Matthew Schwartz
Statistical Mechanics, Spring 2021

Lecture 3: Equilibrium

1 Introduction

Now we have a little bit of sense of how things simplify when large numbers of events are sampled from a probability distribution. The next thing to do is apply this simplification to general physical systems. The key observation, which allows statistical mechanics to be useful at all, is that systems equilibrate. This means that at some point the macroscopic properties of the system stop changing. Note that equilibrium does not mean the system is static: a gas at fixed temperature still has moving molecules, but its macroscopic properties, pressure, temperature, etc. are not changing. From a microscopic perspective, the probability distribution of the states does not change. In this lecture, we will show that probabilities eventually stop changing.

It is not hard to build intuition for equilibration of probabilities. For example, take a deck of cards and pick two consecutive cards from somewhere in the middle. What is the probability that the cards have the same suit after the deck has been shuffled t times? When you first open a new deck, it is all in order, so the chance that the two cards have the same suit pretty high, $P(0)\approx 1$. Then you shuffle it. After shuffling once, the probability is of two suited cards in a row is lower than before, $P(t=1)\lesssim 1$ but probably still pretty high since one shuffle doesn't mix them much. Eventually, after shuffling a bunch of times, the probability of finding two consecutive suited cards is going to stop changing ($\lim_{t\to\infty}P(t)=\frac{12}{52}$). Note that the cards change with each shuffle, but the probabilities don't: if the deck shuffles continuously, then the identity of the top card is not constant, but the probability that the top card is the ace of spades is constant, $P=\frac{1}{52}$.

One of the most important properties of equilibrium is that in equilibrium, all possible states are equally likely. This is known as the *postulate of equal a priori probabilities*. For example, with a shuffled deck, the chance of the top card being any of the 52 cards is the same $P=\frac{1}{52}$.

In this lecture we introduce a number of important concepts related to equilibration

- Chaos: the state of a system is uncontrollably sensitive to small perturbations.
- Molecular chaos: correlations among states are lost over time.
- Coarse-graining: averaging over nearby position/momenta.
- Ergodicity: the time average of a system is the same as the average over possible states.

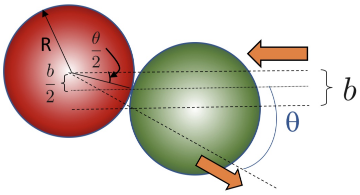
The example we focus on most in this lecture, and for much of the course, is that of an **ideal gas**. An ideal gas is one for which all the interactions are short-ranged and collisions are elastic. In the simplest ideal gas, the gas molecules have no internal structure. The state of the ideal gas is specified classically by giving the positions $\vec{q}_i(t)$ and velocities $\vec{v}_i(t)$ (or momenta $\vec{p}_i(t)=m\vec{v}_i(t)$) of the particles $i=1...N$ at any time t. The set of allowed $\vec{q}_i$ and $\vec{p}_i$ for the N particles is called the **phase space** of the gas. The phase space is $6N$ dimensional. The equations of motion of the gas determine the trajectory through phase space. Thinking of the time-evolution of a system as a trajectory in phase space provides a useful language for discussing (and proving) results about classical systems.

2 Chaos

A key to understanding equilibration (and statistical mechanics!) is to appreciate why we go from knowing exactly what is going on in a system to accepting that we can only discuss the probability of what is going on. This is the transition from mechanics (classical, where you know positions and velocities or quantum where you know the wavefunction) to statistical mechanics.

The easiest way to see why we *must* go to a probabilistic treatment is that, it is literally *impossible* to keep track of every particle in the system. In fact, systems with large numbers of degrees of freedom are always **chaotic**: they are uncontrollably sensitive to infinitesimal inaccuracies of the specification of the system. Chaos is sometimes called the "butterfly effect", since a butterfly flapping its wings in Australia can affect the weather in Boston. We can understand the basic observation about chaos with a simple example.

Let's treat the molecules in a gas as hard sphere particles of radius R. If you follow one sphere, it will bounce off another sphere after travelling, on average, a distance ℓ (the mean free path). The angle θ it deflects will depend on the impact parameter b, defined as the distance between the sphere's centers perpendicular to the direction of motion



(1)

By working out the geometry as in the figure, we see that b, R and θ are related by:

$$\frac{b}{2} = R\sin\frac{\theta}{2} \tag{2}$$

Let us follow one ball. It first collides with an impact parameter b_1 deflecting at an angle θ_1, then hits another ball, with impact parameter b_2 at angle θ_2 and so on. We are interested in how the trajectory of this molecule changes upon a really weak force. If the force changes the impact parameter of the first collision by some small Δb_1, then the after the first collision, the scattering angle will change by $\Delta\theta_1$ where, by Taylor expanding around θ_1 we get

$$\frac{b+\Delta b_1}{2} = R\sin\left(\frac{\theta_1+\Delta\theta_1}{2}\right) \approx R\sin\left(\frac{\theta_1}{2}\right) + R\frac{\Delta\theta_1}{2}\cos\frac{\theta_1}{2} + \cdots \tag{3}$$

The situation is shown in Fig. 1:

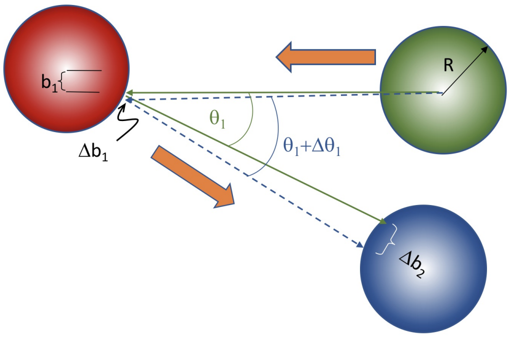


Figure 1. The green ball hits the red ball at an impact parameter b_1, scatters at an angle θ_1, then hits the blue ball at impact parameter b_2. A small change Δb_1 in b_1 leads to a change $\Delta\theta_1$ in θ_1 and a change Δb_2 in b_2.

Assuming $\frac{\Delta b_1}{R} \ll 1$ we can then solve Eq. (3) for $\Delta\theta_1$ using Eq. (2), giving $\Delta\theta_1 \approx \frac{2\Delta b_1}{R\cos\theta_1}$. Let us not assume anything special about θ_1, so that $\cos\theta_1$ is not unusually big or small, and therefore $\Delta\theta_1 \approx \frac{\Delta b_1}{R}$ up to some number of order 1. This makes sense: if Δb_1 is small, then $\Delta\theta_1$ is small too. However, after the first collision, the sphere moves a distance ℓ to the next collision. Then the impact parameter for the second collision changes by $\Delta b_2 = \ell\Delta\theta_1$. This implies $\Delta\theta_2 \approx \frac{\Delta b_2}{R} = \frac{\ell}{R}\Delta\theta_1$.

In this way, the angle change has grown by a factor of $\frac{\ell}{R}$. After the next collision, we would similarly find $\Delta\theta_3 = \frac{\ell}{R}\Delta\theta_2 = \left(\frac{\ell}{R}\right)^2 \Delta\theta_1$ and so on. Thus after N collisions,

$$\Delta\theta_N \approx \left(\frac{\ell}{R}\right)^N \Delta\theta_1 \tag{4}$$

Even if $\Delta\theta_1$ is very very small this factor of $\left(\frac{\ell}{R}\right)^N$ quickly becomes very large. For a gas at room temperature, $\ell \approx 10^{-7}$ m and $R \approx 10^{-10}$ m. Then $\frac{\ell}{R} \approx 10^3$. Thus after just a few collisions, this $\left(\frac{\ell}{R}\right)^N$ factor can make small effects very very big.

For a concrete example, let's estimate the effect on the trajectory of a gas molecule from waving your arm from far away. Displacing your arm by the distance Δr will change the gravitational force by $\Delta F = \frac{\partial F}{\partial r}\Delta r$. So,

$$\Delta F = \frac{d}{dr}\left(G_N \frac{m_1 m_2}{r^2}\right)\Delta r = -2G_N \frac{m_{\text{atom}} m_{\text{arm}}}{r^3}\Delta r \tag{5}$$

Say $m_{\text{atom}} = 10^{-27}\text{kg}$, $m_{\text{arm}} = 1\text{kg}$ and $\Delta r = 1\text{m}$ and you are standing $r = 10\text{km}$ away. Then

$$\Delta F \approx \left(10^{-11}\frac{Nm^2}{\text{kg}^2}\cdot\right)\frac{(10^{-27}\text{kg})(1\text{kg})}{(10^4 m)^3}(1m) = 10^{-50}N \tag{6}$$

This will cause an acceleration of $a = \frac{\Delta F}{m_{\text{atom}}} = 10^{-23}\frac{m}{s^2}$ on the atom. Over a time $\tau \sim 10^{-9}s$ between two collisions, this hand-waving has moved the atom by around $\Delta b_1 = a\tau^2 \approx 10^{-32}m$, much much less than the size of an atom, and so $\Delta\theta_1 \sim \frac{\Delta b_1}{R} \approx 10^{-22}$. This is a tiny tiny angle. However, after N collisions

$$\Delta\theta_N = 10^{-22}\left(\frac{\ell}{R}\right)^N = 10^{-22}(10^3)^N \tag{7}$$

Thus after only 8 collisions, the change in angle is of order 1! This is chaos.

Suppose instead of a hand waving on earth, we consider the wing-waving of a flea across the universe, at a distance of $r \approx 10^{27}m$ with $\Delta r = 10^{-6}\ m$. Once the gravitational waves from the flea arrive to our gas, the would change the initial deflection to $\Delta\theta_1 \approx 10^{-97}$. Even for such a tiny tiny angle, it would take only 33 collisions (10^{-7} seconds) for the trajectory of the atom to change completely. After 1 second, every molecule has a vastly different position and velocity from what it would have had if the flea across the universe had not waved its hand.

A chaotic system is one in which the late time behavior is exponentially sensitive to initial conditions: changing the initial condition by a little bit has an enormous effect on the outcome. A flea across the universe changing the trajectory of an atom by order one after one microsecond illustrates this point. Because of chaos, we can never hope to know the state of a gas exactly. There is simply no physical limit in which a state can be isolated and well defined. We *must* resort to probabilities if we are to make any physical predictions for gases.

In addition, even if we pretend a system is completely isolated – turn off gravity and fleas, etc – the system is still chaotic due to (classical) uncertainty. Even if there is some exponentially small uncertainty on the initial condition, the final state after long enough time will be completely unknown. Using the same numbers as above, if we specify the initial condition to one part in 10^{97} after $1\,\mu s$ the state has order 1 uncertainty. In quantum mechanics, the relevant uncertainty on the initial state is not the Heisenberg uncertainty (on knowing position given momentum), but on how well we can actually know the initial state wavefunction. Again, there is no hope of solving the time-evolution exactly.

3 Maxwell and Molecular Chaos

Next, we want to show that systems tend towards flat probability distributions. This was first understood by Maxwell, who developed his understanding through the kinetic theory of gases.

3.1 Equilibration of molecular velocities

Say we have a gas with different types of molecules in it, of different masses, m_1, m_2, etc. If the gas is in equilibrium, then there will be well-defined probabilities for the velocities of the different molecule types. What can we say about these probabilities?

Pick randomly one molecule of mass m_1 and one of mass m_2. Say these two molecules come in to hit each other with incoming velocities $\vec{v}_1$ and $\vec{v}_2$, respectively. If there is no preferred direction in the system, then the velocities are equally likely to point in any direction. Thus the dot product $\vec{v}_1 \cdot \vec{v}_2 = |\vec{v}_1||\vec{v}_2|\cos\theta$ is equally like to be positive or negative and therefore the expectation value of the dot product of the incoming velocities must be zero:

$$\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = \langle |\vec{v}_1| \rangle \langle |\vec{v}_2| \rangle \langle \cos\theta \rangle = \langle |\vec{v}_1| \rangle \langle |\vec{v}_2| \rangle \frac{1}{\pi} \int_0^\pi d\theta \cos = 0 \tag{8}$$

That is, two random initial velocities are **uncorrelated**.

For collisions, it is often helpful to work in the center of mass frame. The velocity of the center of mass is

$$\vec{v}_{\text{cm}} = \frac{1}{m_1 + m_2} (m_1 \vec{v}_1 + m_2 \vec{v}_2) \tag{9}$$

Note that $\vec{v}_{\text{cm}}$ does not change as a result of the collision, by momentum conservation. We can shift to the center of mass frame by $\vec{v}_1 \rightarrow \vec{v}_1 - \vec{v}_{\text{cm}}$ and $\vec{v}_2 \rightarrow \vec{v}_2 - \vec{v}_{\text{cm}}$. The relative velocity:

$$\overrightarrow{\Delta v} = \vec{v}_1 - \vec{v}_2 \tag{10}$$

is unchanged by this shift. Since $\vec{v}_1$ and $\vec{v}_2$ pointed in random directions, $\overrightarrow{\Delta v}$ also points in a random direction as does $\vec{v}_{\text{cm}}$. Moreover: $\overrightarrow{\Delta v}$ and $\vec{v}_{\text{cm}}$ are totally uncorrelated: one can have any relative velocity with any $\vec{v}_{\text{cm}}$. Thus,

$$\langle \overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} \rangle = 0 \tag{11}$$

Next, write

$$\overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} = \frac{1}{m_1 + m_2} (\vec{v}_1 - \vec{v}_2) \cdot (m_1 \vec{v}_1 + m_2 \vec{v}_2) \tag{12}$$

$$= \frac{m_1 v_1^2 - m_2 v_2^2 + (m_2 - m_1) \vec{v}_1 \cdot \vec{v}_2}{m_1 + m_2} \tag{13}$$

Now take the average value of the terms in this equation over all possible choices for molecules 1 and 2 with masses m_1 and m_2. Since we already established that $\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = 0$ and $\langle \overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} \rangle = 0$ we must therefore have

$$\langle m_1 \vec{v}_1^2 \rangle = \langle m_2 \vec{v}_2^2 \rangle \tag{14}$$

Dividing by two, we conclude that the **average kinetic energy** of any molecular species in the gas is the same. This calculation of Maxwell's was one of the first theoretical results demonstrating progress towards equilibrium.

I think Maxwell's calculation is remarkable because it is so simple, yet so profound. All we used was that the velocities point in random directions, and we learned something non-trivial about the magnitudes of the velocities. Not only do heavier molecules move slower, on average, but the average kinetic energy for any molecules is a universal quantity determined by the state of the gas, independent of the mass.

3.2 Molecular chaos

Note that Maxwell's argument depended on $\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = 0$, which we justified with the logic that there is no preferred direction in the system, so the angle between two randomly chosen velocities should be evenly distributed. This justification is very reasonable, but it is still an assumption:

- The **assumption of molecular chaos**: velocities of colliding particles are independent of each other, and independent of the position of the collision.

This assumption is an excellent approximation for most physical systems. It is however, never exactly true.

To turn the assumption into an equation, it is helpful to refer to the state of the entire gas at time t as a point $(\vec{q}_i, \vec{p}_i)$ in phase space. Note that phase space is enormous; it is $6N \approx 10^{24}$ dimensional. We are interested in the probability $P(\vec{q}_i, \vec{p}_i, t)$ of finding all the particles with given positions and momenta at the same time t; i.e. if there are N molecules in the gas then P is a $6N+1$ dimensional function: $P = P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t)$. The assumption of molecular chaos lets us write

$$P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t) = P_1(\vec{q}_1, \vec{p}_1, t) \times \cdots \times P_1(\vec{q}_N, \vec{p}_N, t) = \prod_j P_1(\vec{q}_j, \vec{p}_j, t) \tag{15}$$

for a simpler function $P_1(\vec{q}, \vec{p}, t)$ of just 7 variables. Now, the expectation value of observables, like the average velocity-squared, can be computed just with 7-dimensional function P_1 rather than the 10^{23} dimensional function P:

$$\langle \vec{v}_k^2 \rangle(t) = \int d^{3N}q\, d^{3N}q\, \frac{\vec{p}_k^2}{m^2} P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t) = \int d^3q_k\, d^3p_k\, \frac{\vec{p}_k^2}{m^2} P_1(\vec{q}_k, \vec{p}_k, t) \tag{16}$$

In the last step, we have used $\int d^3q d^3p P_1(p, q) = 1$ for all the $p's$ and q's other than p_k, q_k. So the velocity distribution is the same for any particle, and determined by a single 7-dimensional function.

To see the subtlety in the assumption of molecular chaos, return to the example of two colliding particles of masses m_1 and m_2 with uncorrelated initial velocities $\vec{v}_1$ and $\vec{v}_2$. Are the outgoing velocities $\vec{v}_1'$ and $\vec{v}_2'$ uncorrelated as well? If you think about it for a moment, it is easy to see that the answer must be no. For example, say $m_1 \gg m_2$, like a truck hitting a bicycle. After the collision, the truck and the bicycle will be going close to the truck's initial direction. So the angle θ measured with respect to the incoming truck's direction will likely be close to zero and therefore $\langle \vec{v}_1' \cdot \vec{v}_2' \rangle \neq 0$. In other words, after a collision, two uncorrelated velocities become correlated.

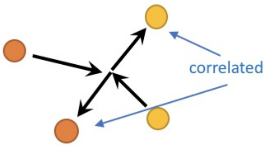


Figure 2. When two molecules collide, the outgoing velocities are correlated.

Since outgoing particles eventually become incoming particles when they collide again, the assumption of molecular chaos must not strictly hold. Why then can we use it?

The key to tracking what happens to the correlations is that the dynamics of multibody systems are chaotic, as discussed in Section 2. Since we never know exactly what the initial state is of any physical system – there is some measurement uncertainty or uncertainty due to motion of fleas across the universe – we should properly specify the state not as a point in phase space but as a region R in phase space around the point (p_i, q_i) of volume $\Delta V = (\Delta q)^{3N} (\Delta p)^{3N}$ with Δq and Δp our (classical) uncertainty on the position and momentum. To be concrete, say we have a gas of hard spheres, and let us track two of them that collide head on at $t = 0$. Right after the collision, their outgoing momenta will be highly correlated (still back-to-back). If we shift the initial momenta p_i or positions q_i by a little bit, the outgoing momenta will be slightly different, but still essentially back-to-back. So molecular chaos, Eq. (15) does not seem to apply after one collision: given one outgoing momentum, we have a pretty good idea of what the other one is.

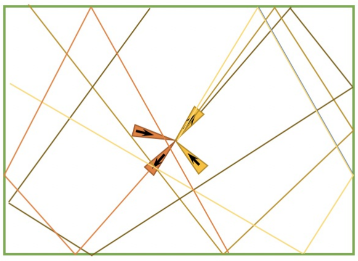


Figure 3. After a short time, nearby points in phase space follow highly-correlated trajectories through phase space.

But now let's wait a bit longer. Suppose for the exact point $(\vec{q}_i, \vec{p}_i)$ there are 9 collisions in $10^{-8}\,s$. Now consider a point $10^{-32}m < \Delta q$ away still within the region R. As we saw in Eq. (7) the trajectory of such a point will be off by an angle of order 1 after the 8 collisions, so will miss the 9th collision and move on to very different region in phase space. Thus, as time moves on, the original region R of volume ΔV gets fragmented and split up into an enormous number of disconnected regions:

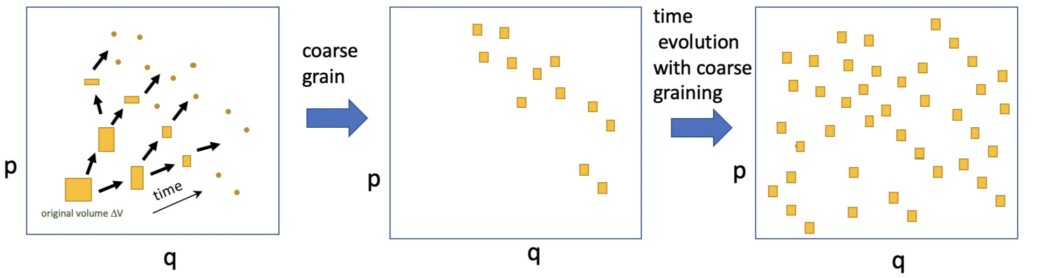


Figure 4. Over time, a phase space region R of size ΔV fragments into an enormous number of small regions with the same total volume (left). When we coarse grain, the phase space volume increases (middle). Further time evolution with coarse graining fills up more and more of phase space (right).

This doesn't yet explain molecular chaos – the small disconnected regions are still highly correlated with each other.

Now we invoke a result from classical mechanics called **Liouville's theorem**, which says that the sum of the phase-space volumes of all the fragments is the same as the original volume ΔV of the region R (it is easy to prove Louiville's theorem using Hamilton's equations of motion[1]). This means the volumes of the fragments are getting smaller and smaller after each collision. This is illustrated in the first panel of Fig. 4. However, we already said that we cannot possibly know what point in phase space our system is in with a precision better than ΔV. So we must **coarse-grain** these small phase space volumes, treating them all the same way. That is, we must accept that we cannot distinguish nearby points. Although the original small fragmented regions were highly correlated, nearby fragments that we absorb through coarse graining are not. In other words, the correlations are still there, but we cannot ever measure anything sensitive to them: it would be beyond our experimental resolution. So the effective phase space volume, from the point of view of things we can physically distinguish, is increasing (middle panel). When we coarse-grain, the correlations are completely washed out.

In summary, if we had perfect knowledge of a system, the trajectories of all the particles would be highly correlated. However, with any arbitrarily small amount of uncertainty, as we *must* have do to the lack of knowledge of fleas across the universe, those correlations get dispersed into small phase space fragments due to the chaotic nature of multibody systems. When these fragments are coarse-grained, again due to our classical uncertainty, the correlations are lost forever.

This simple fact, that chaos forces us to discard correlations, is our first indication of the arrow of time. Microscopic laws of physics are time-reversible invariant (a video of two point masses orbiting each other looks realistic in reverse), but average properties of the system made up of a large number of microscopic particles change in a fixed time direction (a video of a gas expanding does not look realistic in reverse). We saw that irreversibility arises because the correlations, which would be there if we had the exact solution to the equations of motion, become dilute in phase

---

[1] To prove Liouville's theorem, think about an infinitesimal time evolution as a change of variables, from q, p to $q' = q + \dot{q}\,dt$ and $p' = p + \dot{p}\,dt$. Then, to leading order in dt, the phase space volume $dq dp$ changes to

$$dq'\,dp' = \frac{\partial q'}{\partial q}\frac{\partial q'}{\partial p} dq dp = \left(1 + \frac{\partial \dot{q}}{\partial q} dt + \frac{\partial \dot{p}}{\partial p} dt\right) dq dp = \left(1 + \frac{\partial}{\partial q}\frac{\partial H}{\partial p} dt - \frac{\partial}{\partial q}\frac{\partial H}{\partial p} dt + \cdots\right) dq dp = dq dp \qquad (17)$$

where Hamilton's equations of motion $\dot{q} = \frac{\partial H}{\partial p}$, $\dot{p} = -\frac{\partial H}{\partial q}$ were used. Thus the phase space volume does not change upon time evolution. The generalization to many q_i, p_i just involves adding indices.

space and inaccessible. In other words, the information in the correlations becomes lost. This loss of information is a key feature of the progression of time. The arrow of time is the direction in which our ignorance grows.

4 Boltzmann's *H* theorem

Not long after Maxwell's work, Ludwig Boltzmann attempted to make the arguments for approaching equilibrium more robust and general.

Say we have some state a. For concreteness, think of a as a specification of all the momenta and positions of all the particles in a gas. Let's denote by $P_a(t)$ the probability of finding the state in a. Let's denote the rate for a to turn into some other state b by T_{ab}, and the rate for b to turn into a by T_{ba}. Then the rate of change of $P_a(t)$ is given by summing over possible states b by the equation

$$\frac{d}{dt}P_a(t) = \underbrace{\sum_b P_b(t)T_{ba}}_{\text{transitions } b\to a} - \underbrace{P_a(t)\sum_b T_{ab}}_{\text{transitions } a\to b} \tag{18}$$

If a state b in the sum can never turn into a (so $T_{ab}=0$) or if a can never turn into b (so $T_{ab}=0$), there is no point in including b in this sum together. So we can separate the problem into exclusive sets of states that can turn into each other. So let's just assume all the $T_{ab}\neq 0$.

The key property of physical systems that allows equilibrium to be approached is **the principle of detailed balance**: the transition rate from one state a to b is the same as the rate for b going to a: $T_{ab}=T_{ba}$. In classical mechanics, this follows from time-reversal invariance of the equations of motion; in quantum mechanics, it follows from the fact that the Hamiltonian is Hermitian, i.e.

$$T_{ab} = |\langle a|H_{\text{int}}|b\rangle|^2 = \langle a|H_{\text{int}}|b\rangle\langle b|H_{\text{int}}^\dagger|a\rangle = \langle b|H_{\text{int}}|a\rangle\langle a|H_{\text{int}}^\dagger|b\rangle = |\langle b|H_{\text{int}}|a\rangle|^2 = T_{ba} \tag{19}$$

In quantum field theory, it follows from unitarity (probability conservation). The principle of detailed balance is often used in chemistry: in equilibrium the rate for a reaction $A\to B$ must be the same as the rate for the reverse reaction $B\to A$.

Once we know that $T_{ab}=T_{ba}$ it follows from Eq. (18) that

$$\frac{d}{dt}P_a(t) = \sum_b T_{ab}[P_b(t) - P_a(t)] \tag{20}$$

This is a powerful equation. For example, say there are only two states. Then this equation says that if $P_a(t) > P_b(t)$ then $P_a(t)$ will go down, and if $P_b(t) > P_a(t)$ then $P_a(t)$ will go up. Thus over time, the probabilities will become the same: $\lim_{t\to\infty}P_a(t) = \lim_{t\to\infty}P_b(t)$.

To see what happens when there are N states, we consider the quantity[2]

$$H(t) = -\sum_a P_a(t)\ln P_a(t) \tag{21}$$

Then

$$\frac{d}{dt}H(t) = -\sum_a \left[\frac{d}{dt}P_a(t)\right]\ln P_a(t) - \frac{d}{dt}\sum_a P_a(t) \tag{22}$$

Since $\sum_a P_a(t) = 1$ the second term is zero. Thus

$$\frac{d}{dt}H(t) = \sum_a\sum_b T_{ab}\,[P_a(t) - P_b(t)]\ln P_a(t) \tag{23}$$

Switching the labels a and b we also have

$$\frac{d}{dt}H(t) = \sum_a\sum_b T_{ab}[P_b(t) - P_a(t)]\ln P_b(t) \tag{24}$$

---

[2] This definition of H may seem like it was pulled out of thin air. We will see in the next lecture that it is in fact related to the number of configurations and to entropy.

Averaging these two equations gives

$$\frac{d}{dt}H(t) = \frac{1}{2}\sum_a \sum_b T_{ab}[\ln P_a(t) - \ln P_b(t)][P_a(t) - P_b(t)] \tag{25}$$

Now, $\ln x$ is a monotonic function of x, so if $P_a > P_b$ then $\ln P_a > \ln P_b$. This means every term in the sum is non-negative and therefore

$$\boxed{\frac{d}{dt}H(t) \geqslant 0} \tag{26}$$

This is known as the **Boltzmann H theorem**.

If $H(t)$ is changing, then the probabilities must also be changing and we cannot be in equilibrium. Thus, equilibrium is only possible if $\frac{d}{dt}H(t) = 0$ which only happens if $P_a(t) = P_b(t)$ for all states a and b. That is,

- In equilibrium, the probabilities of finding the system in any two states a and b for which transitions can possibly occur ($T_{ab} \neq 0$) are the same

This is the postulate of equal a priori probabilities.

For a simple example, imagine you have 5 coins in a box and they all start heads up. Then you start throwing golf balls into the box, one by one. Each golf ball could hit a coin and flip it so that it may then land heads or tails. So initially, $P(\text{HHHHH}) = 1$. But after a long enough time, the probability of any configuration will be the same, $\frac{1}{2^5} = \frac{1}{32}$ and $P(\text{HHHHH}) = \frac{1}{32}$.

Note that the Boltzmann H theorem is not time-reversal invariant: H increases as we move forward in time, not backwards in time. How did this happen? The microscopic equations of motion are time-reversal invariant, so where did this arrow of time come from? In other words, for each sequence of events which takes $a \to b \to c$ there is exactly one sequence of events which goes $c \to b \to a$. For colliding molecules, we simply reverse the velocities of the outgoing molecules and then we get the initial velocities back. This mystery is known as **Loschmidt's paradox**.

To understand Loschmidt's paradox, we have to decode the implicit assumptions in Boltzmann's H theorem. First note that if we knew exactly what the state was and evolved it with time perfectly, we would always know the state, so $P_a(t)$ would either be zero or 1 for all time, just for different a at different times. For example, with the coins, if they start as HHHHH and after one hit go to HHTTH then HTHTT and so on, there is only ever one configuration possible. So $P_a(t) = 1$ for that configuration and $P_a(t) = 0$ for the others and thus $H = 0$ for all time. Thus, $\frac{dH}{dt} = 0$, which is consistent with Boltzmann's H theorem, but does not lead to the postulate of equal a priori probabilities.

So why do we say the probabilities change with the coins example? In that example, what we mean by the probability is that if threw the ball in a bunch of times, but didn't keep count, and didn't try to calculate everything, then the chance of finding any given configuration at some random later time is the same. One way to isolate our assumption is that we are implicitly talking about the time-averaged probabilities: over an interval T, what fraction of the time is a given configuration present? For large enough T these time-averaged probabilities will average out. Alternatively, we can try the experiment over and over again. Because it's hard to control the ball, we would get a different answer each time; there is inherent chaos in the system that makes it impossible to actually predict what happens. So when we repeat the experiment, what we mean by the probability is an average over the unknown parts of the initial conditions: it's not just HHHHH, but HHHHH with a flea in Australia flapping its wings, etc. Either way, the probabilities are changing because we do not have perfect information, either by choice (the time averaging) or necessity (chaos).

The key assumption in Boltzmann's H theorem is that a state a can transition to multiple states b. This is generally not possible in a unitary causal theory, since the time evolution should uniquely determine $a(t)$. So even though the theory is causal at the microscopic level, the apparent violation of causality, allowing $H(t)$ to grow, comes about because we discard information that is in principle available but practically inaccessible. We do this by time-averaging, by averaging over unknown parts of the initial conditions, and by coarse graining. These, often implicit, operations are the key to the Boltzmann H theorem, the postulate of equal a priori probabilities, and the increase of entropy (Lecture 6).

5 The ergodic hypothesis

Another concept related to chaos and coarse graining that we will need is **ergodicity**.

- An **ergodic system** is one for which the average over all possible states (the ensemble average) is the same as the average over the states that a given state will evolve into over time (the time average).

In other words, we can find the probabilities of a system being in a state at a given time t by looking at the possible states a system passes through over time.

We used ergodicity already in describing diffusion when we equated the probability of finding a particle at position x at time t, $P(x, t)$ with the number density: $n(x, t) = NP(x, t)$. Strictly speaking, $P(x,t)$ is a smooth function but $n(x,t)$ is not. At any time $n_{\text{true}}(x,t) = \frac{1}{V}\sum_i \delta(x - x_i(t))$ since the molecules are only ever at some precise points. When we write $n(x,t) = NP(x,t)$ what we mean is the time-averaged number density $n(x,t) = \frac{1}{T}\int_0^T dt' n_{\text{true}}(x, t + t')$ for some time T greater than the typical collision time τ will agree with $NP(x, t)$. Similarly, $P(x, t)$ is an average: we average over the possible random-walk paths any molecules could have taken. Each walk for each molecule constitutes a microstate. So the ensemble average over microstates, $P(x,t)$, is the time-average for a particular microstate, $\langle n(x,t)\rangle$. We usually write $n(x,t)$ instead of $\langle n(x,t)\rangle$ with the use of the ergodic hypothesis implicit.

The idea behind ergodicity that a classical trajectory through phase space $\vec{q}(t), \vec{p}(t)$ will eventually pass close to any other accessible phase space point $\vec{q}_0$, $\vec{p}_0$. For example, a gas molecule bouncing around a room, will eventually go everywhere and eventually have any momentum. Unfortunately, most systems are not ergodic, in the strict mathematical sense. Hence ergodic "hypothesis". An example of a non-ergodic system is one with closed orbits in phase space. It is not hard to find such systems. For example, imagine a circular pool table. A billiard ball bouncing around this table would never reach points closer than a certain distance from the center. It is non-ergodic. A cardioid pool table is ergodic:

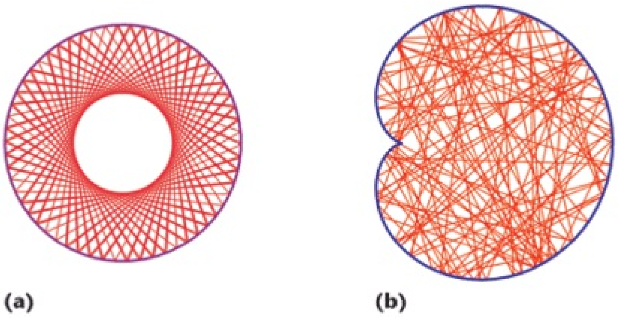


Figure 5. A circular container (left) is non-ergodic while a cardioid container (right) is. If let to continue indefinitely, the left trajectory would never reach the center, while the right trajectory would eventually fill the entire volume.

Most systems are believed to have regions of phase space which do not mix, so generally ergodicity is not exact. Moreover, even in systems where it does hold, the time for all the points in phase space to be passed through is astronomical. This is simply because the volume phase space is astronomical: $6N \sim 10^{24}$ dimensional (in constrast to the pictures in Fig. 5) which are 2 dimensional. For the trajectory of molecules in a gas to fill out a 10^{24} dimensional space will take a very very long time.

The fact that systems are not strictly ergodic and take a long time to be approximately ergodic is largely irrelevant. The reason is that we coarse grain the phase space. So instead of a $6N \sim 10^{24}$ dimensional space, we treat phase space as essentially 6 dimensional as in Eq. (15). Although it takes forever for the entire collection of 10^{24} molecules to pass near any point in the 10^{24} dimensional phase space, it does not take long at all for *one* of the 10^{24} molecules to get close to any given position and velocity.

6 Counting states

Boltzmann's H theorem immediately implies the

- **Postulate of equal a priori probabilities**: all accessible microstates are equally likely.

This postulate is really a theorem, to the extent that Boltzmann's H theorem is a theorem. It is rigorously true, as long as we coarse-grain.

The word "accessible" is required because we only proved Boltzmann's H theorem for the sets of states for which $T_{ab} \neq 0$, as in Eq. (18). For example, if we have a box of gas, a state with all the molecules on the left side and a state with them all on the right side are both accessible to each other. However, if our box had a partition in the middle then the two states would not be accessible. Accessibility can be limited by physical barriers, or by conservation laws (number of particles, energy, charge, etc.).

To make the postulate precise, we need to know how to compare probabilities. There is always some measure for the probabilities. For example, if the states are the phase space points of a gas $(\vec{q}_i, \vec{p}_i)$ there is some intrinsic resolution $(\Delta q)^{3N}(\Delta p)^{3N}$ to how well we can determine the points. Ultimately, the phase space resolution is limited by quantum uncertainty: $\Delta q \Delta p \geqslant \hbar$. For classical statistical mechanics, one does not need to invoke Planck's constant,[3] but one does need some notion of Δq and Δp – the probability of finding a system at an *exact* phase space point is necessarily zero. Thus, we will stick to the general notation of Δq and Δp for our phase space resolution, and plan to take $\Delta p \Delta q \to 0$ at the end to recover the infinite precision by which a classical system can in principle be specified. It is only in situations that are very dense, so that more than one particle might be in the same phase space point, that the actual value of $\Delta p \Delta q$ is relevant. For these situations, quantum statistical mechanics is necessary, as we will see starting in Lecture 10.

Let's take an example: an ideal gas in a box with energy E. An **ideal gas** is one where all the collisions are perfectly elastic. We treat it classically, so that positions and momenta are continuous. We assume that there are no external forces or external potential, so the energy of a particle is independent of position $\vec{q}_i$. Then the number of states is the product of the number of choices for momenta and number of states of position

$$\Omega = \Omega_q \Omega_p \tag{27}$$

Then for every state $(\vec{q}_i, \vec{p}_i)$ there is another state with the same $\vec{q}_i$ but different $\vec{p}_i$.

To count the $\vec{q}_i$, let us assume that each gas molecule can be some box of size Δq, so that there are $\Omega_1 = \frac{L}{\Delta q}$ choices for one molecule in 1 dimension and $\frac{V}{(\Delta q)^3}$ choices in 3 dimensions. We assume that any of the molecules can be in any position. Thus, for the whole ensemble,

$$\Omega_q = \left[\frac{V}{(\Delta q)^3}\right]^N \tag{28}$$

where N is the number of particles. Although Ω_q depends on the arbitrary scale Δq, we will be able to take $\Delta q \to 0$ once we have used Ω_q to compute a physical quantity.

For momenta, the calculation is harder since energy is involved. Say we have a classical monatomic gas where all the energy is in kinetic energy. Then if we know the total energy E we have a constraint:

$$E = \sum_j \frac{\vec{p}_j^{\,2}}{2m} \tag{29}$$

There are going to be many choices of $\vec{p}_j$ for this constraint equation to hold. How do we count them?

Let's start with 2 particles in 1 dimension. Then

$$2mE = p_1^2 + p_2^2 \tag{30}$$

---

[3] In classical mechanics, the properties of phase space and probability distributions on it have been very thoroughly studied. We will try to keep our discussion intuitive and non-technical as much as possible. See Grandy's book in pdf form on the Canvas site for more discussion.

This is the equation for a circle of radius $R=\sqrt{2mE}$. Thus the number of states, normalized to a unit size of Δp is $\Omega_p=\frac{2\pi R}{\Delta p}=\frac{2\pi\sqrt{2mE}}{\Delta p}$.

If there are 3 particles in 1 dimension, then

$$2mE=p_1^2+p_2^2+p_3^2 \tag{31}$$

the number of states is determined by the surface area of a sphere: $\sigma_3=4\pi$. Then we get that $\Omega_p=\frac{4\pi R^2}{\Delta p^2}=\frac{4\pi\,(2mE)}{\Delta p^2}$.

If there are N particles in 3 dimensions, then

$$2mE=p_{1x}^2+p_{1y}^2+p_{1z}^2+\cdots+p_{Nx}^2+p_{Ny}^2+p_{Nz}^2 \tag{32}$$

this is the sum of the squares of $3N$ independent momenta. Then

$$\Omega_p=\sigma_{3N}\left(\frac{\sqrt{2mE}}{\Delta p}\right)^{3N} \tag{33}$$

where σ_d is the surface area of a d-dimensional sphere with radius $r=1$. For $d=2$, $\sigma_2=2\pi$, for $d=3$, $\sigma_3=4\pi$. For general d the result is[4]

$$\sigma_d=\frac{2(\sqrt{\pi})^d}{\left(\frac{d}{2}-1\right)!} \tag{36}$$

Therefore,

$$\Omega_p=\frac{2(\sqrt{\pi})^{3N}}{\left(\frac{3}{2}N-1\right)!}\left(\frac{\sqrt{2mE}}{\Delta p}\right)^{3N} \tag{37}$$

For large N, we can write $3N-1\approx 3N$ and also Stirling's approximation $N!\approx e^{-N}N^N$ so that $\left(\frac{3}{2}N\right)!\approx e^{-\frac{3}{2}N}\left(\frac{3}{2}N\right)^{\frac{3N}{2}}$ giving

$$\Omega_p=e^{\frac{3}{2}N}\left(\frac{4\pi mE}{3N(\Delta p)^2}\right)^{\frac{3N}{2}} \tag{38}$$

Combining with the phase space for position in Eq. (28), we get

$$\boxed{\Omega(N,V,E)=e^{\frac{3}{2}N}\left(\frac{V}{(\Delta q\Delta p)^3}\right)^N\left(\frac{4\pi mE}{3N}\right)^{\frac{3N}{2}}} \tag{39}$$

A key feature of the number of states is that it is an extremely rapidly growing function of energy – it grows like energy to the power 10^{24}. That is,

$$\Omega(E)\sim E^{10^{24}} \tag{40}$$

So when you add energy to a system, the number of states grows exponentially. For example, say we increased the energy by 0.00001% ($E\rightarrow E+10^{-7}E$). Then the number of states grows by

$$\#\text{new states}=\Omega(E+10^{-7}E)-\Omega(E)=10^{-7}\partial_E E^{10^{24}}=10^{18}\,\Omega(E) \tag{41}$$

This is an enormous number, 10^{18}, times a ridiculously enormous number $E^{10^{24}}$.

---

[4] To derive Eq. (36) we first compute the 2D integral

$$\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,e^{-x^2-y^2}=2\pi\int_0^{\infty}r\,dr\,e^{-r^2}=\pi \tag{34}$$

Then we can generalize to d dimensions:

$$(\sqrt{\pi})^d=\left[\int_{-\infty}^{\infty}dp\,e^{-p^2}\right]^d=\int_{-\infty}^{\infty}dp_1\cdots dp_n e^{-p_1^2-\cdots-p_d^2}=\sigma_d\int dr\,r^{d-1}\,e^{-r^2} \tag{35}$$

The last integral is a 1D integral that we can do with mathematica. It gives $\frac{\sigma_d}{2}\Gamma\left(\frac{d}{2}\right)$ where $\Gamma(x)$ is the Gamma function. The Gamma function is a generalization of the factorial function. For positive integers, $\Gamma(n)=(n-1)!$.

In the next lecture, we will consider situations that have more contributions to the energy in Eq. (32), such as from vibrational modes of a molecule. In such situations, the calculation is similar, resulting in $\Omega \sim E^{\frac{f}{2}N}$ rather of $\Omega \sim E^{\frac{3}{2}N}$ with f the number of degrees of freedom in which the molecules can store energy.

7 Maxwell-Boltzmann distribution

After all this rather formal introduction, we are finally ready to compute something observable: the velocity distribution of molecules in a gas.

What is the probability of finding the p_x component of the momentum of one molecule in a small region of size Δp around p_x? According to the postulate of equal a priori probabilities, this probability is proportional to the number of states compatible with this restriction, divided by the total number of states: $P(p_x) = \frac{\Omega_{p_x \text{fixed}}}{\Omega_{\text{total}}}$.

Once we fix p_x the remaining energy is $E' = E - \frac{p_x^2}{2m}$. So the remaining phase space volume, with energy E' is

$$\Omega(E') = e^{\frac{3N-1}{2}} \left(\frac{V}{(\Delta q \Delta p)^3} \right)^N \left(\frac{4\pi m E'}{3N-1} \right)^{\frac{3N-1}{2}} \Delta p \approx e^{\frac{3N}{2}} \left(\frac{V}{(\Delta q \Delta p)^3} \right)^N \left(\frac{4\pi m E'}{3N} \right)^{\frac{3N}{2}} \Delta p \tag{42}$$

The probability of finding the x component of momentum between p_x and $p_x + \Delta p$ is therefore

$$\frac{\Delta P}{\Delta p} \equiv \frac{P(p_x \,\text{to}\, p_x + \Delta p)}{\Delta p} = \frac{\Omega(E')}{\Omega(E)\Delta p} = \frac{(E')^{3N/2}}{(E)^{3N/2}} = \left(1 - \frac{p_x^2}{2mE} \right)^{3N/2} \tag{43}$$

This function ΔP looks a lot like the limit definition of the exponential function

$$e^{-x} = \lim_{N \to \infty} \left(1 - \frac{x}{N} \right)^N \tag{44}$$

To make it match exactly, let us introduce the average energy

$$\bar{\varepsilon} = \frac{E}{N} \tag{45}$$

Writing Eq. (43) in terms of $\bar{\varepsilon}$ we can then take $N \to \infty$ using Eq. (44):

$$\frac{\Delta P}{\Delta p} = \left[\left(1 - \frac{1}{N} \frac{p_x^2}{2m\bar{\varepsilon}} \right)^N \right]^{3/2} \xrightarrow[N \to \infty]{} \left[\exp\left(-\frac{p_x^2}{2m\bar{\varepsilon}} \right) \right]^{3/2} = \exp\left(-\frac{3p_x^2}{4m\bar{\varepsilon}} \right) \tag{46}$$

In taking the limit, we have messed up the normalization. We would like the probabilities to be normalized so that $\sum_{p_x} \Delta P(p_x) = 1$. Rather than working with the discrete sum, it is easier to go straight to the continuum limit. Defining $\frac{dP}{dp_x} = C \frac{\Delta P}{\Delta p}$ and choosing C so that $\int \frac{dP}{dp_x} dp_x = 1$ we get

$$\frac{dP(p_x)}{dp_x} = \sqrt{\frac{3}{4\pi m \bar{\varepsilon}}} \exp\left(-\frac{3p_x^2}{4m\bar{\varepsilon}} \right) \tag{47}$$

Note that Δq and Δp have dropped out of this expression so we take $\Delta p \to 0$ and $\Delta q \to 0$

We could repeat the calculation for p_y and p_z. Since the theory is rotationally symmetric, we get the same answer. Therefore,

$$\frac{d^3 P(\vec{p})}{dp_x dp_y dp_z} = \left(\frac{3}{4\pi m \bar{\varepsilon}} \right)^{3/2} \exp\left(-\frac{3p_x^2}{4m\bar{\varepsilon}} \right) \exp\left(-\frac{3p_y^2}{4m\bar{\varepsilon}} \right) \exp\left(-\frac{3p_z^2}{4m\bar{\varepsilon}} \right) \tag{48}$$

In other words

$$\boxed{\frac{d^3 P(\vec{p})}{dp_x dp_y dp_z} = \left(\frac{3}{4\pi m \bar{\varepsilon}} \right)^{3/2} e^{-\frac{1}{\bar{\varepsilon}} \frac{3\vec{p}^2}{4m}}} \tag{49}$$

This is known as the **Maxwell-Boltzmann distribution**.

As a check, we compute the average value of kinetic energy:

$$\langle\frac{\vec{p}^2}{2m}\rangle = \int d^3p\left(\frac{\vec{p}^2}{2m}\right)\frac{d^3P(\vec{p})}{d^3p} = \int d^3p\frac{\vec{p}^2}{2m}\left(\frac{3}{4\pi m\,\bar{\varepsilon}}\right)^{3/2} e^{-\frac{1}{\bar{\varepsilon}}\frac{3\vec{p}^2}{4m}} = \bar{\varepsilon} = \frac{E}{N} \tag{50}$$

This is as expected, since all energy is kinetic. Note that this result is consistent with Eq. (14) which we derived using kinetic theory.

Now wait a minute ... didn't we say that each state is equally likely, but now we say that each state has probability $e^{-\frac{1}{\bar{\varepsilon}}\frac{3\vec{p}^2}{4m}}$? How are these statements consistent? Let's be careful. What we said is that each microstate with total energy E is equally likely. This is still true. But if we now start grouping the microstates the value of $\vec{p}^2$ for a given molecule, then we find fewer and fewer such microstates with larger values of $\vec{p}^2$. This is because the bigger $\vec{p}$ is, the fewer ways there are to split up the remaining energy among the other molecules, and thus the probability of finding $\vec{p}$ goes down as $\vec{p}^2$ goes up.

If we change variables from $\vec{p}^2$ to speed $v = \frac{\sqrt{\vec{p}^2}}{m}$ (using $d^3p = 4\pi|\vec{p}|^2 d|\vec{p}| = 4\pi m^3 v^2 dv$) and substitute $\bar{\varepsilon} = \frac{3}{2}k_B T$ (a result we'll derive in the next lecture), we get the **Maxwell-Boltzmann** distribution:

$$\boxed{\frac{dP(v)}{dv} = 4\pi v^2\left(\frac{m}{2\pi k_B T}\right)^{3/2} e^{-\frac{mv^2}{2k_B T}}} \tag{51}$$

This looks like

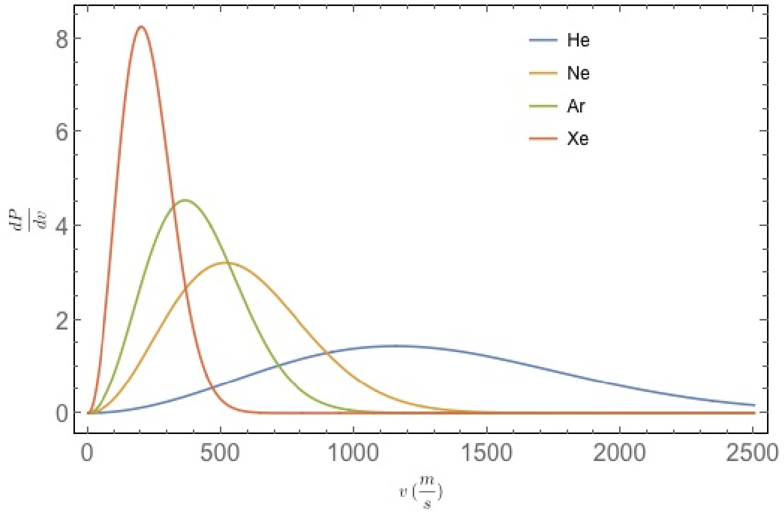


Figure 6. Maxwell-Boltzmann velocity distributions for some gases with the same value of the average energy $\bar{\varepsilon} = \frac{E}{N} = \frac{3}{2}k_B T$.

8 Summary

There were a lot of new concepts introduced in this lecture and some very important results. So let's recap:

- **Chaos**: the trajectory of molecules in a gas are strongly sensitive to even the smallest perturbations, like a flea flapping its wings across the universe.
- **Molecular chaos**: when molecules scatter, their outgoing velocities are correlated. Due to chaos, these correlations are rapidly disperse throughout phase space, dissolving into every smaller, separated regions, like a kind of phase space dust.
- Because the correlations are dilute, they get lost when we **coarse grain** (average over nearby regions in phase space). Coarse graining lets us treat the probability of each molecule occupying a point in its phase space as independent of what the other molecules are doing, as in Eq. (15). It breaks time-reversal invariance.

- The **Boltzmann *H*-Theorem**: with correlations discarded, probabilities tend towards uniformity over phase space.
- The H-theorem follows from the **principle of detailed balance**: the rate for a process and the reverse process is the same.
- **Ergodicity**: the probability of finding a molecule at point $(\vec{p}, \vec{q})$ at a given time t is the same as the probability of finding it at $(\vec{p}, \vec{q})$ averaged over t.
- **Postulate of equal a priori probabilities**: in equilibrium, a system is equally likely to be found in any accessible state.
- **Maxwell-Boltzmann distribution**: the distribution of velocities of gas molecules is computed by counting the number of ways the the total energy of the gas can be distributed among the molecules.

Much of the material in this lecture was abstract and foundational. It built up to the postulate of equal a priori probabilities. We then applied this postulate to derive the Maxwell-Boltzmann distribution in Sections 6 and 7. Going forward, we will do many more applications, referring back to this foundational material when appropriate. If you are confused, don't despair: this is probably the most conceptually difficult lecture in the entire course.