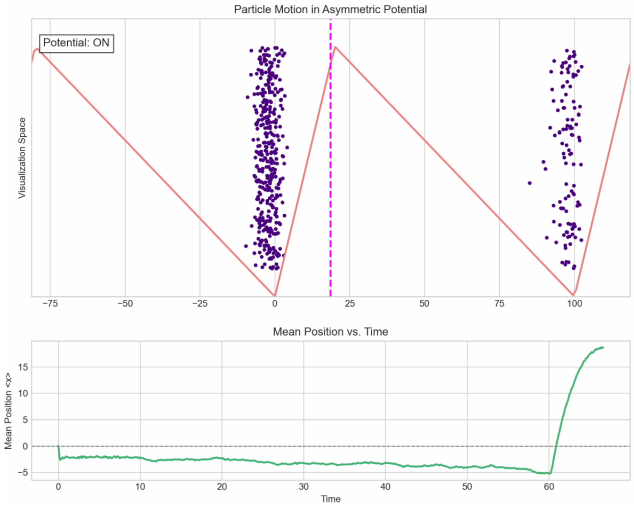
where we have introduced the phase-space coordinate, $\mathbf{X}$, defined as $\dot{\mathbf{X}} = (\dot{q}_1, \dot{q}_2, ..., \dot{q}_N, \dot{p}_1, ..., \dot{p}_N)$. Writing this



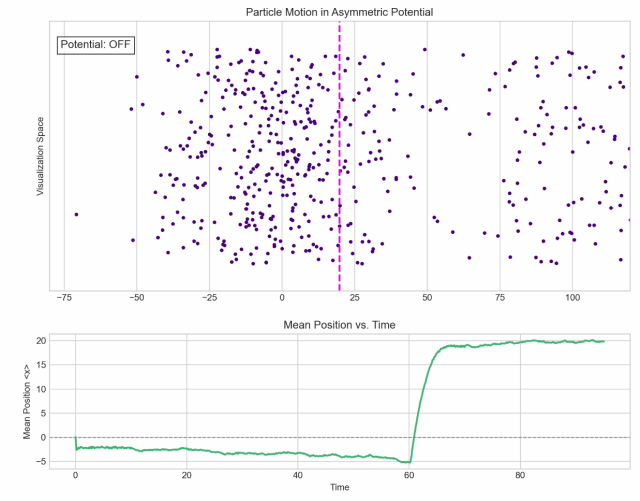
(a) The particles begin clustered around $x = 0$



(b) They undergo Brownian motion, diffusing.



(c) The potential turns on, trapping most particles back at $x = 0$, but some escape into the next potential well.



(d) The particles diffuse again, but the mean position has shifted in the positive direction.

Figure 4: The basic premise of a Brownian motor.

in terms of the positions and momenta, we get

$$
\begin{aligned}
0 &= \frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{q}_i} \cdot (f\dot{\vec{q}}_i) + \frac{\partial}{\partial \vec{p}_i} \cdot (f\dot{\vec{p}}_i) \\
&= \frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{q}_i} \cdot \left(f \frac{\partial H}{\partial \vec{p}_i} \right) + \frac{\partial}{\partial \vec{p}_i} \cdot \left(-f \frac{\partial H}{\partial \vec{q}_i} \right) && \text{Hamilton's equations} \\
&= \frac{\partial f}{\partial t} + \left[\frac{\partial^2 H}{\partial \vec{q}_i \partial \vec{p}_i} f + \frac{\partial f}{\partial \vec{q}_i} \frac{\partial H}{\partial \vec{p}_i} - \frac{\partial^2 H}{\partial \vec{q}_i \partial \vec{p}_i} f - \frac{\partial H}{\partial \vec{q}_i} \frac{\partial f}{\partial \vec{p}_i} \right] && \text{product rule} \\
&= \frac{\partial f}{\partial t} + \left[\frac{\partial f}{\partial \vec{q}_i} \frac{\partial H}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{q}_i} \frac{\partial f}{\partial \vec{p}_i} \right] \\
&= \frac{\partial f}{\partial t} - \{H, f\} && \text{definition of Poisson bracket}
\end{aligned}
$$

This last equality is known as *Liouville's equation*, and it gives the time evolution of the phase space density. As a practical matter, Liouville's equation is not something that we can really solve in statistical mechanics. The phase space has some $\mathcal{O}(10^{23})$ coordinates, and the equation is intractable. However, we can derive a number of important results from this equation, including some lower order equations that are easier to work with. In terms of a total derivative, Liouville's equation is equivalent to saying $\frac{df}{dt} = 0$, which is to say that the phase space distribution behaves as an incompressible fluid. Its shape can change, and it can translate across phase space, but its volume cannot change.

By definition, the system is at equilibrium when its phase space density has no explicit time dependence, and so the phase space distribution Poisson-commutes with the Hamiltonian. The easiest way to achieve this is by making the phase space distribution itself a function of the Hamiltonian (for example, the Boltzmann distribution).

Now, a recurring theme in this section will be that of irreversibility. The macroscopic kinetic equations that we come to will not be time reversible, and so it is a worthwhile exercise to determine exactly which point along this line of reasoning the reversibility condition breaks down. Hamilton's equations are obviously time-reversible; taking derivatives with respect to $-t$ yields the same equations. Liouville's equation follows directly from Hamilton's equations, so we can see that Liouville's equation is also time reversible.

As a last point, for those of you who have not seen Poisson brackets before, it is okay! Our main use for them in this sections is as a notational convenience. It is simply easier to write the bracket than to write out the derivatives. One property that we will use, however, is the fact that $\{H, f(H)\} = 0$. That is, if the distribution function is a function of the Hamiltonian, then their Poisson bracket vanishes.

6.2 The BBGKY Hierarchy

Now, in principle, the Liouville equation tells us everything we need to know about the evolution of a random ensemble. However, it tells us far more than we actually need to know in order to compute macroscopics. Furthermore, since we are directly working in phase space, the Liouville equation very quickly becomes intractable.

Instead, we can try to work with the *one-body* distribution function:

$$f_1(\vec{q}, \vec{p}, t) = N \int \Pi_{i=2}^{N} d^3q_i d^3p_i f(q_i, p_i, t) \tag{48}$$

This is the distribution function for a single particle (in our case the first, but since they are identical, it does not matter). We would like to come up with an equation of motion for this one-body density, as it would, in principle be a lot easier to work with.

$$\begin{aligned} \frac{\partial f_1}{\partial t} &= N \int \Pi_{i=2}^{N} d^3q_i d^3p_i \frac{\partial f(q_i, p_i, t)}{\partial t} \\ &= N \int \Pi_{i=2}^{N} d^3q_i d^3p_i \{H, f\} \end{aligned}$$

Let's assume that the full, N particle Hamiltonian takes the following form:

$$H_N = \underbrace{\sum_{i=1}^{N} \frac{\vec{p}_i^2}{2m} + \sum_{i=1}^{N} V(\vec{q}_i)}_{\text{one-body}} + \underbrace{\sum_{i=1}^{N} \sum_{j<i} U(|\vec{q}_i - \vec{q}_j|)}_{\text{interaction}} \tag{49}$$

Using this, we can explicitly compute the Poisson bracket, which ends up giving

$$\frac{\partial f_1}{\partial t} = \{H_1, f_1\} + N \int \Pi_{i=2}^{N} d^3q_i d^3p_i \sum_{k=2}^{N} \frac{\partial U(\vec{q}_1 - \vec{q}_k)}{\partial \vec{q}_1} \frac{\partial f}{\partial \vec{p}_1} \tag{50}$$

Now what does this tell us? The first term on the right hand side is the same as Liouville's equation. The second term, called the "collision integral", is new. So we see that the one-body density evolves in the typical Liouville manner, plus some extra that depends on the full, many body density. This is unfortunate, since the goal was to get an equation of motion that did not depend on the many body density. However, we shouldn't be surprised to see that the collision integral involves the many-body particle density. After all, it accounts for the *collisions* between particles.

However, we have done something very important here. We have isolated the dependence on the many body term from the dependence on the one-body term. This allows us to much more easily make approximations. In fact, we can make some further headway by noting that all particles are identical, and so every term in the sum is identical. We can choose the second particle to be representative of all k, and then, the collision term can be rewritten as

$$\left(\frac{\partial f}{\partial t}\right)_{coll} = N(N-1) \int \Pi_{i=2}^{N} d^3q_i d^3p_i \frac{\partial U(\vec{q}_1 - \vec{q}_2)}{\partial \vec{q}_1} \frac{\partial f}{\partial \vec{p}_1} \tag{51}$$

If you are confused about why the terms are identical, consider the $k = 2$ term. We are integrating $\frac{\partial U_{12}}{\partial \vec{q}_1} \frac{\partial f}{\partial \vec{p}_1}$ over $d^3q_1 d^3p_1 d^3q_2 d^3p_2 ....$ For the $k = 3$ term, we are integrating $\frac{\partial U_{13}}{\partial \vec{q}_1} \frac{\partial f}{\partial \vec{p}_1}$ over $d^3q_1 d^3p_1 d^3q_2 d^3p_2 ....$ Since we are just permuting what particle's relative position we are integrating, and we know that the particle labels are just arbitrary labels, we can say all terms in the sum are the same. We can write this a little more suggestively as

$$\left(\frac{\partial f}{\partial t}\right)_{coll} = N(N-1) \int d^3q_2 d^3p_2 \frac{\partial U(\vec{q}_1 - \vec{q}_2)}{\partial \vec{q}_1} \cdot \frac{\partial}{\partial \vec{p}_1} \underbrace{\int \Pi_{i=3}^{N} d^3q_i d^3p_i f}_{f_2, \text{ the 2-body distribution}} \tag{52}$$

So we see that the one-body distribution evolves in a Liouvillian manner, with a correction given by the two body distribution, which by the same reasoning as above, itself evolves in a Liouvillian manner, except for a correction given by the three body distribution, and so on, until its collision integrals all the way down. The point is that the n-body distribution function evolves as

$$\frac{\partial f_n}{\partial t} = \{H_n, f_n\} + \sum_{i=1}^{n} \int d^3q_{n+1} d^3p_{n+1} \frac{\partial U(\vec{q}_i - \vec{q}_{n+1})}{\partial \vec{q}_i} \frac{\partial f_{n+1}}{\partial \vec{p}_i} \tag{53}$$

These equations are called the *BBGKY* hierarchy. They are a system of N coupled integro-differential equations. By plugging in the one and two body Hamiltonians and evaluating the Poisson bracket, we see that the first two equations of the hierarchy read

$$\begin{aligned}
\frac{\partial f_1}{\partial t} &= \frac{\partial H}{\partial \vec{q}} \frac{\partial f_1}{\partial \vec{p}} - \frac{\partial H}{\partial \vec{p}} \frac{\partial f_1}{\partial \vec{q}} + \left(\frac{\partial f_1}{\partial t}\right)_{coll} \\
&= \frac{\partial V}{\partial \vec{q}} \frac{\partial f_1}{\partial \vec{p}} - \frac{\vec{p}}{m} \frac{\partial f_1}{\partial \vec{q}} + \left(\frac{\partial f_1}{\partial t}\right)_{coll} \\
&= \frac{\partial V}{\partial \vec{q}} \frac{\partial f_1}{\partial \vec{p}} - \frac{\vec{p}}{m} \frac{\partial f_1}{\partial \vec{q}} + \int d^3\vec{q}_2 d^3\vec{p}_2 \frac{\partial U(|\vec{q}_2 - \vec{q}_1|)}{\partial \vec{q}_1} \frac{\partial f_2}{\partial \vec{p}_1}
\end{aligned}$$

and

$$\frac{\partial f_2}{\partial t} + \frac{\partial V}{\partial \vec{q}_1}\frac{\partial f_2}{\partial \vec{p}_1} + \frac{\partial V}{\partial \vec{q}_2}\frac{\partial f_2}{\partial \vec{p}_2} + \frac{\vec{p}_1}{m}\frac{\partial f_2}{\partial \vec{q}_1} + \frac{\vec{p}_2}{m}\frac{\partial f_2}{\partial \vec{q}_2} - \frac{1}{2}\frac{\partial U(\vec{q}_1 - \vec{q}_2)}{\partial \vec{q}_1}\left(\frac{\partial f_2}{\partial \vec{p}_1} - \frac{\partial f_2}{\partial \vec{p}_2}\right)$$
$$= \int d^3q_3 d^3p_3 \left(\frac{\partial U(|\vec{q}_3 - \vec{q}_1|)}{\partial \vec{q}_1}\frac{\partial f_3}{\partial \vec{p}_1} + \frac{\partial U(|\vec{q}_3 - \vec{q}_2|)}{\partial \vec{q}_2}\frac{\partial f_3}{\partial \vec{p}_2}\right)$$

You notice the equations get messy very fast. Fortunately, we don't need to go any further up the hierarchy for our purposes. The point will be to determine what approximations we must make in order to truncate the hierarchy. Before we move on to some example kinetic equations, the last point I want to make is that the BBGKY hierarchy is fully time reversible, which follows from its direct equivalence to the full Liouville equation. Evidently, we have not yet introduced an arrow of time.

7 Lecture 6 - Boltzmann's Equation

We concluded our discussion of Brownian motion with the observation that, for the most part, we took a very heuristic approach. We leveraged our knowledge of equilibrium statistical mechanics to gain insight into what the appropriate construction of non equilibrium dynamics should be. We then asked if it would be possible to start from classical mechanics, and somehow derive statistical laws. We started from Hamilton's equations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}$$

and derived *Liouville's equation*, which is the equation of motion for the phase space density.

$$\frac{\partial f}{\partial t} - \{H, f\} = 0$$

While this equation looks deceptively simple, we made the observation that in reality, it is nigh impossible to use in statistical mechanics, since the dimension of the phase space we are working in is $6N$, where N is the number of particles. In statistical mechanics, this is frequently $\mathcal{O}(10^{23})$, and so solving Liouville's equation is hopeless. Instead, we asked the question: how does the *average* particle evolve? That is, how does the *one-body* density,

$$f_1(\vec{q}, \vec{p}, t) = N \int \Pi_{i=2}^N d^3q_i d^3p_i f(q_i, p_i, t)$$

Assuming that the Hamiltonian is given by

$$H_N = \underbrace{\sum_{i=1}^N \frac{\vec{p}_i^2}{2m} + \sum_{i=1}^N V(\vec{q}_i)}_{\text{one-body}} + \underbrace{\sum_{i=1}^N \sum_{j<i} U(|\vec{q}_i - \vec{q}_j|)}_{\text{interaction}}$$

we got that

$$\frac{\partial f_1}{\partial t} = \{H_1, f_1\} + N(N-1)\int d^3q_2 d^3p_2 \frac{\partial U(\vec{q}_1 - \vec{q}_2)}{\partial \vec{q}_1} \cdot \frac{\partial}{\partial \vec{p}_1} \underbrace{\int \Pi_{i=3}^N d^3q_i d^3p_i f}_{f_2, \text{ the 2-body distribution}}$$

That is, we saw that the one-body distribution function evolves according to Liouville's equation for the single particle system, plus some correction that depends on the two-body distribution function. This correction is often referred to as the "*collision integral*", denoted $\left(\frac{\partial f}{\partial t}\right)_{coll}$, since it is an integral, and it dictates evolution due to two body collisions. So, in order to solve the hierarchy, we need to know how that two body distribution evolves. I told you that the two body distribution, by similar arguments, can be shown to evolve according to the two particle Hamiltonian, plus some correction due to three body collisions. In general, the n-body distribution evolves according to the n-particle Hamiltonian, plus some correction due to the $n+1$-particle distribution.

$$\frac{\partial f_n}{\partial t} = \{H_n, f_n\} + \sum_{i=1}^{n} \int d^3q_{n+1} d^3p_{n+1} \frac{\partial U(\vec{q}_i - \vec{q}_{n+1})}{\partial \vec{q}_i} \frac{\partial f_{n+1}}{\partial \vec{p}_i}$$

You might rightly ask yourself: How does this help at all? Instead of a single very hard problem (a differential equation for the evolution of a 10^{23} dimensional distribution), we now have 10^{23} problems, most of which don't seem any easier. The point of writing this in terms of a hierarchy is that we have separated out the dependence on the higher order distributions into specific terms of each equation. We can then more easily apply approximations based on our expectations of which equations in the hierarchy are important. We will see how this works today.

Before we do that though, it is worth dwelling for a moment on some of the major themes of this section of the course. In particular, one idea we will repeatedly return to is the notion of *irreversibility*. Classical mechanics is perfectly time reversible, and yet, statistical mechanics is not. We have some intuition by this point that irreversibility is an emergent behavior. It is a result of coarse graining, a symptom of our ignorance about the underlying state of the system. However, we have not yet injected that ignorance into our mathematics. The BBGKY hierarchy is directly equivalent to the full Liouville equation, and so it is also fully time reversible.

7.1 The Collisionless Boltzmann Equation

The first equation in the hierarchy can be rewritten in Poisson bracket notation:

$$\frac{\partial f_1}{\partial t} + \{f_1, H_1\} = \left(\frac{\partial f_1}{\partial t}\right)_{coll}$$

Now, as you may have noticed, the really tricky part to handle is the collision integral. We will deal with this properly in the next section, but a simpler way to deal with annoying things is to pretend they don't exist. So what happens if we set the collision integral to 0? This happens to be a reasonable approximation for systems which are sufficiently dilute, to the point where collisions are negligible. This particular approximation is especially common in plasma physics, where it goes by the name the *Vlasov equation*, and in astrophysics, where it is called the *collisionless Boltzmann equation* (CBE).

Writing the Poisson bracket out explicitly, we have

$$\frac{\partial f_1}{\partial t} - \frac{\partial V}{\partial \vec{q}} \cdot \frac{\partial f_1}{\partial \vec{p}} + \frac{\vec{p}}{m} \cdot \frac{\partial f_1}{\partial \vec{q}} = 0$$

When we add collisions back in, we will get the full Boltzmann equation, but first, let's see what the simpler, collisionless case is good for. It's worth doing, especially since the procedure we will develop carries over almost identically in the collisional case, but is a lot simpler here.

7.1.1 Example: Galactic Dynamics

A cool application of the collisionless Boltzmann equation is in the modeling of *galactic dynamics*. In a typical galaxy, the stars interact via some gravitational potential, but they are so *astronomically* (haha) far apart that they almost never approach close enough to collide. Because of this, the CBE provides a good approximation. (Technically, collisionless is a bit of a misnomer. The proper way to think of this approximation is that the interaction distance is large enough that each star is interacting with many many other stars at any given time, so that this is close to a mean field approximation, as opposed to some sort of collisional system).

In this field, it is conventional to write the CBE in terms of the velocities, and to replace the potential energy V with the gravitational potential, $\phi = V/m$.

$$\frac{\partial f_1}{\partial t} - \frac{\partial \phi}{\partial \vec{q}}\frac{\partial f_1}{\partial \vec{v}} + \vec{v}\frac{\partial f_1}{\partial \vec{q}} = 0$$

Even the single body phase space density contains more information than we are necessarily interested in. In particular, we are generally more concerned with the *spatial* variation in a quantity, not the velocity space variation. To address this, we take the so-called velocity moments of the CBE.

In general the n-th moment is given by

$$\nu_{i_1 i_2 ... i_n}(\vec{x}, t) = \int v_{i_1} ... v_{i_n} f_1(\vec{x}, \vec{v}, t) d^3 v$$

The zeroth moment is the case where we don't multiply by any velocity. Instead, we just integrate the distribution function over velocity space and get

$$\nu(\vec{x}) = \int d^3 v f_1(\vec{x}, \vec{v}, t)$$

You'll note that this is just the number density of particles, since it tells us the average number of particles at any given location. All three of the first moments are

$$\bar{v_i}(x) = \frac{1}{\nu(x,t)} \int d^3 v \; v_i f_1(\vec{x}, \vec{v}, t)$$

That is, we can see that this gives just the average velocity of the distribution at each spatial location. We conventionally divide by the particle density since we are interested in the *average* velocity per particle. Overdense regions should not be overrepresented. The second moments, now, follow in the same manner.

$$\overline{v_i v_j} = \frac{1}{\nu(x,t)} \int d^3 v \; v_i v_j f_1(\vec{x}, \vec{v}, t)$$

This is often written in terms of the so called *velocity dispersion tensor*:

$$\sigma_{ij} = \langle (v_i - \bar{v_i})(v_j - \bar{v_j}) \rangle \rightarrow \overline{v_i v_j} = \sigma_{ij} + \overline{v_i} \cdot \overline{v_j}$$

This is the same as in single variable statistics when you say that the variance is defined as

$$\sigma^2 = \overline{x^2} - (\overline{x})^2$$

Okay, so these are the moments of the distribution function. We use these to define *moment equations*, which are usually easier to solve than the full, 6D collisionless Boltzmann equation. The procedure is as follows: We take the collisionless Boltzmann equation, multiply it by some combination of velocities, and then integrate over space.

The zeroeth moment equation is just given by doing the integration over velocity space:

$$\int d^3v \left[\frac{\partial f_1}{\partial t} - \frac{\partial \phi}{\partial \vec{q}}\frac{\partial f_1}{\partial \vec{v}} + \vec{v}\frac{\partial f_1}{\partial \vec{q}}\right] = 0$$
$$\frac{\partial}{\partial t}\int d^3v f_1 - \int d^3v \, \frac{\partial \phi}{\partial \vec{q}}\frac{\partial f_1}{\partial \vec{v}} + \int d^3v \, \vec{v}\frac{\partial f_1}{\partial \vec{q}} = 0$$
$$\frac{\partial}{\partial t}\int d^3v f_1 - \frac{\partial \phi}{\partial \vec{q}}\int d^3v \, \frac{\partial f_1}{\partial \vec{v}} + \frac{\partial}{\partial \vec{q}}\cdot\int d^3v \, \vec{v} f_1 = 0$$

The integral in the first term is just the number density (zeroth moment), and if we pull the spatial derivative out of the third term, what remains is the density times our first velocity moment, so

$$\frac{\partial}{\partial t}\nu(x,t) - \frac{\partial \phi}{\partial \vec{q}}\int d^3v \, \frac{\partial f_1}{\partial \vec{v}} + \sum_i \frac{\partial}{\partial q_i}(\nu(x,t)\cdot\overline{v_i}(x,t)) = 0$$

Is there a way we can simplify the middle term? We can eliminate it by applying a particular form of the divergence theorem:

$$\iiint_V \mathbf{F}\cdot(\nabla g)\,\mathrm{d}V = \oiint_S g\mathbf{F}\cdot\mathbf{n}\,\mathrm{d}S - \iiint_V g(\nabla\cdot\mathbf{F})\,\mathrm{d}V$$

In this case, $\mathbf{F} = \frac{\partial \phi}{\partial \vec{q}}$, since we are integrating the middle term over velocity space and taking the velocity gradient, and $\nabla g = \frac{\partial f_1}{\partial \vec{v}}$, so

$$\frac{\partial \phi}{\partial \vec{q}}\int d^3v \, \frac{\partial f_1}{\partial \vec{v}} = \oiint_S f_1 \frac{\partial \phi}{\partial \vec{q}}\cdot\hat{\mathbf{n}} - \iiint_V f_1 \cdot \left(\nabla_v \cdot \frac{\partial \phi}{\partial \vec{q}}\right) d^3V$$

The surface integral can be made to vanish by taking the surface out to infinity (in velocity space), where distribution function must vanish in order to be normalizable. The volume integral also vanishes, since the gravitational potential does not depend on velocity, and so $\nabla_v \cdot \frac{\partial \phi}{\partial \vec{q}} = 0$. Then,

$$\frac{\partial}{\partial t}\nu(x,t) + \sum_i \frac{\partial}{\partial q_i}(\nu(x,t)\cdot\overline{v_i}(x,t)) = 0$$

This looks just like a regular continuity equation, which we might have anticipated, since we started with a continuity equation in phase space, but simply integrated out the velocity variables. So, the zeroth moment equation is just the continuity equation. Importantly, notice that the zeroth moment equation is a function of the first moment. What about the first moment equations? We get these by multiplying the CBE through by the velocity in each direction:

$$\int d^3v \, v_i \cdot \left[\frac{\partial f_1}{\partial t} - \frac{\partial \phi}{\partial \vec{q}}\frac{\partial f_1}{\partial \vec{v}} + \vec{v}\frac{\partial f_1}{\partial \vec{q}}\right] = 0$$

Integrating by parts and expressing our results in terms of our average velocities, we have

$$\frac{\partial}{\partial t}(\nu\overline{v_j}) + \frac{\partial}{\partial x_i}(\nu\overline{v_i v_j}) + \frac{\partial \phi}{\partial x_i}\int f\frac{\partial v_j}{\partial v_i}\, d^3\vec{v} = 0$$

In Cartesian (and in fact any orthogonal) coordinates

$$\frac{\partial v_j}{\partial v_i} = \delta_{ij}$$

so the last term on the left hand side becomes $\nu\delta_{ij}$. We can then apply the product rule to the first term, and then use the continuity equation to get

$$\nu\frac{\partial \bar{v}_j}{\partial t} - \bar{v}_j\frac{\partial}{\partial x_i}(\nu\bar{v}_i) + \frac{\partial}{\partial x_i}\left[\nu(\sigma_{ij}^2 + \bar{v}_i\bar{v}_j)\right] = -\nu\bar{v}_i\frac{\partial\phi}{\partial x_j}$$

where we have made use of the relation between the second moments and the velocity dispersion. Finally, we can apply the product rule to the second term on the left hand side, as part of the result cancels part of the third term and we arrive Jeans' equations for a collisionless fluid,

$$\nu\frac{\partial \bar{v}_j}{\partial t} + \bar{v}_i\nu\frac{\partial \bar{v}_j}{\partial x_i} = -\nu\frac{\partial\phi}{\partial x_j} - \frac{\partial}{\partial x_i}\left(\nu\sigma_{ij}^2\right) \tag{54}$$

The first term on the left is the force (per unit volume), the second is the (kinematic) viscosity, the first on the right is the gravitational force, and the second on the right is the pressure.

There's something very important to note about the Jeans equations though. There's 3 of them, plus the equation of continuity. However, there are more unknowns because the dispersion tensor depends on all the second moments. If we derive an equation for the second moments, we would get something that depended on the third moments, and so on! Once again, a hierarchy...

The technical term for this is that the Jeans equations are not *closed*. In the case of collisional systems, we will get around this by using an *equation of state*, or a constitutive equation, to provide closure. In collisionless systems, however, the way this is usually done is by assuming some form for the dispersion tensor. As an example, we can compute the mass surface density of a stellar disk. What's the idea? Well, observationally, we can see the number of stars in various locations, and we want to infer something about the distribution of mass in the galaxy. Let's pretend that the galaxy is an infinite slab, with plane parallel symmetry and reflection symmetry about the xy-plane, and it is in steady state. We will try and compute the surface area mass density, $\Sigma(z)$:

$$\Sigma(z) = \int_{-z}^{z}\rho(z)dz$$

You can think of this as we are taking a vertical column through our slab and coming up with the mass distribution along that line. We start with the $j = 3$ Jeans equation:

$$\nu\frac{\partial \bar{v}_z}{\partial t} + \bar{v}_i\nu\frac{\partial \bar{v}_z}{\partial x_i} = -\nu\frac{\partial\phi}{\partial z} - \frac{\partial}{\partial x_i}\left(\nu\sigma_{zz}^2\right)$$

We then apply plane parallel symmetry to argue that $\frac{\partial v}{\partial y} = \frac{\partial v}{\partial x} = 0$, so

$$\nu\frac{\partial \bar{v}_z}{\partial t} + \bar{v}_z\nu\frac{\partial \bar{v}_z}{\partial z} = -\nu\frac{\partial\phi}{\partial z} - \frac{\partial}{\partial z}\left(\nu\sigma_{zz}^2\right)$$

At steady state, $\frac{\partial v}{\partial t} = 0$, so

$$\bar{v}_z\nu\frac{\partial \bar{v}_z}{\partial z} = -\nu\frac{\partial\phi}{\partial z} - \frac{\partial}{\partial z}\left(\nu\sigma_{zz}^2\right)$$

Finally, reflection symmetry gives us that $\bar{v}_z = 0$, so

$$\nu \frac{\partial \phi}{\partial z} = -\frac{\partial}{\partial z}\left(\nu \sigma_{zz}^2\right) \rightarrow \frac{1}{\nu}\frac{\partial}{\partial z}\left(\nu \sigma_{zz}^2\right) = -\frac{\partial \phi}{\partial z}$$

Note that $-\frac{\partial \phi}{\partial z} = g_z$, the gravitational field in the z-direction. Next, we apply Gauss's law (same as you do in PHYS 23 for an infinite sheet of charge). Gauss's Law for a gravitational field $\mathbf{g}$ through a closed surface S is given by:

$$\oint_S \mathbf{g} \cdot d\mathbf{A} = -4\pi G M_{\rm enc}$$

We define a cylindrical "pillbox" with cross-sectional area A that extends from $-z$ to $+z$ across the midplane. By symmetry, the gravitational field $\mathbf{g}$ points only in the z-direction: $\mathbf{g} = g_z \hat{\mathbf{k}}$. Since the field must point toward the midplane, we have that the total flux through the top and bottom faces is:

$$\oint \mathbf{g} \cdot d\mathbf{A} = (g_z \cdot A)_{\rm top} - (g_z \cdot A)_{\rm bottom} = 2 g_z A$$

The mass enclosed within the pillbox is the surface density $\Sigma(z)$ (the mass per unit area between $-z$ and z) multiplied by the area A:

$$M_{\rm enc} = \Sigma(z) A$$

Substituting these into Gauss's Law:

$$2 g_z A = -4\pi G \Sigma(z) A \tag{55}$$

Dividing by $2A$, we find:

$$g_z = -2\pi G \Sigma(z) \tag{56}$$

Using the definition of the gravitational potential, $g_z = -\frac{\partial \phi}{\partial z}$, we obtain the expression used in the Jeans Equation:

$$-\frac{\partial \phi}{\partial z} = -2\pi G \Sigma(z) \tag{57}$$

$$\nu \frac{\partial \phi}{\partial z} = -\frac{\partial}{\partial z}\left(\nu \sigma_{zz}^2\right) \rightarrow \frac{1}{\nu}\frac{\partial}{\partial z}\left(\nu \sigma_{zz}^2\right) = -2\pi G \Sigma(z)$$

which gives the final form that we will use:

$$-\frac{\sigma_{zz}^2}{2\pi G}\frac{\partial}{\partial z}\left(\ln \nu\right) = \Sigma(z)$$

So why go through all this? Well, the number density, ν, is something we can directly observe. We can look at the population of stars in the plane of a galaxy, and count their distribution. Similarly, we can measure the velocity dispersion. This then gives us a measure of the total (visible) mass per unit area of the disk of a galaxy. For the Milky Way, this yields about 50 $M_\odot {\rm pc}^{-2}$.

7.2 Boltzmann's equation

Recall that the whole point of deriving the BBGKY hierarchy was to come up with an equation of motion for the one-body distribution. The one body distribution lives in 6D phase space, which is obviously far more manageable than the full Liouville equation, but our equation of motion for the one-body distribution is still coupled to the two body distribution via the collision integral. The goal of this section is to present the most common way in which this is handled.

Our next goal is the derivation of Boltzmann's equation, which is the kinetic equation for a rarefied gas. When we say rarefied, we mean that we are assuming that the time between collisions, τ, is much greater than the time over which a collision actually occurs. This is important, since this assumption allows us to take the view that binary collisions are the dominant interaction in the gas, and higher order collisions can be ignored. The BBGKY hierarchy gives us the collision integral in terms of the two-body distribution function. Now our goal is to rewrite this in terms of one body densities. The first approximation we will make is to turn off the external potential, V. This is simply to make our lives easier.

The principled way to do this derivation is to start from the BBGKY hierarchy. By analyzing the various time scales and distance scales in the problem, one can derive Boltzmann's equation. This derivation, however, is rather tricky. I will refer you either to Kardar or to Tong if you are curious. We will instead take a more heuristic route.

Boltzmann's equation is really just a fancy name for the first equation in the hierarchy, where we will give a particular form to the collision integral:

$$\frac{\partial f_1}{\partial t} - \{H_1, f_1\} = \left(\frac{\partial f_1}{\partial t}\right)_{coll}$$

Warning: Derivatives here are somewhat misleading. Recall here that we are working with a system where we have assumed that the time scale of streaming is much longer than the timescale of the collision. In this sense, the derivative *cannot* be understood as a limit where $\Delta t \to 0$, since that would inevitably run into the regime where $\Delta t = \mathcal{O}(\tau_{coll})$.

What we have to do now is come up with an expression for the collision term in terms of the one-body distribution function. As we saw earlier, it really depends on the two body distribution function, but with the help of some physical reasoning, we will be able to make a good bit of headway.

Suppose we have a single particle at some coordinate in phase space $(\vec{r}, \vec{p}_1)$, when it collides with another particle with coordinates $(\vec{r}, \vec{p}_2)$. For the purposes of this argument, we will assume that collisions are local. That is, they occur as a result of two particles reaching the same location in position space, not mediated by some long-range interaction potential.

After the collision, the particles emerge with some outgoing momenta, $\vec{p_1}'$ and $\vec{p}_2'$. We know that the two body density gives us the probability that two particles sit at $(\vec{r}, \vec{p}_1)$ and $(\vec{r}, \vec{p}_2)$, so we can define the rate of this scattering process to be

$$\text{Rate} = \omega(\vec{p}_1, \vec{p}_2 \to \vec{p}_1', \vec{p}_2') \cdot f_2(\vec{r}, \vec{r}, \vec{p}_1, \vec{p}_2) d^3p_2 d^3p_1' d^3p_2'$$

where ω is some function that captures the dynamics of the scattering. This can be computed from classical mechanics, but we won't do that here.

Now of course, particles also can and will be scattered *into* the $(\vec{r}, \vec{p}_1), (\vec{r}, \vec{p}_2)$ state. By the same arguments, the rate is given by

$$\text{Rate} = \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot f_2(\vec{r}, \vec{r}, \vec{p}_1, \vec{p}_2) d^3p_2 d^3p_1' d^3p_2'$$

and so the change to the distribution function induced by the total collisions will be

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int [\omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot f_2(\vec{r}, \vec{r}, \vec{p}_1', \vec{p}_2') - \omega(\vec{p}_1, \vec{p}_2 \to \vec{p}_1', \vec{p}_2') f_2(\vec{r}, \vec{r}, \vec{p}_1, \vec{p}_2)] d^3p_2 d^3p_1' d^3p_2'$$

Now, without going into the details of how scattering works, we can place some more restrictions on the scattering function ω. In particular, we can use every physicists favorite style of argument: *symmetries*. In particular, we can use the symmetries of spacetime.

We know that the scattering dynamics should look the same when time-reversed. If we reverse time, obviously the momenta are reversed. That is $\vec{p}_1$ becomes $-\vec{p}_1$, $\vec{p}_2$ becomes $-\vec{p}_2$ and so on. Additionally, notions of "initial" and "final" state are swapped. This tells us that

$$\omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) = \omega(-\vec{p}_1, -\vec{p}_2 \to -\vec{p}_1', -\vec{p}_2')$$

Next, we can apply parity symmetry to argue that

$$\omega(\vec{p}_1, \vec{p}_2 \to \vec{p}_1', \vec{p}_2') = \omega(-\vec{p}_1, -\vec{p}_2 \to -\vec{p}_1', -\vec{p}_2')$$

since applying a parity transformation just reverses the direction of the momenta.

Finally, we can assume that the scattering function is translation invariant (that is, the scattering rate is the same at different positions), and combine these all to get:

$$\omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) = \omega(\vec{p}_1, \vec{p}_2 \to \vec{p}_1', \vec{p}_2')$$

That is, the scattering function is symmetric under interchange of the incoming and outgoing momenta. All this work gives us enough information to simplify the collision integral to

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_2(\vec{r}, \vec{r}, \vec{p}_1', \vec{p}_2') - f_2(\vec{r}, \vec{r}, \vec{p}_1, \vec{p}_2)] d^3p_2 d^3p_1' d^3p_2'$$

All that work to factor one thing out! And now, the last step. This is the big one. We are going to make the assumption that the particle momenta are *uncorrelated* before the collision. That is,

$$f_2(\vec{r}, \vec{r}, \vec{p}_1, \vec{p}_2) = f_1(\vec{r}, \vec{p}_1) f_1(\vec{r}, \vec{p}_2)$$

This seems like a very reasonable assumption. After all, why would the velocities be correlated before the collision? This assumption, however, is far from simple. It is so important, in fact, that it gets its own name: the *Stosszahlanzatz*, or assumption of *molecular chaos*, and it has been the source of over 150 years of arguments in statistical physics. It is so important because it introduces an *arrow of time* into the picture. The particle momenta are clearly correlated after the collision, but we are assuming that they are not correlated before. This, it turns out, will be a sufficient condition for irreversibility.

$$\boxed{\frac{\partial f_1}{\partial t} - \{H_1, f_1\} = \int \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_1(\vec{r}, \vec{p}_1') f_1(\vec{r}, \vec{p}_2') - f_1(\vec{r}, \vec{p}_1) f_1(\vec{r}, \vec{p}_2)] d^3p_2 d^3p_1' d^3p_2'} \ \text{☺}$$

This is *Boltzmann's equation*. It is, as you can tell, a nonlinear, integro-differential equation, with precious few analytic solutions. However, it is of fundamental importance because of two things: 1] the myriad analytic results derivable from it, and 2] it is the starting point for numerical simulation of a whole zoo of physical phenomena. This one gets a ☺.

Before we move on, I would be remiss if I didn't mention a few things:

1. While we have not compute the scattering function here, it is not hard to do so in simple cases like hard sphere scattering. This is worked out in Tong's lecture notes.

2. In this derivation, we assumed only binary collisions. The stage at which to incorporate the effects of ternary or quaternary collisions is clearer if you read the derivation of Boltzmann's equation from the BBGKY hierarchy. However, it is a *mess* to do analytically, as higher order collisions requires more and more equations in the hierarchy. Proceed at the risk of your own sanity.

3. As mentioned earlier, Boltzmann's equation is nasty to solve. As we will see, we can make some headway (some major headway, actually), by applying a pair of approximations called "the relaxation time approximation" and the "Chapman-Enskog expansion". However, Boltzmann's equation itself is the basis for a swathe of numerical techniques for simulation of various physical quantities. It forms the basis for the lattice Boltzmann method in computational fluid mechanics and the particle-in-cell method in computational plasma physics for example.

7.3 Consequences of Boltzmann's Equation and the Stosszahlanzatz

Before we discuss how to actually solve the Boltzmann equation, there are a host of insights we can make by just studying properties of potential solutions.

7.3.1 Detailed Balance and Local Equilibrium

The first question we might ask about the Boltzmann equation is: What kinds of distributions are the equilibrium distributions? Remember, equilibrium is the special state where the macroscopic quantities are not changing in time. That is,

$$\frac{\partial f_1}{\partial t} = 0$$

Boltzmann's equation tells us that the time evolution of the distribution function comes in two parts: the streaming term and the collision integral. As we already know, the streaming term vanishes for any function of the Hamiltonian. If we look at the case where there is no external potential, then this means that the streaming term vanishes for *any* function of the momentum. Clearly this doesn't fit with our intuitions about unique equilibrium states. So what does the collision term do?

One obvious way to make the collision integral vanish is to insist that

$$f_1^{eq}(\vec{p}_1) f_1^{eq}(\vec{p}_2) = f_1^{eq}(\vec{p}_1') f_1^{eq}(\vec{p}_2')$$

or, equivalently

$$\log f_1^{eq}(\vec{p}_1) + \log f_1^{eq}(\vec{p}_2) = \log f_1^{eq}(\vec{p}_1') + \log f_1^{eq}(\vec{p}_2')$$

Recalling that the unprimed coordinates are the momenta of the particles before the collision, and the primed coordinates are the momenta after it, we can see that this is just saying that this combination of logs is conserved during a collision.

But we know what things are conserved during a collision: energy and momentum! So, we can make the identification that

$$\log f_1^{eq}(\vec{p}_1) = \beta(\mu - E(\vec{p}_1) + \vec{u} \cdot \vec{p}_1)$$

Here, β, μ, and $\vec{u}$ are constants, but the symbols contain foreshadowing. Since there is no external field, we also know that $E(\vec{p}) = \frac{\vec{p}^2}{2m}$. Then, writing $\vec{p} = m\vec{v}$, exponentiating both sides, and fixing μ by normalization gives:

$$f_1^{eq}(\vec{r}, \vec{p}_1) = \frac{N}{V} \left(\frac{\beta}{2\pi m} \right)^{\frac{3}{2}} e^{-\frac{1}{2}\beta m (\vec{v} - \vec{u})^2}$$

We used the identity here that

$$\vec{a} \cdot \vec{b} = \frac{||\vec{a}||^2 + ||\vec{b}||^2 - ||\vec{a} - \vec{b}||^2}{2}$$

The coefficient out front is fixed by normalization, so we recognize this to be our usual Boltzmann distribution! So, we've learned that the collision term requires that the equilibrium distribution be of the Boltzmann form. Here, $\vec{u}$ allows for an overall drift velocity. For example, I could have set my entire gas to be moving with some constant velocity, but it still will equilibrate.

One more point that will be important later: If we forget the requirement that the streaming term vanish, then there are many more functions that satisfy detailed balance. In particular, all of the constants, $\mu, \vec{u}$, and β are now allowed to be functions of space and time. This is because the requirement that they only be functions of momentum came from the requirement that the distribution function Poisson commuted with the Hamiltonian. These more general functions are said to be in *local equilibrium*.

7.4 Recap

A short summary of our discussion of kinetic theory thus far. We started with Liouville's equation, which dictated the evolution of a probability distribution over phase space, and from there derived a hierarchy of equations that iteratively marginalized over a single particle at a time, until we were left with an equation of motion for the single particle distribution function:

$$\frac{\partial f_1}{\partial t} - \{H_1, f_1\} = \left(\frac{\partial f_1}{\partial t}\right)_{coll}$$

That is, we determined that the single body distribution function evolved according to something that looks quite similar to Liouville's equation, but modified by a collision integral. We then looked at the collisionless Boltzmann equation, which is what one gets if we just set the collision integral to 0. In particular, the two most important things to remember about this section was a] the fact that without collisions, the equation of motion is symmetric under time reversal, and b] that we could derive moment equations that are simpler to solve than the full CBE. We then turned our attention to the collision integral. We made a series of symmetry arguments to derive that

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_2(\vec{r}, \vec{p}_1', \vec{r}, \vec{p}_2') - f_2(\vec{r}, \vec{p}_1, \vec{r}, \vec{p}_2)] d^3p_2 d^3p_1' d^3p_2'$$

Finally, we invoked the assumption of *molecular chaos* to argue that the two body distribution should factorize, giving:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_1(\vec{r}, \vec{p}_1') f_1(\vec{r}, \vec{p}_2') - f_1(\vec{r}, \vec{p}_1) f_1(\vec{r}, \vec{p}_2)] d^3p_2 d^3p_1' d^3p_2'$$

This assumption that the particle momenta are uncorrelated before the collision is, in some sense, the root of all evil. We feel that it is physically unfounded. We feel that since the total system dynamics are in principle known, the momenta before a collision should somehow be correlated. This is a good objection, and is true. However, molecular chaos is an assumption about our ignorance. We don't know the system's true state, and so we have no better choice than to assume that the momenta are uncorrelated.

Finally, we asked ourselves what sorts of distributions are the equilibrium distributions of the Boltzmann equation, and we showed that a Maxwellian meets the criteria:

$$f_1^{eq}(\vec{r}, \vec{p}_1) = \frac{N}{V}\left(\frac{\beta}{2\pi m}\right)^{\frac{3}{2}} e^{-\frac{1}{2}\beta m(\vec{v}-\vec{u})^2}$$

We will now turn our attention to the consequences of Boltzmann's equation.

7.5 The H-theorem

We are now in a position to tackle the thorny question of how reversible classical dynamics lead to irreversible statistical dynamics. We touched on this a moment ago when we discussed molecular chaos, but as you might expect, there is a lot more to say.

Let's start by just briefly discussing what exactly molecular chaos is saying. Recall that we are looking for the time evolution of the distribution function. What molecular chaos does is it says that the particles' momenta are *uncorrelated* before the collision. Obviously the particles momenta become correlated during a collision, but Boltzmann's equation promptly forgets about such correlations, since we make the independence assumption for all future collisions as well.

So why is this an acceptable physical model? Obviously, correlations develop over time and grow increasingly complex. So what are we really saying here? We can gain some insight by returning, conceptually to Liouville's equation. In the case of Liouville's equation, the system starts in some state, and then evolves over time to a deterministic final state. In that sense, its entropy is constant. The problem, of course, is that we can't keep track of all of those correlations. Molecular chaos is just a way of mathematically injecting our ignorance into the picture.

Boltzmann's H-theorem, then, is one particular way of demonstrating the irreversibility of Boltzmann's equation. The argument goes as follows. Define H as

$$H = \int d^3p d^3r f_1(\vec{r}, \vec{p}, t) \log f_1(\vec{r}, \vec{p}, t)$$

Some of you may recognize this as the negative of the *information entropy*, or Shannon entropy, of f_1. *Historical aside: Boltzmann originally called this quantity E for entropy, but some historical hijinks later led to it being called H. Shannon then, in a nod to Boltzmann, denoted his entropy as H, which is the modern, information theoretic notation for entropy.*

We can then compute how H changes with time:

$$\begin{aligned}
\frac{dH}{dt} &= \frac{d}{dt} \int d^3p d^3r f_1(\vec{r}, \vec{p}, t) \log f_1(\vec{r}, \vec{p}, t) \\
&= \int d^3p d^3r \frac{\partial f_1}{\partial t} \log f_1 + \frac{f_1}{f_1} \cdot \frac{\partial f_1}{\partial t} \\
&= \int d^3p d^3r \frac{\partial f_1}{\partial t} (\log f_1 + 1)
\end{aligned}$$

where from the second line onward, I dropped the arguments of f_1. The first line to the second uses the product rule, the second to the third is just algebra. We can drop the $+1$ since

$$\int d^3p d^3r \frac{\partial f_1}{\partial t} = \frac{\partial}{\partial t} \int d^3p d^3r f_1 = \frac{\partial}{\partial t} 1 = 0$$

by normalization of the distribution function, so that leaves us with

$$\frac{dH}{dt} = \int d^3p d^3r \frac{\partial f_1}{\partial t} \log f_1$$

Some of you may be confused how the total derivative became a partial derivative. This is because we are not differentiating along a particle path. We are averaging over the entire phase space. We can substitute in Boltzmann's equation to get

$$\frac{dH}{dt} = \int d^3r d^3p \log f_1 \cdot \left[\{H_1, f_1\} + \left(\frac{\partial f_1}{\partial t} \right)_{coll} \right]$$

If you were to actually write out the Poisson bracket, you could confirm for yourself that that term vanishes, so that leaves us with

$$\begin{aligned}\frac{dH}{dt} &= \int d^3r d^3p \log f_1(\vec{p}_1) \cdot \left(\frac{\partial f_1}{\partial t}\right)_{coll} \\ &= \int \log f_1 \cdot \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_1(\vec{p}_1')f_1(\vec{p}_2') - f_1(\vec{p}_1)f_1(\vec{p}_2)] d^3r d^3p_1 d^3p_2 d^3p_1' d^3p_2'\end{aligned}$$

where I dropped the arguments of $\vec{r}$ for clarity. There is one thing to note about this expression: we integrate over all four momenta. We can swap the subscripts of 1 and 2, and we can swap the primed and unprimed momenta, and nothing changes. In particular, all the while, the scattering function remains unchanged, thanks to the symmetry properties we argued above, so lets do that.

Swapping the 1s and 2s only changes the log:

$$\frac{dH}{dt} = \int \log f_1(\vec{p}_2) \cdot \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_1(\vec{p}_1')f_1(\vec{p}_2') - f_1(\vec{p}_1)f_1(\vec{p}_2)] d^3r d^3p_1 d^3p_2 d^3p_1' d^3p_2'$$

Adding these two expressions for $\frac{dH}{dt}$ gives

$$\frac{dH}{dt} = \frac{1}{2}\int \log\left(f_1(\vec{p}_1)f_1(\vec{p}_2)\right) \cdot \omega(\vec{p}_1', \vec{p}_2' \to \vec{p}_1, \vec{p}_2) \cdot [f_1(\vec{p}_1')f_1(\vec{p}_2') - f_1(\vec{p}_1)f_1(\vec{p}_2)] d^3r d^3p_1 d^3p_2 d^3p_1' d^3p_2'$$

Now, swapping the primes and the unprimes in the above expression and again adding them together gives:

$$\frac{dH}{dt} = -\frac{1}{4}\int \omega \left[\log f_1(\vec{p}_1')f_1(\vec{p}_2') - \log f_1(\vec{p}_1)f_1(\vec{p}_2)\right] \cdot [f_1(\vec{p}_1')f_1(\vec{p}_2') - f_1(\vec{p}_1)f_1(\vec{p}_2)] d^3r d^3p_1 d^3p_2 d^3p_1' d^3p_2'$$

I dropped the argument in ω, but its the same as before. The result of all this chicanery is that we get an expression of the form $-(\log x - \log y)(x - y)$, which is strictly nonpositive. Thus,

$$\frac{dH}{dt} \leq 0 \to \frac{dS}{dt} \geq 0$$

There are a few caveats here. The first is that any distribution function that satisfies detailed balance will not reduce H with time. The second is that this derivation relied on molecular chaos, but we already established that molecular chaos is not the most well founded of assumptions. The takeaway is that molecular chaos is a *sufficient* condition for the second law to hold, but it is not a *necessary* one.

7.6 Paradoxes and the Kac Ring

Almost immediately after the publication of Boltzmann's equation and H-theorem, a number of his contemporaries objected quite vigorously to its derivation and implications, much like you likely are as well. The two most famous objections are due to Loschmidt and Zermelo. *Loschmidt's paradox* is the objection that classical dynamics are time-reversible, and thus cannot give rise to the manifestly irreversible dynamics of Boltzmann's equation. *Zermelo's paradox* is the objection that H cannot monotonically decrease, on the grounds that given sufficient time, the system will always return to its original state. Zermelo's argument hinged on something called *Poincare's*

recurrence theorem, which states that a dynamical system in a compact phase space *must* eventually get arbitrarily close to its initial state for almost any initial condition.

To shed some light on the resolution of these paradoxes, we turn to an ingenious model introduced by the mathematician Mark Kac: the eponymous *Kac* ring.

Imagine a ring of $N \gg 1$ sites, each occupied by a white or a black ball. In between $n < N$ pairs of sites, there are "markers.' At each time step, the balls all shift clockwise by one site, and those that pass a marker flip their color. The Kac ring is time reversible: if you followed the same laws but rotated counter clockwise, you would reverse the dynamics. It is also periodic. After $2N$ steps, each ball passes each marker twice, and so it ends up on its original spot with its original color. Some configurations might repeat faster than that, but the point is just that the system is periodic. Let $W(t)$ be the number of white balls at time t, and $B(t)$ be the number of black balls. Our goal will be to come up with an expression for $\delta(t)$, the difference between these, normalized by the number of balls: $\delta(t) = W(t) - B(t)$

The beauty of the Kac ring is that it is a system which can be solved both *exactly*, using a master equation, and by means of a kinetic equation. We will start with the kinetic approach. First, we will define the auxiliary quantities, $w(t)$ and $b(t)$, which are the number of white and black balls one step before a marker at time t. Then, we can clearly see that

$$W(t+1) = W(t) - w(t) + b(t)$$

and

$$B(t+1) = B(t) - b(t) + w(t)$$

We then have that

$$\delta(t+1) = W(t+1) - B(t+1) = W(t) - B(t) - 2w(t) + 2b(t)$$

Then problem is that we don't know $w(t)$ or $b(t)$ based on the macroscopic quantities, W, B, and δ. This is an important feature of the Kac ring: the time evolution of the macroscopic quantities depends on the microscopic quantities, and cannot be determined from the macroscopics alone. Thus, we are left to make assumptions. We would expect that for a typical initial configuration, with randomly placed balls,

$$\mu \equiv \frac{n}{N} = \frac{b}{B} = \frac{w}{W}$$

These relations are not strictly true for every possible initial configuration, but we might expect them to be correct on average. This is our equivalent of the molecular chaos assumption in the derivation of Boltzmann's equation. Then,

$$\delta(t+1) = \delta(t) - 2\mu W(t) + 2\mu B(t) = (1 - 2\mu)\delta(t)$$

which immediately yields

$$\delta(t) = (1 - 2\mu)^t \cdot \delta(0)$$

This equation is clearly inexact. It yield physically nonsensical results like $\delta(t) \notin \mathbb{Z}$ and it clearly exhibits an arrow of time, since $\delta(t)$ is decreasing. We can make some sense of this by noting that the assumption about μ was true for a *typical* ring, suggesting that we can interpret this as a sort of ensemble average. Let's say that we fix a particular initial configuration for the balls, and then we look at the average behavior of the system for random configurations of the markers.

Let $\chi_i(t)$ denote the color of the i-th ball at time t, with $\chi_i(t) = 1$ denoting a white ball and $\chi_i(t) = -1$ denoting a black ball. Let $m_i(t) = 1$ if there is a marker between sites i and $i+1$, and $m_i(t) = -1$ if there is none.

Then, we have that

$$\chi_i(t+1) = m_i \chi_i(t)$$

and

$$\delta(t) = \sum_i^N m_{i-1}\chi_i(t-1) = \sum_i^N m_{i-1}m_{i-2}...m_{i-t}\chi_{i-t}(0)$$

We can now take an average over all possible configurations of markers:

$$\langle\delta(t)\rangle = \sum_i^N m_{i-1}\chi_i(t-1) = \sum_i^N \langle m_{i-1}m_{i-2}...m_{i-t}\rangle\chi_{i-t}(0)$$

Now, we can take advantage of the fact that $\langle m_{i-1}m_{i-2}...m_{i-t}\rangle = \langle m_1 m_2 ... m_t\rangle$ since the probability of a marker on each edges is equal. Then,

$$\langle\delta(t)\rangle = \langle m_1 m_2 ... m_t\rangle \sum_i^N \chi_{i-t}(0) = \langle m_1 m_2 ... m_t\rangle \delta(0)$$

The last task is coming up with an equation for the quantity in angle brackets. When $0 \leq t < N$, we never revisit the same edge, so $m_1, \ldots, m_t$ are independent. The product is 1 for an even number of markers, and -1 for an odd number of markers, so we get

$$\langle m_1 m_2 \cdots m_t\rangle = \sum_{j=0}^{t} (-1)^j p_j(t) \tag{14}$$

where $p_j(t)$ denotes the probability of finding j markers on t consecutive edges. The markers follow a binomial distribution with parameter $\mu = n/N$, so that

$$p_j(t) = \binom{t}{j}\mu^j(1-\mu)^{t-j}$$

and, using the binomial theorem,

$$\langle m_1 m_2 \cdots m_t\rangle = \sum_{j=0}^{t}\binom{t}{j}(-\mu)^j(1-\mu)^{t-j} = (1-2\mu)^t$$

Inserting this expression into our equation for $\langle\delta(t)\rangle$, we obtain

$$\langle\delta(t)\rangle = (1-2\mu)^t\Delta(0)$$

which is the same result as we got from our kinetic equation, except this time, we explicitly carried out an ensemble average. The kinetic equation is not microscopically exact, but it is macroscopically correct for typical initial conditions and on appropriate time scales. In the same way, Boltzmann's equation is not exactly correct. It is a statement about the *typical* behavior of a system on typical timescales. The point of this model is to see, in gory, explicit detail, how the process of *coarse-graining* leads to irreversibility. The appearance of irreversibility is a statement about our ignorance, not the dynamical laws that govern our universe.

8 Lecture 7 - Hydrodynamics (Thermodynamics makes a Splash!)

Hydrodynamics is sometimes called the "long wavelength" limit of kinetics. What we mean by this is that it is the study of systems that are in *local equilibrium*, and thus systems where the thermodynamic variables change slowly over space and time. Again, "slowly" here means relative to the collision time. We know that collisions induce extremely fast relaxation back to equilibrium, and so the question will be if we can identify quantities that don't relax super fast. If so, we might expect that low energy, long wavelength excitations of those quantities might persist for long enough to exhibit some interesting behavior.

The way we will do this is by going back to the idea of *moment equations* of the Boltzmann equation.

8.0.1 Conservation Laws and Collisional Invariants

Let's start by considering a function over the single particle phase space, $\phi(\vec{r},\vec{p})$. In many cases, we don't actually care about how things vary in momentum space, just real space. Quantities like spatial density, the spatial temperature distribution, and the spatial pressure distribution are more important to understand than the corresponding densities in velocity space.

Because of this, we can follow the same procedure as we did when studying the collisionless Boltzmann equation, and integrate over momentum space:

$$\langle\phi(\vec{r},t)\rangle = \frac{\int d^3p\phi(\vec{r},\vec{p})f_1(\vec{r},\vec{p},t)}{\int d^3p f_1(\vec{r},\vec{p},t)}$$

The quantity in the denominator is simply the particle number density,

$$\langle\phi(\vec{r},t)\rangle = \frac{1}{n(\vec{r},t)}\int d^3p\ \phi(\vec{r},\vec{p})f_1(\vec{r},\vec{p},t)$$

Our goal is to come up with an expression for how $\langle\phi\rangle$ evolves in time. We went through this procedure for the Liouville evolution, but now we are asking the same question of the Boltzmann equation. It turns out, as you might expect, that the time evolution of $\langle\phi\rangle$, much like in the Liouvillian case, contains a streaming piece and a collision piece. We know what the collisions do: they cause a fast relaxation back to equilibrium. Since we want to look at interesting non-equilibrium behavior, this motivates us to look at functions ϕ, which vanish when integrated against the collision integral. That is,

$$\int d^3p\ \phi(\vec{r},\vec{p})\left(\frac{\partial f}{\partial t}\right)_{coll} = 0$$

Substituting in the expression for the collision integral, we get

$$\int d^3p_1 d^3p_2 d^3p_1' d^3p_2' \omega(\vec{p}_1',\vec{p}_2' \to \vec{p}_1,\vec{p}_2)\cdot[f_1(\vec{r},\vec{p}_1')f_1(\vec{r},\vec{p}_2') - f_1(\vec{r},\vec{p}_1)f_1(\vec{r},\vec{p}_2)]\phi(\vec{r},\vec{p})$$

This looks quite similar to our earlier expression, at the start of the derivation of the H-theorem, except that we have replaced $\log f_1$ with an arbitrary function, ϕ. Following the same arguments, we can rewrite the equation above as

$$\int d^3p_1 d^3p_2 d^3p_1' d^3p_2'\ \omega(\vec{p}_1',\vec{p}_2' \to \vec{p}_1,\vec{p}_2)$$
$$\cdot[f_1(\vec{r},\vec{p}_1')f_1(\vec{r},\vec{p}_2') - f_1(\vec{r},\vec{p}_1)f_1(\vec{r},\vec{p}_2)]\cdot[\phi(\vec{r},\vec{p}_1)+\phi(\vec{r},\vec{p}_2)-\phi(\vec{r},\vec{p}_1')-\phi(\vec{r},\vec{p}_2')]$$

which, if we wish for this quantity to vanish for *any* distribution function, then it is clear that

$$[\phi(\vec{r},\vec{p}_1)+\phi(\vec{r},\vec{p}_2)-\phi(\vec{r},\vec{p}_1')-\phi(\vec{r},\vec{p}_2')]=0$$

or, put in words, ϕ must be unchanged before and after the collision. Quantities such as this are known as *collisional invariants.* We know what they are: energy, momentum, and of course any constant. Now that we know that ϕ vanishes when integrated against the collision integral, we realize that we can repeat the exact same procedure as we did with the collisionless Boltzmann equation, and derive *moment equations.* Recall the procedure: take the CBE, multiply it by some combination of velocities, and then integrate over momentum. In that case, we multiplied by velocities because we wanted velocity moment equations. In principle, however, we could have multiplied by *any* function.

Not so in the case of the Boltzmann equation. This is because we need whatever quantity we multiply the equation by to vanish when integrated over the collision integral. This is why the collisional invariants are so important. Let's do this now. You will find that it looks quite familiar:

$$\int d^3p\,\phi(\vec{r},\vec{p})\left(\frac{\partial f_1}{\partial t}+\frac{\vec{p}}{m}\frac{\partial}{\partial\vec{r}}-\frac{\partial V}{\partial\vec{r}}\frac{\partial}{\partial\vec{p}}\right)f_1=0$$

In this step, I've multiplied Boltzmann's equation by ϕ and integrated over velocity space. The right hand side of Boltzmann's equation vanishes because ϕ is a collisional invariant. We can rewrite this equation in the following way:

$$\frac{\partial}{\partial t}\int d^3p\,\phi f+\frac{\partial}{\partial\vec{r}}\cdot\int d^3p\,\frac{\vec{p}}{m}\,\phi f-\int d^3p\,\frac{\vec{p}}{m}\cdot\frac{\partial\phi}{\partial\vec{r}}\,f+\int d^3p\,\frac{\partial V}{\partial\vec{r}}\cdot\frac{\partial\phi}{\partial\vec{p}}\,f=0$$

which may seem messier, but if we rewrite it in terms of our notation for averages, $\langle\cdot\rangle$, we get

$$\frac{\partial}{\partial t}\langle\phi n\rangle+\frac{\partial}{\partial\vec{r}}\cdot\langle n\vec{v}\phi\rangle-n\langle\vec{v}\cdot\frac{\partial\phi}{\partial\vec{r}}\rangle+n\langle\frac{\partial V}{\partial\vec{r}}\frac{\partial\phi}{\partial\vec{p}}\rangle=0 \tag{58}$$

Up to this point, we have worked entirely with an arbitrary collisional invariant, ϕ. Now, we need to plug in specific quantities. The easiest one to start with is $\phi=1$. This is clearly a collisional invariant, as a constant is unchanged by a collision. Plugging this in to Eq. 58, we get

$$\frac{\partial}{\partial t}\langle n\rangle+\frac{\partial}{\partial\vec{r}}\cdot\langle n\vec{v}\rangle=0$$

The second two terms vanish because $\phi=1$ is being differentiated. Recall that n is already an averaged quantity, so it doesn't change under $\langle\cdot\rangle$:

$$\frac{\partial}{\partial t}n+\frac{\partial}{\partial\vec{r}}\cdot n\langle\vec{v}\rangle=0$$

At this point, we can introduce the helpful shorthands that $\vec{u}\equiv\langle\vec{v}\rangle$, and $\rho=mn$. Multiplying the above equation through by m, we then get

$$\frac{\partial}{\partial t}\rho+\frac{\partial}{\partial\vec{r}}\cdot\rho\vec{u}=0$$

Make sure to keep in mind the fact that ρ and $\vec{u}$ are functions of space and time. Next, we look at our second collisional invariant, $\vec{p}=m\vec{v}_i$. If we substitute that into Eq. 58, we get that

$$\frac{\partial}{\partial t}\rho u_i + \frac{\partial}{\partial r_j}\cdot \rho\langle v_j v_i\rangle + n\langle\frac{\partial V}{\partial \vec{r}}\rangle = 0$$

The second term depends on the second moment of the velocity distribution, which we can rewrite as

$$\langle v_i v_j\rangle = \langle (v_i - u_i)(v_j - u_j)\rangle + u_i u_j$$

We can define the symmetric *pressure tensor*,

$$P_{ij} = P_{ji} = \rho\langle (v_i - u_i)(v_j - u_j)\rangle$$

and rewrite the equation as

$$\frac{\partial}{\partial t}\rho u_i + \frac{\partial}{\partial r_j}P_{ij} + \frac{\partial}{\partial r_j}\rho u_i u_j + n\langle\frac{\partial V}{\partial \vec{r}}\rangle = 0$$

or,

$$\rho\left(\frac{\partial}{\partial t} + u_j\frac{\partial}{\partial r_j}\right)u_i = -\left(\frac{\rho}{m}\langle\frac{\partial V}{\partial \vec{q}}\rangle + \frac{\partial}{\partial r_j}P_{ij}\right)$$

The left hand side is the acceleration of a fluid element. The right hand side has a sum of forces: the external force, and the internal pressure experienced by the fluid. It is important in interpreting this equation to remember that the acceleration is that of a particular parcel of fluid, not the rate of change of the fluid velocity at a fixed point in space.

This is known as the *Lagrangian frame*, which can be contrasted with the Eulerian frame, in which we view the fluid as moving through some fixed coordinate system. In fact, the combination of derivatives on the left,

$$D_t \equiv \frac{\partial}{\partial t} + u_j\frac{\partial}{\partial r_j}$$

is sometimes called the *Lagrangian derivative* (or material derivative, convective derivative, substantive derivative, and many other names). Similar to the case of the Jeans equations from last lecture, we also see that these equations are not *closed*. The continuity equation depends on us having an equation for $\vec{u}$, and the Euler equation depends on us having an expression for the pressure tensor. Both of these would require some knowledge about the distribution function.

Our final collisional invariant is kinetic energy. The slightly easier quantity to work with is the *relative* kinetic energy of the particles:

$$\phi = \frac{1}{2}m(\vec{u} - \vec{v})^2$$

Is this still a collisional invariant? Yes. Why? Expand it out. $\vec{u}(\vec{r}, t)$ is just a function of space, and so all four terms in the condition for collisional invariance,

$$[\phi(\vec{r}, \vec{p}_1) + \phi(\vec{r}, \vec{p}_2) - \phi(\vec{r}, \vec{p}_1') - \phi(\vec{r}, \vec{p}_2')] = 0$$

are the same. So the $\vec{u}^2$ term is a collisional invariant. Then, we know that $m\vec{v}$ and $m\vec{v}^2$ are the momentum and energy of the particles, which we know to be collisional invariants. So, plugging this into our master formula, Eq. 58, we get

$$\frac{1}{2}m\left[\frac{\partial}{\partial t}\langle(\vec{u}-\vec{v})^2 n\rangle + \frac{\partial}{\partial \vec{r}}\cdot\langle n\vec{v}(\vec{u}-\vec{v})^2\rangle - n\langle\vec{v}\cdot\frac{\partial(\vec{u}-\vec{v})^2}{\partial\vec{r}}\rangle + n\langle\frac{\partial V}{\partial\vec{r}}\frac{\partial}{\partial\vec{p}}(\vec{u}-\vec{v})^2\rangle\right] = 0 \tag{59}$$

The last term vanishes, since the momentum derivative can be written in terms of velocity, and then gives $-2(\vec{u}-\vec{v})$, which has zero mean. The force does not affect anything, since remember that the angle brackets are integrals over momentum, and the potential is a function of space, and so can be pulled out. We then get that

$$\frac{1}{2}\left[\frac{\partial}{\partial t}\langle\rho(\vec{u}-\vec{v})^2\rangle + \frac{\partial}{\partial\vec{r}}\cdot\langle\rho\vec{v}(\vec{u}-\vec{v})^2\rangle - \rho\langle\vec{v}\cdot\frac{\partial(\vec{u}-\vec{v})^2}{\partial\vec{r}}\rangle\right] = 0$$

The idea is to now define temperature such that

$$\frac{3}{2}kT(\vec{r},t) = \frac{1}{2}m\langle(\vec{v}-\vec{u}(\vec{r},t))^2\rangle$$

This coincides with the temperature that we expect from equipartition, but not that here, we have not enforced anything about equilibrium. Is this then still a valid definition of temperature? One thing you can do to convince yourself of this is to assume the system is in *local equilibrium*, and so the distribution function is a Maxwellian. Then, you can show that the statistical mechanical definition of temperature, $\frac{1}{T} = \frac{\partial S}{\partial U}$ is satisfied. However, this definition is not restricted to the case of local equilibrium. We also define the *heat flux*,

$$\vec{q} = \frac{1}{2}\rho\langle(\vec{v}-\vec{u})(\vec{v}-\vec{u})^2\rangle$$

Then, we can rewrite the middle term in our equation for the kinetic energy's rate of change as

$$\frac{\partial}{\partial\vec{r}}\cdot\langle\rho\vec{v}(\vec{u}-\vec{v})^2\rangle = \frac{\partial}{\partial\vec{r}}\cdot\langle\rho(\vec{v}-\vec{u})(\vec{u}-\vec{v})^2\rangle + \frac{\partial}{\partial\vec{r}}\cdot\vec{u}\langle\rho(\vec{u}-\vec{v})^2\rangle = 2\frac{\partial}{\partial\vec{r}}\cdot\vec{q} + 3kTn\frac{\partial}{\partial\vec{r}}\cdot\vec{u}$$

Then, we can rewrite our equation for the evolution of the kinetic energy in terms of temperature, to give

$$\left[\frac{3}{4}\frac{k}{m}\frac{\partial}{\partial t}\rho T + \frac{\partial}{\partial\vec{r}}\cdot\vec{q} + \frac{3}{2}kTn\frac{\partial}{\partial\vec{r}}\cdot\vec{u} - \frac{1}{2}\rho\left\langle\vec{v}\cdot\frac{\partial(\vec{u}-\vec{v})^2}{\partial\vec{r}}\right\rangle\right] = 0$$

If we use the same $+u$ and $-u$ trick, we can rewrite this in terms of the pressure tensor.

$$\left[\frac{3}{4}\frac{k}{m}\frac{\partial}{\partial t}\rho T + \frac{\partial}{\partial\vec{r}}\cdot\vec{q} + \frac{3}{2}kTn\frac{\partial}{\partial\vec{r}}\cdot\vec{u} - P_{ij}\frac{\partial u_j}{\partial r_i}\right] = 0$$

We can expand the first term and use the continuity equation to get

$$\frac{3\rho}{2m}\left(\frac{\partial}{\partial\vec{r}}\cdot\vec{u} + \frac{\partial}{\partial t}\right)kT + \frac{\partial}{\partial\vec{r}}\cdot\vec{q} - P_{ij}\frac{\partial u_j}{\partial r_i} = 0$$

Finally, we can multiply through by $\frac{2m}{3}$ to get

$$\rho\left(\frac{\partial}{\partial\vec{r}}\cdot\vec{u} + \frac{\partial}{\partial t}\right)kT + \frac{1}{m}\frac{\partial}{\partial\vec{r}}\cdot\vec{q} - \frac{2}{3}mP_{ij}\frac{\partial u_j}{\partial r_i} = 0$$

So, finally we have an equation for the evolution of the density, n, the velocity, $\vec{u}$, and the temperature T. However, just like in the case of the Jeans equations, we see that these equations are not closed. The equation for n depends on $\vec{u}$, the equation for $\vec{u}$ depends on P_{ij}, and the equation for T depends on $\vec{q}$. All of these quantities could

be computed from the distribution function, but to get the distribution function, we need to solve the Boltzmann equation. What to do?

As a first approximation, we can guess a form for the distribution function:

$$f_1^{(0)}(\vec{r},\vec{p},t) = \frac{n(\vec{r},t)}{(2\pi mkT(\vec{r},t))^{\frac{3}{2}}} \exp\left(-\frac{m(\vec{v}-\vec{u}(\vec{r},t))^2}{2kT(\vec{r},t)}\right)$$

Why guess this? Well, we are looking for slow variables. The distribution above is the one describing *local equilibrium* that we discussed earlier, and so it vanishes when integrated against the collision integral. Since collisions induce fast relaxation to equilibrium, we might expect that the equations that we derive here would exhibit some kind of persistent behavior.

We can plug this distribution into the definitions of P_{ij} and $\vec{q}$, and we get that

$$P_{ij} = kn(\vec{r},t)T(\vec{r},t)\delta_{ij} \equiv P(\vec{r},t)\delta_{ij}$$

We have defined the pressure as a function of space in the above equation. You will note that the equation coincides with the ideal gas equation of state, which is expected.

$$\vec{q} = 0$$

Plugging these expressions into our conservation laws, we get:

$$\frac{\partial}{\partial t}\rho + \rho\frac{\partial}{\partial \vec{r}}\cdot\vec{u} = 0$$

$$\left(\frac{\partial}{\partial t} + u_j\frac{\partial}{\partial r_j}\right)u_i = -\frac{1}{m}\langle\frac{\partial V}{\partial \vec{q}}\rangle - \frac{1}{\rho}\frac{\partial}{\partial r_j}P(\vec{r},t)$$

and

$$\left(\frac{\partial}{\partial \vec{r}}\cdot\vec{u} + \frac{\partial}{\partial t}\right)T - \frac{2T}{3}\frac{\partial}{\partial \vec{r}}\cdot\vec{u} = 0$$

Taken together, these are the *Euler equations* for fluid mechanics. They describe the behavior of a so-called *ideal fluid*, without dissipation. One of the most immediate consequences of the Euler equations we just derived is that they predict the existence of sound waves. To see this, we are going to look at small perturbations around a fluid in equilibrium.

Let's assume our fluid is initially at rest, with a uniform, constant density ρ_0 and pressure P_0. We then introduce a small disturbance, so that the density, pressure, and velocity become:

$$\begin{aligned}\rho(\vec{r},t) &= \rho_0 + \delta\rho(\vec{r},t)\\ P(\vec{r},t) &= P_0 + \delta P(\vec{r},t)\\ \vec{u}(\vec{r},t) &= 0 + \delta\vec{u}(\vec{r},t)\end{aligned}$$

We assume that these perturbations $(\delta\rho, \delta P, \delta\vec{u})$ are very small, which allows us to linearize our fluid equations by dropping any terms that are quadratic or higher in the perturbations (e.g., terms like $\delta\rho\delta\vec{u}$ or $(\delta\vec{u}\cdot\nabla)\delta\vec{u}$).

Let's plug these into our continuity equation:

$$\frac{\partial}{\partial t}(\rho_0 + \delta\rho) + \nabla\cdot((\rho_0+\delta\rho)\delta\vec{u}) = 0$$

Since ρ_0 is constant, its time derivative is zero. We can expand the divergence

$$\nabla \cdot (\rho_0 \, \delta\vec{u} + \delta\rho \, \delta\vec{u}) = 0$$

keeping only the first-order linear terms gives our *linearized continuity equation*:

$$\frac{\partial \delta\rho}{\partial t} + \rho_0 \nabla \cdot \delta\vec{u} = 0 \tag{60}$$

Next, we do the same for the Euler momentum equation, assuming no external potential

$$(\rho_0 + \delta\rho)\left(\frac{\partial \delta\vec{u}}{\partial t} + (\delta\vec{u} \cdot \nabla)\delta\vec{u}\right) = -\nabla(P_0 + \delta P)$$

The convective derivative term $(\delta\vec{u}\cdot\nabla)\delta\vec{u}$ is second-order small, so we drop it. The gradient of the constant pressure P_0 is zero. This leaves the *linearized momentum equation*:

$$\rho_0 \frac{\partial \delta\vec{u}}{\partial t} = -\nabla \delta P \tag{61}$$

Now we have two equations, but three unknowns $(\delta\rho, \delta\vec{u}, \delta P)$. We need an equation of state to close the system by relating the pressure perturbation to the density perturbation. Assuming the sound waves oscillate fast enough that there is no time for heat exchange (an adiabatic process), we can write the pressure as a Taylor expansion around the background density:

$$\delta P = \left(\frac{\partial P}{\partial \rho}\right) \delta\rho \equiv c_s^2 \delta\rho$$

where we have suggestively defined the constant $c_s^2 = (\partial P/\partial\rho)$. Substituting this into our linearized momentum equation gives:

$$\rho_0 \frac{\partial \delta\vec{u}}{\partial t} = -c_s^2 \nabla \delta\rho \tag{62}$$

To finish the job, we take the time derivative of the linearized continuity equation:

$$\frac{\partial^2 \delta\rho}{\partial t^2} + \rho_0 \nabla \cdot \left(\frac{\partial \delta\vec{u}}{\partial t}\right) = 0$$

and we take the divergence of our new linearized momentum equation:

$$\rho_0 \nabla \cdot \left(\frac{\partial \delta\vec{u}}{\partial t}\right) = -c_s^2 \nabla^2 \delta\rho$$

Substituting the second equation into the first precisely eliminates the velocity term, leaving us with:

$$\frac{\partial^2 \delta\rho}{\partial t^2} - c_s^2 \nabla^2 \delta\rho = 0 \tag{63}$$

This is the classic wave equation! It tells us that density perturbations propagate through the fluid as waves with a speed c_s. Of course, real fluids do have dissipation, so what went wrong? Well, we assumed the system was in local equilibrium, and while this makes the collision integral vanish, but it doesn't solve the streaming term in Boltzmann's equation. In fact, if you plug it in to Boltzmann's equation, you can show that

$$\frac{\partial f_1^{(0)}}{\partial t} - \{H_1, f_1^{(0)}\} = \left[\frac{1}{T}\left(\frac{m}{2k_BT}(\vec{v}-\vec{u})^2 - \frac{5}{2}\right)(\vec{v}-\vec{u})\cdot\nabla T \right. \\ \left. + \frac{m}{2k_BT}\left((v_i-u_i)(v_j-u_j) - \frac{1}{3}(\vec{v}-\vec{u})^2\delta_{ij}\right)\left(\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i}\right)\right] f_1^{(0)} \tag{64}$$

which is not necessarily 0. However, we can see that the two remaining terms on the right hand side are proportional to ∇T and $\frac{\partial \vec{u}}{\partial r_i}$, so we can see now that if the variations in temperature and pressure are long wavelength variations, then we are close to having a solution. So let's assume that the correction to the distribution function is small:

$$f_1 = f_1^{(0)} + \delta f_1$$

As a quick aside, the more formal way to do this, called the *Chapman-Enskog expansion*, does this more precisely by defining a small parameter, called the *Knudsen number*, which is the ratio of the mean free path to the length scale of excitations in temperature and momentum, and then expanding the distribution function in this parameter. We won't do this because it's quite difficult to do properly. Instead we will just make to the approximation above. Since our distribution function is no longer a Maxwellian, it will contribute to the collision integral. In particular,

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{coll}} = \int d^3p_2 d^3p_1' d^3p_2' \, \omega(\vec{p}_1', \vec{p}_2'|\vec{p}_1, \vec{p}_2)\left[f_1(\vec{p}_1')f_1(\vec{p}_2') - f_1(\vec{p}_1)f_1(\vec{p}_2)\right] \\ = \int d^3p_2 d^3p_1' d^3p_2' \, \omega(\vec{p}_1', \vec{p}_2'|\vec{p}_1, \vec{p}_2)\left[f_1^{(0)}(\vec{p}_1')\delta f_1(\vec{p}_2') + \delta f(\vec{p}_1')f_1^{(0)}(\vec{p}_2') \right. \\ \left. - f_1^{(0)}(\vec{p}_1)\delta f_1(\vec{p}_2) - \delta f(\vec{p}_1)f_1^{(0)}(\vec{p}_2)\right]$$

In the second equation, we have used the fact that $f_1^{(0)}$ does not contribute to the collision integral, and have dropped terms of order δf_1^2. Now, we in principle could come up with a way to properly treat this expression. However, a much simpler approach that captures the relevant physics, at least at a high level, is the so-called *relaxation time approximation.* The idea is that the collisions are supposed to induce a fast relaxation to equilibrium, and so the contribution of the collision integral can be modeled as being proportional to the deviation from local equilibrium:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = -\frac{\delta f_1}{\tau}$$

where the constant of proportionality is a timescale, called the *relaxation time.* If we put this in the Boltzmann equation, we get that

$$\frac{\partial}{\partial t}(f_1^{(0)} + \delta f_1) - \{H, f_1^{(0)} + \delta f_1\} = -\frac{\delta f_1}{\tau}$$

The contribution of δf_1 on the left hand side turns out to be second order small. If we were to formally expand the distribution function, we would see that the rate of change of the first order perturbation depends on the second order collision term, and so we can drop the δf_1's from the left hand side. Then, we get a simple expression for δf_1:

$$\delta f_1 = -\tau \left[\frac{1}{T} \left(\frac{m}{2k_B T} (\vec{v} - \vec{u})^2 - \frac{5}{2} \right) (\vec{v} - \vec{u}) \cdot \nabla T + \frac{m}{2k_B T} \left((v_i - u_i)(v_j - u_j) - \frac{1}{3} (\vec{v} - \vec{u})^2 \delta_{ij} \right) \left(\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i} \right) \right] f_1^{(0)}$$

We can now use this new distribution function to recompute the heat flux and pressure tensor. We can use the definition of the heat flux,

$$\vec{q} = \frac{1}{2} \rho \langle (\vec{v} - \vec{u})(\vec{v} - \vec{u})^2 \rangle$$

where the expectation is taken against our corrected distribution function. We already know from the previous part that the heat flux when computed against the local equilibrium distribution $f_1^{(0)}$ vanishes. Thus, we only care about the expectation against δf_1.

$$\vec{q} = \frac{1}{2} \rho \int d^3 p \, (\vec{v} - \vec{u})(\vec{v} - \vec{u})^2 \delta f_1$$

Actually performing this integral is just a matter of calculation, and it yields that

$$\vec{q} = -\frac{5}{2} n(\vec{r}, t) \tau k T(\vec{r}, t) \nabla T$$

We identify

$$\kappa = \frac{5}{2} n(\vec{r}, t) \tau k T(\vec{r}, t)$$

as our revised expression for heat conductivity. It turns out, (see Tong for more discussion of this), that the relaxation time is inversely proportional to the density and to the square root of temperature. With these two facts, we can see that we can recover, up to a constant factor, the expression for heat conductivity that we got earlier in the course:

$$\kappa = \frac{1}{2} n k l \langle v \rangle$$

If we now tried to compute the pressure tensor with our new distribution function, we would see that the result is that we would have an additional term:

$$P_{ij} = P(\vec{r}, t) \delta_{ij} + \Pi_{ij}$$

where Π is a new tensor called the *deviatoric stress tensor* of the fluid. I won't explain exactly how the calculation proceeds, as again, its somewhat involved, and not very insightful, but if you would like, you can look at Tong again. Suffice to say, this takes you to (one form of) the famous Navier-Stokes equation:

$$\left(\frac{\partial}{\partial t} + \vec{u} \cdot \nabla \right) \vec{u} = \frac{\vec{F}}{m} - \frac{1}{\rho} \nabla P + \frac{\eta}{\rho} \nabla^2 \vec{u} + \frac{\eta}{3\rho} \nabla (\nabla \cdot \vec{u})$$

This equation was derived in the case where we assume the relaxation time approximation. If you keep terms to higher orders in the Chapman-Enskog expansion, you get higher order corrections to Navier Stokes, called the Burnett equations. As a final comment, to partially motivate why our next topic is so interesting, it is rather remarkable that the behavior of fluids can be captured by simply considering the behavior of three collisional

invariants. We justified this a little bit by discussing how collisional invariants are *slow variables*, but nevertheless it is odd that the slow variables tell us almost all we need to know about the behavior of the fluid. As a food for thought, one could ask: What happens if the variations in temperature and pressure and density are not low frequency? What if they are extremely high frequency and rapidly oscillating, but low amplitude? Do the equations of hydrodynamics still work well? We will explore this idea in greater detail next time, when we discuss projection operators.

9 Lecture 8 - Memory

What better way to end the quarter than with reminiscing?

In this course, we have broadly taken two approaches to non-equilibrium statistical physics. The first approach centered around stochastic processes and the Langevin equation. The second approach started from classical mechanics, from which we derived the Boltzmann equation for transport. When we discussed kinetic theory, you will recall that while we started by working with the distribution function, which in principle contains all the information needed to characterize the system, we eventually switched to working with a few, select quantities, like temperature, pressure, and density.

A highly non obvious and very important fact is that many systems far from equilibrium can be described with the use of a small number of *relevant* variables. Our goal today will be to use the so-called "projection operator formalism" to bridge these two understandings of statistical physics, and show why the Langevin equation can be good approximation given what we know about classical mechanics. A brief word of warning before we begin: the literature on projection operator methods in statistical physics is both dense and messy. There are many different ways to define these projections, which lead to very subtly different mathematics. Given the limited time that we have, I won't go into much details on the various different constructions and their differences, but instead will focus on the big picture of projection operators, and some simple, illustrative examples of how non-Markovianity emerges as a result of coarse graining. The main formalisms are due to the work of three physicists: Sadao Nakajima, Hajime Mori, and Robert Zwanzig, and so the central equation that we will come to is called the Mori-Zwanzig equation, and its quantum generalization is called the Nakajima-Zwanzig equation.

9.1 Hilbert Space and Projections in Statistical Mechanics

The starting point of the projection operator formalism is to consider the set of all observables of a system to be a Hilbert space. To be more precise, we think of the state of a system as being described by a coordinate in its phase space. An observable is a function of the state, or equivalently, a function of the phase space coordinates. We think of these functions as forming a Hilbert space, which is (roughly speaking) a vector space equipped with an inner product. We will denote this inner product with $(\cdot, \cdot)$. The exact form of the inner product can be chosen in a number of ways, but one common choice is the following:

$$(A, B) = \int d\Gamma A(\Gamma) B(\Gamma) \rho_{eq}(\Gamma)$$

where ρ_{eq} denotes the equilibrium probability density. Given such a vector space, we can define a *projection* operator P that takes a vector, X as input and returns the component of X along some other vector A, in the following way:

$$PX = (A, A)^{-1}(X, A)A$$

If we want to project onto a subspace spanned by a *set* of observables, $\{A_i\}$, we can generalize the above formula in the following way:

$$PX = \sum_{i,j} (A_j, A_i)^{-1} (X, A_i) A_j$$

The idea now is that we can employ these projection operators to describe the dynamics of a particular subspace of relevant observables (think temperature, pressure, or density). We define the complementary projection $Q \equiv 1 - P$. The projection operators have the following properties:

$$\begin{aligned} P^2 &= P \\ Q^2 &= Q \\ PQ = QP &= 0 \\ P &= P^\dagger \\ Q &= Q^\dagger \end{aligned}$$

The first two relations are easily understood, given the nature of a projection. A projection extracts the component of the input that lies within the subspace that we are projecting onto. If the input vector lies entirely in that subspace, then the projection leaves it unchanged. Thus, we see that projecting N times with the same projection operator is the same as just projecting once. The third relation is a consequence of the fact that P and Q, by construction, project onto orthogonal subspaces. The last two relations just express the idea that the projectors are Hermitian with respect to the inner product for the space. This whole idea can be rather abstract, so I think it helps to have a few concrete examples in mind. As a first example, imagine we have a function, defined over some length of time, $t \in [0, t_{\max}]$. For example, imagine we have a fluid, and we are describing its temperature at some point as a function of time. If we are measuring this experimentally, say with a thermal camera, it has a finite frame rate, and so we can't resolve temporal variations with infinite resolution. That is, if we were to look at the temporal Fourier transform of the temperature, there is a frequency $n_{\max}$ beyond which we can't resolve the variations in temperature. One could model this via a projection operator that is essentially a low-pass filter.

$$Pf = \frac{1}{t_{\max}} \sum_{n=0}^{n_{\max}} \int_{-\infty}^{\infty} dt \ \exp(-2\pi i n / t_{\max}) f(t)$$

Another, perhaps simpler, example is the case of partitioning. Let's say that we have N billiards balls on a table. The phase space for such a system has $2N$ coordinates, representing the position and momentum of the balls: $X = (q_1, p_1, ..., q_{2N}, p_{2N})$. Let's say that we, for some reason, can only see the first N balls. We can represent this has a projection matrix, $P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes I^{N \times N}$. Think for a moment about what the dynamics of such a system might look like. The intuition, though we of course will need to be more formal about this, is that there are two components to the dynamics. The first component is just the usual dynamics, when the observable balls are moving without interacting with the invisible ones. This looks like normal motion. The second component is dynamics of the observable balls due to their interaction with the invisible balls. We might see them abruptly change direction or something like that.

9.2 Some Useful Formulae from Hamiltonian Mechanics

In classical mechanics, we learn that the time evolution of any dynamical variable, or function of the phase space coordinates, is given by

$$\frac{d}{dt}A(q,p,t) = \frac{\partial}{\partial t}A(q,p,t) + \{A,H\}$$

Some of you may be more well acquainted with the quantum mechanical version of this statement:

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} - \frac{i}{\hbar}[A,H]$$

For observables which do not explicitly depend on time, this equation is often written in terms of the *Liouville operator*, $L = \{H, \cdot\}$, so

$$\frac{dA}{dt} = LA$$

One can show that the Liouville operator, as defined above, is anti-Hermitian. Formally, one can then write the solution to this equation as

$$A(t) = \exp(Lt)A(0)$$

This operator, $\exp(Lt)$, is called the *propagator* of the system, as it takes the initial state, and propagates it forward in time to the state at time t. Why do we care about doing all this? The propagator equation above depends on the full state of the system, since the Liouville operator depends on the system's Hamiltonian, which in turn depends on the full state. We know, empirically, however, that we can generally understand the dynamics of various macroscopic observables in a self-determined way. For example, Navier-Stokes describes the transport of density, temperature, and pressure in a fluid, without any need to rely on the microscopic state of the system. Projection operators, then, give us a way to determine this

9.3 Two variable system

The simplest example that illustrates the mechanism at play is a two variable system. Let's say that our system is composed of two observables, a_1 and a_2. We can then represent its evolution via the matrix equation

$$\frac{d}{dt}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Say now that we only care about the evolution of the variable a_1. Then, we can eliminate the dependence on $a_2(t)$ from its equation of motion by solving for $a_2(t)$:

$$\begin{aligned}
\frac{d}{dt}a_2(t) &= L_{21}a_1(t) + L_{22}a_2(t) \\
\frac{d}{dt}a_2(t) - L_{22}a_2(t) &= L_{21}a_1(t) \\
e^{-L_{22}t}\frac{d}{dt}a_2(t) - e^{-L_{22}t}L_{22}a_2(t) &= e^{-L_{22}t}L_{21}a_1(t) \\
\frac{d}{dt}e^{-L_{22}t}a_2(t) &= e^{-L_{22}t}L_{21}a_1(t) \\
\int_0^t ds\left[\frac{d}{ds}e^{-L_{22}s}a_2(s)\right] &= \int_0^t ds e^{-L_{22}s}L_{21}a_1(s) \\
e^{-L_{22}t}a_2(t) - e^{-L_{22}\cdot 0}a_2(0) &= \int_0^t ds e^{-L_{22}s}L_{21}a_1(s)
\end{aligned}$$

which gives the result

$$\begin{aligned}
a_2(t) &= e^{L_{22}t}\int_0^t ds e^{-L_{22}s}L_{21}a_1(s) + e^{L_{22}t}a_2(0) \\
a_2(t) &= \int_0^t ds e^{L_{22}(t-s)}L_{21}a_1(s) + e^{L_{22}t}a_2(0)
\end{aligned}$$

So, we can now substitute this in to the equation for a_1:

$$\begin{aligned}
\frac{d}{dt}a_1(t) &= L_{11}a_1(t) + L_{12}a_2(t) \\
&= L_{11}a_1(t) + L_{12}\int_0^t ds e^{L_{22}(t-s)}L_{21}a_1(s) + L_{12}e^{L_{22}t}a_2(0)
\end{aligned}$$

We can see here that eliminating one of the variables leads to an explicitly non-Markovian equation of motion for the other.

9.4 Mori-Zwanzig Equation

The Mori-Zwanzig equation is essentially just the result of applying an operator identity, called the *Dyson identity*, to the propagator. The idea is as follows. We start with the equation

$$A(t) = \exp(Lt)A(0)$$

which implies that

$$\frac{dA}{dt} = \exp(Lt)\cdot LA(0)$$

One then inserts the identity, in the form $I = Q + P$, into the equation:

$$\frac{dA}{dt} = \exp(Lt)\cdot(Q+P)LA(0) = \exp(Lt)QLA(0) + \exp(Lt)\cdot PLA(0)$$

In the second term, we can insert the definition of the projector:

$$\exp(Lt) \cdot PLA(0) = \exp(Lt) \cdot \frac{(LA, A)}{(A, A)} A$$

where in the second step, I dropped the argument $A(0)$. The value

$$\Omega \equiv \frac{(LA, A)}{(A, A)}$$

The first term is treated using the following identity, called the *Dyson identity*.

$$e^{-(A+B)t} = e^{-At} - \int_0^t ds e^{-As} B e^{-(A+B)(t-s)}$$

If you want to prove this identity, you can do so by differentiating both sides with respect to time, and using Leibniz's rule to exchange the order of integration and differentiation on the right hand side.

In our case, we use $A = QL$ and $B = PL$, and get

$$e^{-(P+Q)Lt} = e^{-QLt} - \int_0^t ds e^{-QLs} PL e^{-L(t-s)}$$

We then take the adjoint of this equation, using the fact that the Liouville operator is anti Hermitian,

$$e^{Lt} = e^{QLt} + \int_0^t ds e^{L(t-s)} PL e^{QLs}$$

Plugging this into the first term, we get

$$\begin{aligned} \exp(Lt) QLA(0) &= \left[e^{QLt} + \int_0^t ds e^{L(t-s)} PL e^{QLs} \right] QLA(0) \\ &= QL e^{QLt} A + \int_0^t ds e^{L(t-s)} PL e^{QLs} QLA \end{aligned}$$

We define the quantity $F(t)$:

$$F(t) \equiv QL e^{QLt} A(0)$$

We also insert the definition of the projection operator, P, in the second term to give

$$PL e^{QLs} QLA = \frac{(L e^{QLt} QLA, A)}{(A, A)} A$$

Defining the memory kernel,

$$K(t) \equiv -\frac{(L e^{QLt} QLA, A)}{(A, A)} = -\frac{(LF(t), A)}{(A, A)}$$

This gives the Mori-Zwanzig equation, also called the *generalized Langevin equation*:

$$\frac{dA}{dt} = \Omega e^{Lt} A(0) - \int_0^t ds e^{QLs} PLe^{L(t-s)} QLA(0) + F(t)$$
$$= \boxed{\Omega A(t) - \int_0^t ds\ K(s)A(t-s) + F(t)}$$

Now, in reality all that we have done is rearranged some terms and defined some new quantities, so it's very much unclear that what we have done is useful. Part of the purpose can be seen by considering how one can interpret each term. The first term in the MZ equation is the contribution to the rate of change of the observable as due to its current value, while the second isolates the effect of its prior history. The last term depends entirely on the initial condition of the unresolved subspace, and so it is often interpreted as a sort of noise. We will see exactly how this comes into play in a moment.

Before we do that though, there is an important result that one can derive from the above definitions. You may get the feeling at the moment that all of this operator algebraic manipulation has provided very little of real utility, since in order to do anything useful with it, we need to know how to perform these projections and how to solve the resulting equations. In particular, the noise term relies entirely on the unresolved variables, and so we need to know how those were initially distributed in order to do anything useful. However, the utility of this formalism comes from the fact that we can make novel statistical statements about these quantities. Starting from the definition of the memory kernel,

$$K(s) = -\frac{(LF(s), A)}{(A, A)} = \frac{(F(s), LA)}{(A, A)} = \frac{(QF(s), LA)}{(A, A)} = \frac{(F(s), QLA)}{(A, A)} = \frac{(F(s), F(0))}{(A, A)}$$

The first step is the definition of the memory kernel. The second uses the anti-Hermitian nature of the Liouville operator. The third inserts a factor of Q in front of $F(s)$, which we can do since $F(s)$ already lives entirely in the unresolved subspace, so projecting again does not change it. The fourth step uses the fact that Q is Hermitian, and the final step uses the definition of F. So what is going on? This result is called the *second fluctuation dissipation theorem*, and it tells us the strength of the memory term is directly proportional to the autocorrelation of the noise. If the memory kernel decays very fast, the noise we need looks like white noise. If the memory kernel decays very slowly, then the noise term also displays some strong temporal correlation.

One important detail that may be non obvious is that the formalism, as developed so far, is only really useful for systems near equilibrium. This is because we have assumed in our derivation that the Hamiltonian, and thus the Liouvillian, is time-independent. If the phase space distribution is stationary, then this is the case, but if it is not, then we need some time dependence. The formalism has been generalized to this case. See the book by Grabert.

The immediate question that you might have is how we might actually compute the memory kernel or the streaming term, since generally it can be difficult to obtain analytic expressions for things involving the Liouville operator. We will now derive one useful formula for doing this. Starting from the Mori-Zwanzig equation, we can take the inner product of both sides with $A(0)$:

$$\left(\frac{dA(t)}{dt}, A(0)\right) = \left(\Omega A(t) - \int_0^t ds\ K(s)A(t-s) + F(t), A(0)\right)$$
$$\frac{d}{dt}\Big(A(t), A(0)\Big) = \Omega\Big(A(t), A(0)\Big) - \int_0^t ds\ K(s)\Big(A(t-s), A(0)\Big) + \Big(F(t), A(0)\Big)$$

If the inner product is the equilibrium average, we see that $\Big(A(t), A(0)\Big)$ is simply the equilibrium time correlation function that we encountered earlier in the course. If we denote this as $C(t) \equiv \Big(A(t), A(0)\Big)$, we get

$$\frac{d}{dt}C(t) = \Omega C(t) - \int_0^t ds\ K(s)C(t-s)$$

since $\Big(F(t), A(0)\Big) = 0$. To see why, we just plug in the form of $F(t)$:

$$\Big(F(t), A(0)\Big) = \Big(e^{QLt}A(0), A(0)\Big) = -\Big(e^{QLt}A(0), QLA(0)\Big)$$

but

$$QLA(0) = (1-P)A(0) = A - \frac{(A,A)}{(A,A)}A = 0$$

so we get the relationship that

$$\Omega = \frac{d}{dt}C(t)\bigg|_{t=0}$$

A more complicated derivation yields

$$\tilde{K}(s) = \Omega - s + \frac{1}{\tilde{C}(s)}$$

where $\tilde{C}(s)$ is the Laplace transform of $C(s)$ and $\tilde{K}(s)$ is the Laplace transform of $C(s)$. I don't expect you all to remember how to take a Laplace transform, but I am assuming that you have heard the term before, maybe in your high school math courses? If not, it is just an integral transform, much like the Fourier transform. The point being we can express the memory kernel in terms of time correlation functions of the corresponding observable, which in many cases is either computable or measurable.

9.5 The Markovian Approximation

In general, the Mori-Zwanzig equation is hard to solve. The streaming term is not too hard to deal with, but the convolution term and the noise term are more complicated. We will consider how to deal with the memory term first. In many cases of interest, like for example hydrodynamics, the relevant variables are chosen such that they evolve on a slow time scale as compared to the unresolved variables. In particular, because of this, we expand the middle term in powers of LA. To keep things a little cleaner, let's say that $LA = \mathcal{O}(\lambda)$, where λ is some small parameter. We can expand the MZ equation, keeping only terms up to second order in λ. Ω is first order in λ, and the memory kernel, K is $\mathcal{O}(\lambda^2)$, since

$$K(t) \equiv -\frac{(Le^{QLt}QLA, A)}{(A,A)} = \frac{(e^{QLt}QLA, LA)}{(A,A)}$$

Simultaneously, we can expand $A(t-s)$ in time

$$A(t-s) \approx A(t) - s\dot{A}(t) + \mathcal{O}(\lambda^2)$$

The second term in this expansion is first order in LA, since $\dot{A} = LA$, and so we are left with the realization that we may truncate this series after the first term. This gives the following, approximate equation:

$$\frac{dA}{dt} = \Omega A - \left[\int_0^t ds\ K(s)\right] A(t) + F(t)$$

Finally, we can invoke the second fluctuation dissipation relation to argue that, in the case where the relevant variables evolve slowly as compared to the orthogonal space, the memory kernel must decay quickly, and so we are able to raise the bound on the integral to infinity.

$$\frac{dA}{dt} = \Omega A - \left[\int_0^\infty ds\ K(s)\right] A(t) + F(t)$$

This is called the *Markovian approximation* to the generalized Langevin equation.

A note about units: At a number of points during this lecture, we have either added quantities with seemingly inconsistent units, or Taylor expanded in dimensionful quantities. In general, this is because of some details that are being brushed under the rug to keep the presentation here consistent with the broader literature. For example, at the very start, we proposed that there was a Hilbert space of dynamical variables. This is problematic because different functions of the phase space variables have different units. One can resolve this by non-dimensionalizing everything. In the case of Taylor expanding in LA, we can see that this has whatever units A has, divided by time. Thus, the terms in the Taylor expansion are dimensionally consistent.

9.6 Example: The Langevin Equation Revisited

It is a little bit useful to see how some of these ideas can be revisited in the case of Brownian motion. As a first sanity check, let's see if we can recover the Langevin equation from the GLE. In that case, the relevant observable is the velocity of the Brownian particle, $v(t)$. Its correlation function, as one can compute directly from the standard Langevin equation, is $C(t) = e^{-\gamma t/m}$. *You may notice that this is off by a factor of v_0^2 from what you expect from the Langevin equation. This because $C(t)$ includes a normalization factor to be consistent with our earlier definition.* This immediately presents a problem, since we also know that time correlation functions should be even functions of time. The resolution to this is that the formula for $C(t)$ given above is true for $t > 0$ and reversed for $t < 0$. Unfortunately, this gives a cuspy function where the derivative is ill defined at $t = 0$. For reasons that I won't get into, it turns out that $\Omega = 0$ in this case.

What about the memory kernel? The Laplace transform of $C(t) = e^{-\frac{\gamma}{m}t}$ is $\tilde{C}(s) = -\frac{m}{\gamma + ms}$, so

$$\tilde{K}(s) = -s + \frac{ms + \gamma}{m} = \frac{\gamma}{m}$$

The inverse Laplace transform of a constant is a delta function, so we get

$$K(t) = \frac{\gamma}{m}\delta(t)$$

which is exactly what we expect for Brownian motion! Plugging these results into the GLE, we get

$$\frac{dv}{dt} = -\frac{\gamma}{m}v(t) + F(t)$$

which is exactly our old Langevin equation. All that remains is to argue that $F(t)$ shares the same characteristics as our white noise force, which we can do by simply applying the second fluctuation dissipation theorem, to see that

$$(F(t), F(0)) = v_0^2 \cdot \delta(t)$$

And so we recover the usual Langevin equation, only using the assumption that the correlation function decays exponentially. This also ought to signal to you why the Langevin equation is so, so fundamental. We see that for any observable with exponentially decaying correlations, the Langevin equation is the correct descriptor.

As a final example, we will consider the case of Brownian motion in a harmonic potential. The equations of motion are

$$\frac{dx}{dt} = \frac{p}{m}$$
$$\frac{dp}{dt} = -m\omega^2 x - \frac{\gamma p}{m} + \delta F(t)$$

Let's assume that at $t = -\infty$, the momentum was 0. Then, we can solve the second equation by use of an integrating factor

$$\begin{aligned}
\frac{d}{dt}p(t) + \frac{\gamma}{m}p(t) &= -m\omega^2 x(t) + \delta F(t) \\
e^{\frac{\gamma t}{m}}\left[\frac{d}{dt}p(t) + \frac{\gamma}{m}p(t)\right] &= e^{\frac{\gamma t}{m}}\left[-m\omega^2 x(t) + \delta F(t)\right] \\
\frac{d}{dt}\left(e^{\frac{\gamma t}{m}}p(t)\right) &= e^{\frac{\gamma t}{m}}\left[-m\omega^2 x(t) + \delta F(t)\right] \\
e^{\frac{\gamma t}{m}}p(t) &= \int_{-\infty}^{t} ds\, e^{\frac{\gamma s}{m}}\left[-m\omega^2 x(s) + \delta F(s)\right] \\
p(t) &= \int_{-\infty}^{t} ds\, e^{\frac{\gamma(s-t)}{m}}\left[-m\omega^2 x(s) + \delta F(s)\right] \\
&= \int_{-\infty}^{t} ds\, e^{\frac{-\gamma(t-s)}{m}}\left[-m\omega^2 x(s) + \delta F(s)\right] \\
&= \int_{0}^{\infty} ds\, e^{\frac{-\gamma s}{m}}\left[-m\omega^2 x(t-s) + \delta F(t-s)\right]
\end{aligned}$$

If we plug this into the differential equation for position, we get that

$$\frac{dx}{dt} = -\int_0^\infty ds\, e^{\frac{-\gamma s}{m}}\omega^2 x(t-s) + \frac{1}{m}\int_0^\infty ds\, e^{\frac{-\gamma s}{m}}\delta F(t-s)$$

We can define

$$K(s) = \omega^2 e^{\frac{-\gamma s}{m}}$$

and

$$F(t) = \frac{1}{m}\int_0^\infty ds\, e^{\frac{-\gamma s}{m}}\delta F(t-s)$$

to get the non-Markovian Langevin equation,

$$\frac{dx}{dt} = -\int_0^\infty ds\, K(s)x(t-s) + F(t)$$

This example also illustrates that in addition to it being the case that elimination of degrees of freedom leads to non-Markovian behavior, the converse is also possible. If we are to instead *add* a variable to this non Markovian equation for position, we can then convert it into a system of Markovian equations. This procedure, of adding variables to make an equation Markovian is called *Markovian embedding*, and can be applied to processes where the correlations decay exponentially in time.

The examples that we have treated thus far have been really simple, and so it may not be clear to you why this formalism is all that useful. Unfortunately, due to time constraints, we won't be able to go into more examples, but if you are curious, here are some resources that you can look at to see other ways in which the MZ formalism can be applied:

- *Nonequilibrium Statistical Mechanics*, Robert Zwanzig, Section 1.5 and Chapter 8: These sections cover the examples of Brownian motion in a harmonic oscillator heat bath, self-diffusion, and some discussion of hydrodynamics.
- *Projection operators in statistical mechanics: a pedagogical approach* by Michael te Vrught and Raphael Wittkowski: Sketches out a derivation of the Bloch equations and dynamic density functional theory using projection operators. A slightly different, but much more fleshed out derivation of the Bloch equations is presented in Zwanzig's book in section 6.4.