Concepts in Theoretical Physics

Lecturer: Dr. Eugene A. Lim
King's College London
Department of Physics
June 19, 2022

Acknowledgments

This module was hugely inspired by a similar course taught in the Cambridge DAMTP Mathematical Tripos. I am particularly grateful to Daniel Baumann for not just providing great insights and suggestions on how to teach this course and making his course notes available for me to shamelessly steal, but also for the many years of discussions we had on how to teach fundamental concepts to students with the minimal of scaffolding. I would also like to thank the 2020-2021 academic 1st year physics students of King's College London for patiently sitting through a prototype of this course delivered on zoom – after all the zoom lectures they have to sit through during the pandemic, I am grateful for them to show up for this optional course with no credit to provide me with useful feedback. And finally, I would like to thank Sophie and Astra, for their patience while I obsess over these course notes.

What is this course about?

This course is an attempt to teach modern theoretical concepts to 1st year undergraduate students, assuming that the only mathematical preparation they have is linear algebra and differential/integral calculus. Why do we want to do that? I can think of several reasons. The first reason is simply that physics is suppose to be *fun*, and I think many students took up physics because they were inspired to learn more by all the exciting things they may have heard in the popular media, only to find out that to get to the good stuff requires a huge amount of pre-requisite knowledge. So we can think of this course as a way to nourish this excitement while they plow through the background knowledge before they get to the really good stuff. The second reason is that it can serve as signposts for the students as they navigate the enormous amount of material that we regularly throw at them in their standard modules – and hopefully helping them to organize their learning better. If they know *why* studying those nasty functional integrals are important, they are more likely to learn it with sharper focus. Finally, the third reason is that providing such "big picture" landscape of theoretical physics will help them to plan their career earlier, especially if they want to pursue a career in physics.

The course is designed to teach students the fundamental concepts so that the problem of quantum gravity can be described, thus the subjects are chosen with that focus in mind. So right now, there is about 10-15 hours of lecturing material. Obviously, there are things that can be added in (and probably should be added in) – Cosmology, Particle Physics, Superconductivity, Electromagnetism, dimensional analysis, to name some examples. But, we'll leave that for the future.

Contents

Chapter 1

Quantum Mechanics

> *I think I can safely say that nobody understands quantum mechanics.*
>
> Richard Feynman

1.1 Introduction : The Classical World

Until the early 20th century, scientists believed that the laws of nature are determined by the (retroactively named) so-called the **classical physics**, mostly due to the Isaac Newton. This viewpoint asserts that objects are immutable, and fully deterministic in the sense that you can exactly measure all their properties to as accurate as you want, at any time you want. For example, consider the simple case where you see a cannonball of mass m flying in the air at some time t_0. You measure its position $\mathbf{x}(t_0)$ and momentum $\mathbf{p} = m\dot{\mathbf{x}}(t_0)$ (where each dot denotes a derivative with respect to time t, e.g. $\dot{x} \equiv d\mathbf{x}/dt$). Then, by using the Newton's 2nd Law of motion

$$\ddot{\mathbf{x}} = \mathbf{g} \ , \tag{1.1}$$

where $\mathbf{g} = 9.81\text{m s}^{-2}$ is the gravitational acceleration of Earth, you can then calculate and *predict* the exact trajectory of this cannonball $\mathbf{x}(t)$ and its momentum $\mathbf{p}(t)$ as far to the future as you like. Furthermore, you can calculate *backwards in time* to determine the path of the cannonball as far past as you like. At any time t, once we know the position and its momentum $(\mathbf{x}(t), \mathbf{p}(t))$, or the **state** of the cannonball, we know its entire history far into the future, and far into the past.

This viewpoint assumes that the act of measurement does not alter the trajectory of the cannonball. Given our everyday experience, this seems a reasonable viewpoint. We, the **observers**, are omnipotent – we exist outside of the life of the things we measure (in the above example, the cannonball never felt our act of measurement). Furthermore, we are allowed to measure as accurate as we want its state, at anytime we like.

However, this viewpoint is wrong. As 20th century rolls onwards and technology improved, scientists began to probe deeper into the microscopic world to study the structure of atoms and the properties of elementary particles such as the electrons. The classical viewpoint began to unravel, and scientists realized that classical physics is just an *approximation* of the real world, which is actually described by a more fundamental theory : **quantum mechanics**. The key difference between the classical and quantum world is the role of the observer. In the quantum world, the observer is not omnipotent – whether or not they like it or not, *to make a measurement, they must participate in the experiment.*

In our short 2 hour lecture on quantum mechanics, I will not be retelling the story of how classical physics unraveled – I will leave it for you as assignments instead. Instead, I want to try to teach you the fundamentals of quantum mechanics itself. In this spirit, let's consider two of the most dramatic consequences of quantum mechanics, which you might have read or heard somewhere else before.

1.1.1 The Heisenberg Uncertainty Principle

Now, let's imagine we actually try to do the experiment of measuring the momentum and position of the cannonball that we described above. You pay a lot of money for the world's best instruments in making measurements of momentum and position of cannonballs. You have a friend shoot a cannonball into the air, and in midflight at some predetermined time t seconds after the cannonball was shot, you measure its position $\mathbf{x}$ and its momentum $\mathbf{p}$. You do this experiment many times – you ask your friend to shoot the same cannonball at the same direction with the same velocity many times. You make many measurements at the same time t seconds after the cannonball is shot. According to classical physics, you should always get the same values of $\mathbf{x}$ and $\mathbf{p}$ at time t – after all classical physics say that the state $(\mathbf{x}(t), \mathbf{p}(t))$ can be precisely measured and would always obey Newton's 2nd law of motion. However, despite your best attempts with the very best equipment money can buy, you found that you never get exactly the values $(\mathbf{x}(t), \mathbf{p}(t))$. Instead, you find that you keep getting some small errors from the expected values of $(\mathbf{x}, \mathbf{p})$. Now you will find that measurements very far away from your expected values of $(\mathbf{x}(t), \mathbf{p}(t))$ are rarer than those closer to it. You can plot the results of your measurements in a plot as shown in Fig. 1.1. The widths $\Delta\mathbf{p}$ and $\Delta\mathbf{x}$ of the so-called "Gaussian" (or Normal) distributions are called their *variance*. You find that, despite your best efforts, you will aways end up with the non-zero values of $\Delta\mathbf{x}$ and $\Delta\mathbf{p}$. Even more intriguingly, you find that

$$\Delta\mathbf{x}\Delta\mathbf{p} \geq \frac{\hbar}{2} , \tag{1.2}$$

where $\hbar = 1.054 \times 10^{-34}$ J s is known as the **Planck's constant**.

What happens if you try to improve the accuracy of your measurements? Suppose you bought even more expensive instruments, and you find, sure enough your measurements of $\mathbf{x}$ got better so the error variance $\Delta\mathbf{x}$ becomes smaller. But strangely, even though you have bought equally expensive instrument to measure $\mathbf{p}$, you find that as your $\Delta\mathbf{x}$ got smaller, your $\Delta\mathbf{p}$ got bigger such that Eq. (1.2) is *always* obeyed! Even more strange, you find that whether $\Delta\mathbf{x}$ or $\Delta\mathbf{p}$ got smaller doesn't depend on how good your instruments are[1]? How do you even explain this behaviour?

This experiment actually has been done many times and this is indeed how nature behaves (although we don't use cannonballs, we use smaller particles)! The relation Eq. (1.2) is called the **Heisenberg Uncertainty Principle** and is not just an experimental fact, but is a *theoretical* fact that can be derived from the laws of quantum mechanics.

1.1.2 The Paradox of the Schrödinger's Cat

You may have heard of the story of the Schrödinger's Cat. **Erwin Schrödinger** propose this *gedanken-experiment* ("thought experiment") in 1935 to illustrate the weirdness of the fact that the observer has to participate in any measurement. A cat is put inside a closed box. There is a vial of poison gas in the box. A hammer will strike the vial if a certain amount of radioactivity is detected in the box, thus killing

[1] In real world quantum measurements, since measurement actively disturb the system you can never reach the Heisenberg limit, but only roughly twice the limit. This is known as the *standard quantum limit.* This is a subtle point – there are two sources of error in measurements, one coming from instruments that will disturb the system, and one coming directly from inherent quantum uncertainty, and they are not the same thing. Indeed, Heisenberg himself got confused, and thought that the explanation for the uncertainty principle is due to the fact that we have to disturb the system to measure it, and that's incorrect.

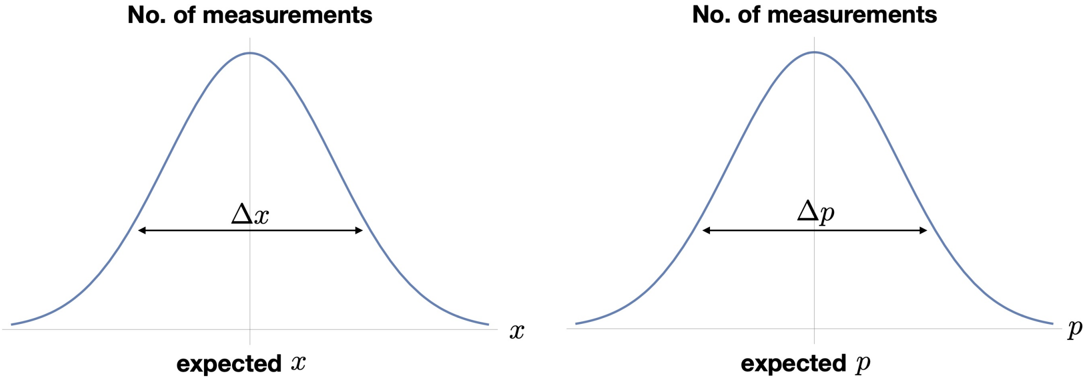


Figure 1.1: Results of Measurements of position and momentum of cannonball. The rough width of the errors $\Delta \mathbf{x}$ and $\Delta \mathbf{p}$ are called its variance, and obey the Heisenberg Uncertainty Principle $\Delta \mathbf{x} \Delta \mathbf{p} = \hbar/2$.

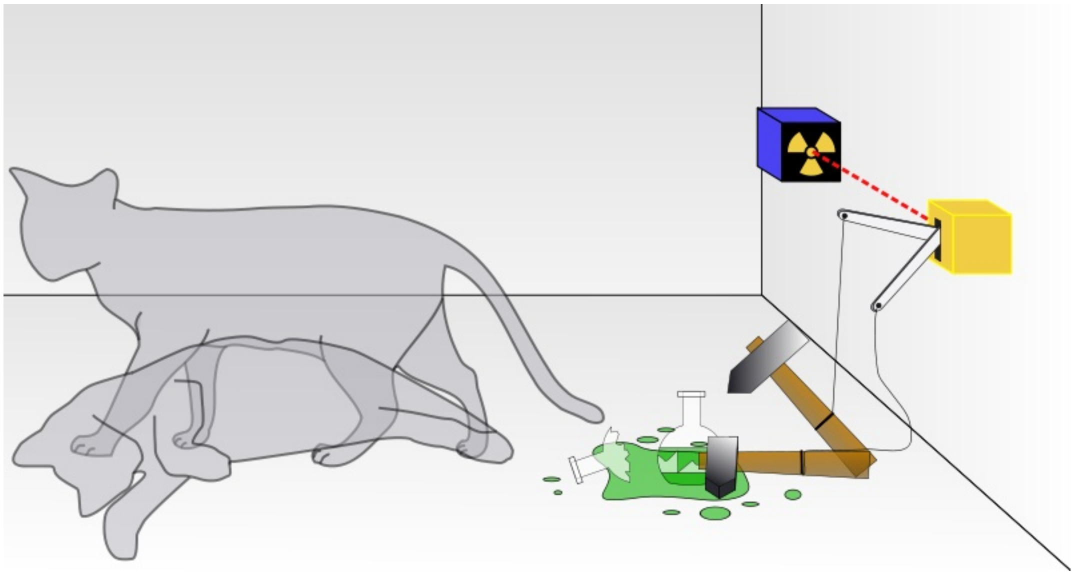


Figure 1.2: Schrödinger's Cat and its sad/happy fate. Stolen from Wikipedia.

the cat. An observer outside the box has no way of finding out if this sad affair has occured without opening the box. Hence the cat is in the curious state of being *both alive and dead at the same time* according to the observer before the opening. Once the box is opened, the observer has a 50% chance of finding a dead cat, and a 50% chance of finding a live cat. What the observer *never* see however, is a cat that is neither dead nor alive – the act of opening the box, or making a measurement – "forces" nature to "choose" an outcome for our poor cat.

This strange state of affairs is actually experimentally tested, not with cats and poison of course (since that would violate animal cruelty laws), but with photons. How do we even *describe* such a state for the cat?

1.2 The Four Rules of Quantum Mechanics

The two "paradoxes" in the previous Section are strange. So strange that classical physics has no way of explaining them. To explain them, we not only need new equations, we need a totally new way of describing reality. Undergraduate quantum mechanics courses usually take a whole year to teach, so we don't really have time to do that. However, what we are going to do is to teach you the core of it, so that

INPUT	OUTPUT
↓	↑
↑	↓

Table 1.1: A NOT gate

you get a sense of what quantum mechanics is. The goal is to teach you sufficient quantum mechanics to show you how to resolve the two paradoxes above – they are only paradoxes if we use our "classical intuition".

As it turns out, there are actually only **Four** Rules of quantum mechanics. These Rules are **postulated**, and are fundamental axioms that everything about quantum mechanics are derivable from. By "Postulate", we mean that they are *not derivable* – or at least nobody at this moment know how to derive them (and people have been trying for more than a century.) In these lecture, I will now tell you these rules.

1.2.1 Classical Bit vs Quantum Bit

In order to show you the key points of quantum mechanics, we will use the simplest possible quantum system, which is that of a **qubit**. As children of the computer revolution, you must be familiar with the idea of a **bit** of information. The bit is a system that can only has two possible states: 1/0 or ↑ / ↓ or on/off or dead cat/live cat etc. Let's use ↑ / ↓ for now. So instead of the cannonball, whose state is described by the functions for position and momentum, $(\mathbf{x}(t), \mathbf{p}(t))$, a bit's state is described simply by two possible states of either 1 or 0. Such binary systems are also called (obviously) *two-state* systems.

The cannonball obey Newton's Law of motion. What about the bit? We can endow this bit with some set of physical rules which when acted upon the system, may change it from one state to another. What kind of rules can we write down for a bit? The set of rules for a bit can be something simple like a NOT gate. This rule simply flips an ↑ to a ↓, and a ↓ to an ↑. A NOT gate rule is shown in Table 1.1. Another rule we can write ↓ is the "do nothing" gate, which just returns ↑ if acted on ↓ , and ↓ if acted on ↑. Mathematically, we can define the following column matrices to represent the ↑ / ↓ states

$$\chi_{\uparrow} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \chi_{\downarrow} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.3}$$

so a NOT gate can be described by the 2×2 matrix

$$\hat{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{1.4}$$

while a "do nothing" gate is obviously the identity

$$\hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.5}$$

"Acting" then means usual matrix multiplication of the column vector from the left by the gate matrix

$$\text{result} = \text{gate matrix} \times \text{state}. \tag{1.6}$$

You can check that acting from the left with $\hat{P}$ and $\hat{I}$ on an up/down state gets you the right results, e.g. acting on up state with NOT gate yields a down state

$$\chi_{\downarrow} = \hat{P}\chi_{\uparrow}. \tag{1.7}$$

We will now introduce the high-brow mathematical word **operator**, which is an object that acts on a state and give you (possibly the same) another state. For example, suppose $\hat{O}$ is an operator acting on the state ψ to give you another state ψ', we write

$$\psi' = \hat{O}\psi. \tag{1.8}$$

Thus $\hat{P}$ and $\hat{I}$ are both operators. As we will see later, there are a lot more of these.

A bit is a classical quantity, so we can measure with arbitrary accuracy whether it is ↑ or ↓. For example, a classical cat is either dead or alive (just check its pulse). We can also *predict* with arbitrary accuracy what would happen when we act on the bit with the rules: if we start with a ↑, acting on it with a NOT gate we predict that it will become a ↓ (and then we can measure it to confirm that our prediction is true).

Now, a classical bit is a perfect model for a classical cat : we can say "↑" means live cat, and "↓" means dead cat. What about a quantum cat? It is clear that we cannot describe a quantum cat with a bit – we need a way to mathematically describe the ambiguity of a cat in the box which can be both alive and dead. This leads to the **First Rule** of Quantum Mechanics.

Rule 1 (State): A *qubit*, ψ, is described by the **state vector**

$$\psi = \alpha\chi_{\uparrow} + \beta\chi_{\downarrow}\ , \text{ where } \alpha, \beta \in \mathbb{C}. \tag{1.9}$$

α and β are called **probability amplitudes** for finding the ↑ and ↓ state, for reasons we will soon see. The important point here is that the coefficients α and β are *complex numbers* (this is what "$\in \mathbb{C}''$ means in proper mathematic language) – this means that the information encoded in the state has been enlarged when compared to the binary classical bit[2]. Rule 1 tells us that the state can be neither ↑ nor ↓ ; it is some **linear superposition** beteween two possible states – hence the cat can be both dead and alive. Notice that ψ is now in general a complex 2×1 matrix – although in this lecture we will just do examples where ψ is real. By convention, we normalize the state vector $(\psi^T)^*\psi = 1$, hence $|\alpha|^2 + |\beta|^2 = 1$, where the superscript T denotes transpose and $*$ denotes complex conjugration[3].

Now that we have mathematically model a cat state that can be both dead and alive, how do we model the how probable the cat is alive/dead when we open the box? This leads us to the **Second Rule**:

Rule 2 (Born's Rule[4]): The *probability* of measuring an ↑ / ↓ state is the absolute square of the inner product of the desired outcome with the state, i.e.

$$\text{Probability of measuring } \uparrow \text{ state} = |\chi_{\uparrow} \cdot \psi|^2 = |\alpha|^2, \tag{1.12}$$

$$\text{Probability of measuring } \downarrow \text{ state} = |\chi_{\downarrow} \cdot \psi|^2 = |\beta|^2. \tag{1.13}$$

Note that since the qubit has to be in some state, the probability must add up to unity $|\alpha|^2 + |\beta|^2 = 1$ – this is the reason why the state vectors are normalized to one. More generally, state vectors must be

[2] *Technically, the space in which a two-state quantum mechanically system live in is a S_2 sphere called the **Bloch Sphere** where the ↑ / ↓ state reside at the North and South poles of this sphere.*

[3] The combination of these two operations is called **Hermitian Conjugation**, which we denote with a $\dagger$ i.e. for any complex matrix $\hat{A}$

$$(\hat{A}^T)^* \equiv \hat{A}^\dagger \tag{1.10}$$

This operation occurs so often in Quantum Mechanics that we will define the **inner product** (or "dot product") of two state vectors the following way. Given two state vectors ϕ and ψ, the inner product is then defined as

$$\phi \cdot \psi \equiv \phi^\dagger \psi. \tag{1.11}$$

[4] Named after physicist Max Born (1882-1970) who won a Nobel Prize for this work. Fun fact, he is the grandfather of singer Olivia Newton-John, who won 4 Grammy awards.

normalizable to be valid quantum mechanical states. A note on jargon: note that probability amplitudes are complex, while probabilities are real.

Finally, what happens when we "make a measurement" or, in this case, open the box with the cat in? This is given by the **Third Rule**:

Rule 3 (Measurement): Once a measurement has been made and ↑ / ↓ has been obtained, the state vector ψ **collapses** into the measured state

$$\psi \overset{\text{measure}}{\longrightarrow} \chi_{\uparrow/\downarrow}. \tag{1.14}$$

While Rule 1 tells us that a qubit can be neither up nor down, Rule 2 tells us the probability of measuring either of the two states. Rule 3 then tells us that once the measurement has been made, follow up measurements will yield identical results (as long as we have not act on the state other than make a measurement). In particular, Rule 3 implies that *the very act of measurement affects the system.* This is often called the **Collapse of the State Vector**.

So the story of the cat is now the following: the state of aliveness/deadness of the cat is carried by a qubit due to the quantum mechanical nature of radioactivity, and the state is described by the following qubit

$$\psi = \frac{1}{\sqrt{2}}\chi_\uparrow + \frac{1}{\sqrt{2}}\chi_\downarrow \tag{1.15}$$

which is just Eq. (1.9) with probability amplitude $\alpha = \beta = 1/\sqrt{2}$. According to Rule 2, the probability of finding the cat to be dead or alive when we open the box is given by $|\alpha|^2 = 1/2$ or $|\beta|^2 = 1/2$, i.e. 50% either way. Once the box is opened, the cat's state will collapse into one of the two states depending on which is measured. In fact, such up/down quantum states are now known as **cat states**, and it can be created in the laboratory, not with real cats, but for example, with ^{87}Rb atoms interacting with an weak optical field.

At this stage, you can rightfully ask – well, if all we want is to mathematically model a cat which can have some probability of being alive and some probability of being dead, why do we need the complex numbers α and β, why can we just model it as something like

$$\psi \overset{?}{=} a\chi + b\chi \ , \ a, b \in \mathbb{R} \ , \tag{1.16}$$

where a and b are real numbers? This is an excellent question, and is unfortunately a discussion that we don't have the time go into. The short (possibly unhelpful answer) is that this is due to the fact that quantum states exhibit a expermentally verified property called **wave-particle duality**, which is that quantum states can behave like a wave and sometimes like a particle. The wave nature of quantum states is modeled by having complex numbers. You can read more about this and do an assignment if you like!

1.2.2 Quantum Entanglement

Before we discuss Rule 4 of quantum mechanics, let's talk about one of the most remarkable consequences of quantum mechanics known as **quantum entanglement**.

A classical bit can have 2 possible (↑ or ↓) state. As you probably know, *two* classical bits can then have 4 possible states ↑↑, ↑↓, ↓↓ and ↓↑. What about 2 qubits? Recall that a qubit can be in a superposition of 2 states $\psi = \alpha\uparrow + \beta\downarrow$ – where we have now for simplicity written $\chi_{\uparrow/\downarrow}$ as simply ↑ / ↓ – 2 qubits can be in superposition of 4 possible states

$$\Psi = \alpha \uparrow\uparrow + \beta \uparrow\downarrow + \gamma \downarrow\downarrow + \sigma \downarrow\uparrow \ , \ \alpha, \beta, \gamma, \sigma \in \mathbb{C} \tag{1.17}$$

where again $\alpha, \beta, \gamma, \sigma$ are all complex numbers. Note that the left arrow denote the first qubit, and the right arrow denote the second qubit, i.e. ↑↑ really means $\uparrow_1\uparrow_2$ etc, but we drop the labels to make the notation less messy. So far, so good. But now, let's say we set up the following state

$$\Psi = \sqrt{\frac{1}{2}} \uparrow\downarrow + \sqrt{\frac{1}{2}} \downarrow\uparrow . \tag{1.18}$$

i.e. where $\alpha = \gamma = 0$ and $\beta = \sigma = 1/\sqrt{2}$. We are given a detector that can make measurement on a *single* qubit. Let's use it on the *first* qubit. According to Rule 2, the probability of measuring an $\uparrow$ state is the square of the probability amplitude of the $\uparrow\downarrow$ term in Eq. (1.18), which is $|1/\sqrt{2}|^2 = 1/2$ or 50% chance.

But now according to Rule 3, once a measurement is made, the state collapsed to its measured state, which in this case is $\uparrow\downarrow$. In other words the following sequence of events has occured

$$\Psi \xrightarrow{\uparrow_1} \uparrow\downarrow \tag{1.19}$$

which hopefully by now you are not surprised – the 2nd qubit has jumped to its $\downarrow$ state! Similarly, there is a 50% probability of measuring an $\downarrow$ state in the first qubit, and consequently by Rule 3, the second quibit will jump to the $\uparrow$ state in this case.

A long time ago, this strange state of affairs was incredibly troubling to Einstein. He, together with Podolsky and Rosen, propose the so-called **Einstein-Podolsky-Rosen** paradox : prepare a 2 qubit state as in Eq. (1.18), and keep the two qubits in two unopened boxes. Send one qubit to Alice at one end of the universe, and the other qubit to Bob at the other end of the universe.

Now Alice wants to open the box. At this stage, Ψ is "uncollapsed", so her probability of finding an $\uparrow$ in her box is 1/2 as we just calculated above. If she now opens the box, and found an $\uparrow$ state, then Bob will open his box and find his state to be $\downarrow$ with probability 1. If instead, Alice opens her box and found $\downarrow$ then Bob will find that his state is $\uparrow$! This is just a story version of the calculation we did above, but EPR were very upset because it seems to imply that information has traveled at the instance Alice opened her box to Bob's qubit in his box instantly. But there is no paradox – the states are *correlated* in such a way that measurement of one imply the other. You might say (as Einstein did), "but wait, hasn't information traveled instantaneously, which violate the light speed limit?". The answer is no – Alice (or Bob) cannot use this entanglement to send a message to Bob (or Alice) faster than the speed of light because *she has no way of deciding which $\uparrow$ or $\downarrow$ state she would measure.*

In honor of our angst-ridden physicists, nowadays we call the entangled state Eq. (1.18) an **EPR pair**, which goes to show that as long as you are famous enough, even drawing the wrong conclusions can get you recognition with things named after you.

1.3 What happens when we make a measurement?

Finally, let's talk about the **Fourth Rule**, which in my opinion is the trickiest rule of quantum mechanics. To motivate the reason for its existence, let's think back to the Heisenberg Uncertainty Principle back in section 1.1.1. There, we described the maddening problem of trying to measure the position $\mathbf{x}$ and momentum $\mathbf{p}$ as accurately as possible and found that nature won't let you, subjecting to errors enforced by the Heisenberg Uncertainty Principle Eq. (1.2). Somehow, measuring $\mathbf{x}$ has affected $\mathbf{p}$ (and *vice versa*) – perhaps $\mathbf{x}$ and $\mathbf{p}$ are related to each other in some deep unknown way. To understand this, we need to dig deep into what happens when we actually make a measurement.

Going back to our discussion on the qubit again. We have said that a general qubit state is given by Eq. (1.9), and that we can make measurements and either get an $\chi_\uparrow$ or $\chi_\downarrow$. Thus, clearly, $\chi_{\uparrow/\downarrow}$ states are possible after a measurement *but what we have not said is what our instruments actually tell us.*

Recall from section 1.2.1 that a NOT gate (or operator) Eq. (1.5) flips $\uparrow$ / $\downarrow$ to $\downarrow$ / $\uparrow$. What happens when we act on the qubit ψ with the NOT gate? Viz.

$$\begin{aligned}\hat{P}\psi &= \hat{P}\alpha\begin{pmatrix}1\\0\end{pmatrix}+\hat{P}\beta\begin{pmatrix}0\\1\end{pmatrix} &(1.20)\\ &= \alpha\begin{pmatrix}0\\1\end{pmatrix}+\beta\begin{pmatrix}1\\0\end{pmatrix}, &(1.21)\end{aligned}$$

i.e. we flip the probability amplitudes of measuring ↑ and ↓.

On the other hand, we can ask "are there any other operator that wouldn't change $\chi_\uparrow$ or $\chi_\downarrow$ when you act on them? $\hat{I}$ is obviously one, which is trivial. But are there any others? In fact, there is! Let's consider the following operator

$$\hat{N}=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{1.22}$$

Acting $\hat{N}$ on $\chi_\uparrow$ and $\chi_\downarrow$, does the following interesting things

$$\hat{N}\chi_\uparrow=\chi_\uparrow\ ,\ \hat{N}\chi_\downarrow=-\chi_\downarrow\ , \tag{1.23}$$

i.e. it doesn't change the matrix but instead "returns" a multiplicative value of $+1$ or -1 depending on whether we have ↑ or ↓ states. One way to think about Eq. (1.23) is that the operator $\hat{N}$ *defines the measurable states $\chi_\uparrow$ and $\chi_\downarrow$, with a "label" of either $+1$ or -1 respectively.* Now we will make an assertion – *what we actually measure is* not $\chi_{\uparrow/\downarrow}$ *but actually the values* $+1$ *or* -1. So, the state $\psi=\alpha\chi_\uparrow+\beta\chi_\downarrow$ jumps to $\chi_\uparrow$ if you measure $+1$ and $\chi_\downarrow$ if you measure -1. Let's call the instrument which measure $\chi_{\uparrow/\downarrow}$ Instrument A, and this instrument is "associated" with the measureable that is defined by the $\hat{N}$ operator.

Why do we bother to make such a distinction? Well, it turns out that there is more than one way to measure a qubit state! To see this, we now take advantage of the fact that the qubit, unlike the bit, can have states that are neither $\chi_\uparrow$ or $\chi_\downarrow$ but some "mixed up states". Consider the pair of states

$$\chi_+=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}\ ,\ \chi_-=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix} \tag{1.24}$$

and you can check that using the NOT gate $\hat{P}$ we discussed earlier this gets us

$$\hat{P}\chi_+=\chi_+\ ,\ \hat{P}\chi_-=-\chi_-\ . \tag{1.25}$$

Thus, the operator $\hat{P}$ *defines* the states χ_+ and χ_-, with labels $+1$ and -1. As it turns out, we can also build an instrument such that when a measurement is make, the state collapses to either χ_+ or χ_-. Let's call this new Instrument B. In other words, instead of Instrument A that measures $\chi_\uparrow$ and $\chi_\downarrow$ states which is associated with the $\hat{N}$ operator, we also have an Instrument B that measures χ_+ and χ_- states which is associated with the $\hat{P}$ operator. The instrument will also return $+1$ if the final state jumps into χ_- and -1 if it jumps into χ_-. In fact, the cat state we talked about earlier Eq. (1.15) is actually χ_+

$$\psi=\frac{1}{\sqrt{2}}\chi_\uparrow+\frac{1}{\sqrt{2}}\chi_\downarrow=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}=\chi_+\ . \tag{1.26}$$

Thus, if we have used Instrument B to measure the cat state, we will always get $+1$!

But which instrument, A or B, is giving us the right result? The answer is *both* – the lesson here is that for a given state, there are many ways to measure it. In fact, notice that

$$\chi_\uparrow=\frac{1}{\sqrt{2}}\chi_+-\frac{1}{\sqrt{2}}\chi_-\ ,\ \chi_\downarrow=\frac{1}{\sqrt{2}}\chi_++\frac{1}{\sqrt{2}}\chi_-\ , \tag{1.27}$$

so suppose we have a $\chi_\uparrow$ state such that Instrument A always measure $+1$, then if we then used Instrument B to make a measurement, we will have an even chance of measuring either $+1$ or -1. Once Instrument B has done its job, the state will jump to either χ_+ or χ_-.

This is weird! Here is a possible sequence of events : We start with a χ_+ state, make a measurement with Instrument A and, say, gets a $+1$ (50% chance) so the state jumps to $\chi_\uparrow$. If we keep measuring with A, we will always get $+1$ thereafter. But now, suppose, we switch to Instrument B, then we have a 50% each chance of measuring $+1$ and -1. Say, we get -1 so now the state jumps from $\chi_\uparrow$ to χ_-. Finally, if we switch back to Instrument A, and again, we get a 50% chance each of getting $+1$ or -1, which we then by chance get $+1$. In this case, the following sequence of events has occured

$$\chi_+ \xrightarrow{A=+1} \chi_\uparrow \xrightarrow{B=-1} \chi_- \xrightarrow{A=-1} \chi_\downarrow\ . \tag{1.28}$$

Of course, we could have measured differently (since the probabilities are 50% for +1 or -1). In general the following sequence of events can occur

$$\chi_+ \xrightarrow{A} \chi_{\uparrow/\downarrow} \xrightarrow{B} \chi_{+/-} \cdots \tag{1.29}$$

The moral of the story is that measurements with A can affect the measurements of B.

What if you try to measure the state *simultaneously* with both instruments A and B? You try to press the buttons of Instruments A and B at the same time, or at least as close to each other in time as possible. Now, depending on whether A goes first of B goes first, we have

$$\text{A goes first : } \chi_+ \xrightarrow{A=+1/-1} \chi_{\uparrow/\downarrow} \xrightarrow{B=+1/-1} \chi_{+/-}\ , \tag{1.30}$$

or

$$\text{B goes first : } \chi_+ \xrightarrow{B=+1} \chi_+ \xrightarrow{A=+1/-1} \chi_{\uparrow/\downarrow}\ . \tag{1.31}$$

If A goes first, then the final state is χ_+ or χ_-, but if B goes first, then the final state is $\chi_{\uparrow/\downarrow}$! In fact, it is more than that : notice that if B goes first, you will always measure $+1$ in the first measurement, but if A goes first then you have a 50/50 chance of getting $+1$ and -1. Thus *the order of the measurement is important!*

You can ask – well, what if I really try very hard to press the buttons of both instruments at the same time? The answer, sadly, is that nature finds a way to forbids it, no matter matter how hard you try. The way nature "forbids" you can be very funny sometimes. In the case of the qubit, if you try to press the buttons on Instruments A and B at the same time, nature will instead of measuring A and B, measure an alternate system between A and B[5]

Perhaps, you may not be surprised that making measurements will affect the measurements of other instruments – after all, experiments are not perfect. The main point here is that this is not a question about how "careful" you are in making a measurement, but that *there is no way you can avoid affecting the system.* In other words, quantum mechanic says that **measurements affect the system**. You are no longer omnipotent observers, but are forced to participate in the dynamics of the very thing you are measuring.

Another way of thinking about this is the following. When you make a measurement of using Instrument A, your state will collapse to either $\chi_\uparrow$ or $\chi_\downarrow$. After this measurement, somebody then asks you "well, can you tell me what Instrument A will measure", you can then confidently tell them "yes", because you now know that the system is in either $\chi_\uparrow$ or $\chi_\downarrow$ as you just measured it. However, if somebody then asks you "what will Instrument B measure", you no longer be so confident – indeed since there is a 50/50 chance of measuring $+1$ or -1 with B, you basically have no idea. Thus, you have gain 100% predictivity in A, and lost all predictivity in B.

[5] It will be akin to a new instrument C where $\theta = \pi/2$ in Eq. (1.33).

Finally, you can ask – OK, but *which instrument is actually the right one?* The answer is *both.* Both instruments are measuring different aspects of the state, if you like, A measures its ↑ / ↓-ness and B measures its $+/-$-ness. In fact, these two are not the only measureables of the humble qubit. As long as you can find an operator $\hat{O}$ and a pair of 2×1 matrices χ_λ such that

$$\hat{O}\chi_\lambda = \lambda\chi_\lambda \tag{1.32}$$

then the values of χ_λ are the measureables, or more precisely, the **observables**. Once a measurement of λ is made, the state collapses to the χ_λ corresponding to either $+1$ or -1. Thus in our example above, $\hat{O}$ is either $\hat{N}$ or $\hat{P}$ and λ is either $+1$ or -1 and χ_λ is either ↑ / ↓ or $+/-$. It can be shown that in general

$$\hat{O} = \begin{pmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} , \tag{1.33}$$

and

$$\chi_\lambda = \frac{1}{1+|\cot\theta+\csc\theta|^2}\begin{pmatrix} -\cot\theta-\csc\theta \\ 1 \end{pmatrix} \text{ or } \chi_\lambda = \frac{1}{1+|\cot\theta-\csc\theta|^2}\begin{pmatrix} -\cot\theta+\csc\theta \\ 1 \end{pmatrix} , \tag{1.34}$$

will satisfy $\hat{O}\chi_\lambda = \pm\chi_\lambda$ for any value of the angle θ. You can see that if you set $\theta = \pi$ you get $\hat{N}$ and if you set $\theta = \pi/2$ you get $\hat{P}$. In other words, for every θ, you can build an instrument which will measure $\lambda = \pm 1$ and results in the correspondent final state χ_λ. Mathematically, we call λ the **eigenvalues** and χ_λ their **eigenvectors**. This is actually the **Fourth Rule** of quantum mechanics. To be very mathematically precise, we will define it as follows (although you have basically seen the main gist of it)

Rule 4 (Observables): Suppose $\hat{O}$ obey

$$\hat{O} = \hat{O}^\dagger = (\hat{O}^T)^* \tag{1.35}$$

then $\hat{O}$ is a **Hermitian Matrix**. An operator associated with an observable $\hat{O}$ is Hermitian. Suppose there exists a set of states χ_λ such that

$$\hat{O}\chi_\lambda = \lambda\chi_\lambda , \tag{1.36}$$

then λ will be real and are called the **eigenvalues** while χ_λ are their respective **eigenvectors**, of $\hat{O}$. The result of a measurement of such an observable associated with $\hat{O}$ on a general state ψ yields one of the eigenvalues, and the state collapses (via Rule 3) into its associated eigenvector.

It might seems strange to you that what started off as a discussion on "what happens when we make a measurement" ended up on Rule 4 which tells you what observables are. However, hopefully you can see the point: for a quantum system, there are many ways to make physical observations of it, and not all of them actually independent. Mathematically, we say that observables which are not independent of each other as **non-commuting**. For example $\hat{P}$ and $\hat{N}$ are non-commuting. What that means is that if you calculate $\hat{P}\hat{N} - \hat{N}\hat{P}$ you will get a non-zero number – go on, try it!

1.3.1 How does the Uncertainty Principle arise?

We are now ready to discuss the uncertainty principle. We have, due to time constraints, have discussed quantum mechanics using the very simple two state qubit system. Of course, the universe is much more complicated than this, and is made out of more than dead/live cats (and its linear superpositions). Going back to the cannonball, our every day lives have taught us that it can be described, at least approximately, by its position $\mathbf{x}$ and momentum $\mathbf{p}$. What about a quantum cannonball?

Just like the qubit state which we can make measurements with either Instrument A or Instrument B (or indeed any other instruments defined by the general operator Eq. (1.33)), it turns out that $\mathbf{x}$ and

$\mathbf{p}$ are both observables corresponding to different operators $\hat{O}$. There is a "position operator" $\hat{X}$ and a "momentum operator" $\hat{P}$, with the corresponding eigenvalues (i.e. the equivalent of $+1$ and -1) $\mathbf{x}$ and $\mathbf{p}$. But just like the observables of the $\hat{P}$ and $\hat{N}$ operators, *making a measurement of position will affect the measurement of momentum* and *vice versa.* The amount each measurement of $\mathbf{x}$ affect on the measurement of $\mathbf{p}$ is given exactly by the Heisenberg Uncertainty Principle

$$\Delta\mathbf{x}\Delta\mathbf{p} \geq \frac{\hbar}{2}\ . \tag{1.37}$$

Thus, $\mathbf{x}$ and $\mathbf{p}$ are not independent, immutable quantities, but are actually different aspects of the quantum state which are related to each other. A measurement of one will affect the other, so there is no way we can simultaneously and independently determine the values of each to arbitrary accuracy. So if you try to pin down the value of $\mathbf{x}$ (i.e. make it more predictable), you will lose knowledge on $\mathbf{p}$ (i.e. becoming less predictable), just like the example of the qubit states above.

In fact, Eq. (1.37) is a formula that can be *derived* from everything you have learned in this lecture already[6]! If we have another 2 hour lecture, we can do it, but sadly we are out of time! Maybe one of you will try to do it in one of the assignment choices?

1.4 Assignment Topics

For this lecture, here are some topics you can choose to write your 2000 word essay and presentation on. * denote a challenging topic! You can also suggest topics to me, and we can discuss whether it will be appropriate!

- *The Bohr Quantization and the Bohr Hydrogen Atom Model* : Describe how the Bohr quantization condition provide a phenomenological model of the Hydrogen Atom.

- *Wave-particle Duality* : Using the two-slit experiment, describe why in quantum mechanics, objects possess both particle and wave properties.

- *The Planck Constant $\hbar$* : The fundamental constant of quantum mechanics is the $\hbar$. Explain how physicist **Max Planck** proposed its existence to explain the spectrum of blackbody radiation.

- *The Photoelectric effect*: Famously **Albert Einstein** won the Nobel Prize not for his theory of gravity, but for describing the photoelectric effect. Describe the photoelectric effect, and why it is such an important step in our understanding of quantum mechanics.

- *Derivation of the Heisenberg Uncertainty Principle** : Can you present a derivation of the uncertainty principle of the qubit observables associated with $\hat{P}$ and $\hat{N}$?

[6] The constant $\hbar$ we can filled in with doing experiments.

Chapter 2

The Principle of Least Action and Conservation Laws

> *The electron does anything it likes, [Richard Feymann] said. It just goes in any direction at any speed, . . . however it likes, and then you add up the amplitudes and it gives you the wavefunction. I said to him, You're crazy. But he wasn't.*
>
> Freeman Dyson

2.1 Why do things move the way they do?

So far in your physics studies, you have learned of Newton's 2nd Law of Motion for a single particle of mass m

$$m\ddot{\mathbf{x}} = \mathbf{F}\ , \tag{2.1}$$

under the action of a force $\mathbf{F}$. These forces include those you can apply on it (say with an engine), or it can be derived from some natural phenomenon such as gravity described by the Newton's Law of Gravity

$$\mathbf{F} = \frac{GMm}{r^2}\frac{\mathbf{x}}{|\mathbf{x}|}\ , \tag{2.2}$$

which is the force $\mathbf{F}$ on the particle m at a distance $r \equiv |\mathbf{x}|$ exerted by an object of mass M, and G is just Newton's constant $G = 6.673 \times 10^{-11}\mathrm{Nm}^2\mathrm{kg}^{-2}\mathrm{s}^{-2}$. Another kind of force you might have learned is the electrostatic **Coulomb Force**,

$$\mathbf{F} = \frac{kqQ}{r^2}\frac{\mathbf{x}}{|\mathbf{x}|}\ , \tag{2.3}$$

between two particles with charges q and Q, and $k = 8.99 \times 10^{10}\ \mathrm{Nm}^2\mathrm{C}^{-2}$ is the Coulomb's constant. Once you have the forces, you can solve Eq. (2.1) and calculate the **dynamics** of your particle, at least in the classical viewpoint. Sometimes you can write Newton's Law of motion Eq. (2.1) as

$$m\ddot{x} = -\nabla V\ , \tag{2.4}$$

where V is called the **potential**. In terms of Eq. (2.4), if the force is driven by gravity, then we can express Newton's law of gravity using the **gravitational potential** defined as

$$V(r) = -\frac{GMm}{r} \ . \tag{2.5}$$

The moniker "potential" literally means the potential energy of the system. In the uniform gravitational field of the surface of the earth, $V(\mathbf{x}) = mgz$, where $g = 9.81$ ms^{-2} is just the acceleration pointing in the $-z$ direction.

Similarly, we can express the electrostatic force using the **Coulomb potential** as

$$V(r) = -\frac{kQq}{r} \ . \tag{2.6}$$

Notice that the potential only depend on the distance r, and not on the direction unlike the force – we get the "vector" by using the gradient derivative ∇. Forces which can be re-expressed using a potential V are known as **conservative forces**. The name may be a bit mysterious to you for now – what is being "conserved"? We will come back to that in the 2nd part of our lecture in section 2.3.

Equations such as Eq. (2.1) and its equivalent Eq. (2.4), which tells us how things move or their **dynamics**, are called **Equations of Motion**. In many ways, physics is the study of how everything in our universe move and evolve, and our attempt to predict the future and understand its past. Thus a large part of the job of the theoretical physicist is to figure out the equations of motion for all the objects in the universe – cannonballs, planets, electrons, quarks, photons, all the way to spacetime and the entire universe itself.

For example, the equations of motion that tell us how light move is given to us by the **Maxwell Equations**

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \partial_t \mathbf{E}, \tag{2.7}$$

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \tag{2.8}$$

$$\nabla \cdot \mathbf{E} = \frac{\rho_c}{\epsilon_0}, \tag{2.9}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{2.10}$$

or, if you have learned some more fancy math, can be condensed into the elegantly looking equation

$$\partial_\mu F^{\mu\nu} = j^\nu \ . \tag{2.11}$$

Meanwhile, the equation that tell us how space and time move is given to us by the **Einstein equation**

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{2.12}$$

Don't worry if you have not seen these equations before or don't understand what they mean! This is just to tell you that there are other equations that describe how things move other than Newton's equation (which itself is actually not fully accurate). There are other equations that tell us how other things like quarks, neutrinos move.

Physicists do spend a lot of time trying to derive these equations from some more fundamental theory, a "mother theory" if you like, for example like string theory. At this moment though, we don't know exactly how this more fundamental "mother theory" would look like – in fact we will discuss in Chapter 5 why it is so hard to find such a theory. So for the moment at least, physics is a collection of seemingly independent set of equations of motion that describe how things in the universe move.

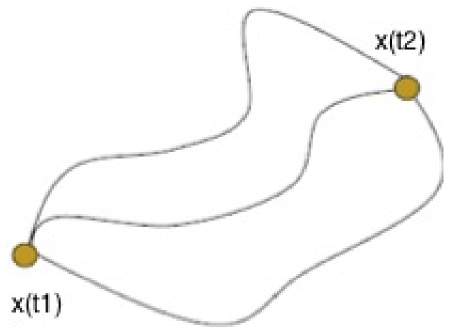


Figure 2.1: We know the starting location $\mathbf{x}(t_1)$ and the ending location $\mathbf{x}(t_2)$, which path will the particle take?

2.1.1 A magical way

On the other hand, there *is* a way to look at all these independent equations such that the equations seem to be "on the same footing". Those words don't mean too much right now, so the best way to see what it means is to dive right in.

As we have discussed in Chapter 1, if you want to calculate the trajectory of a particle of mass m, you need to specify its initial state which is its initial position $\mathbf{x}(t)$ and initial velocity $\mathbf{v}(t)$, and then solve this using Newton's 2nd law of motion.

Let's ask another question. Suppose we know the starting location $\mathbf{x}(t_1)$ and the ending location $\mathbf{x}(t_2)$ of the particle. Consider *all the possible trajectories or paths*, each possible path will then has a "history" $\mathbf{x}(t)$ and $\mathbf{v}(t)$ – but only one of the history will be the "right" one, at least classically[1]. Which one is it?

Let's do something that seems which will sound crazy at first: we will take each possible path, at each point in time t of this possible path, we will subtract the kinetic energy $(1/2)m|\dot{\mathbf{x}}(t)|^2$ from its potential energy $V(\mathbf{x})$, and then we add up all the difference over the entire path, i.e. In equation form, we do the following

$$S[\mathbf{x}(t)] = \int_{t_1}^{t_2} dt \left(\frac{1}{2}m|\dot{\mathbf{x}}(t)|^2 - V(\mathbf{x}(t))\right) \ , \tag{2.13}$$

The S is called the **action**. For each possible path $\mathbf{x}(t)$, we can calculate its action. In words this is

Action = Sum of (Kinetic Energy - Potential Energy) over the entire path

Now here is the astounding claim:

The path that the particle will take is an extremum of the action.

Usually, the extremum here is a *minimum*, which will be the case here. So, our claim here is the particle will take the path which minimizes the action. Let's prove it.

You have learned how to find the extremum of a function, say $f(x)$ – simply take the derivative, set it to zero $df/dx = 0$, and then solve the resulting equation. Thus for example if $f(x) = (x-1)^2$, then $df/dx = 2(x-1)$, and setting $2(x-1) = 0$ we find that the minimum is at $x = 1$ as we expected. However, the action $S[\mathbf{x}(t)]$ is not a function, but a **functional**. In other words, it is a function of a function – so different trajectories $\mathbf{x}(t)$ will be a different function and thus $S[\mathbf{x}(t)]$ will give you different values for different paths. How do we find the extremum of a functional? What we need to learn is really **functional derivative**, but that requires a bit more mathematical background. So instead, let's just prove it directly here!

[1] As we will soon see, quantum mechanically, it will be a lot more interesting!

Consider a path $\mathbf{x}(t)$. Now, let's consider another path that is infinitisimally close, but not exactly the same as this path, say $\bar{\mathbf{x}}(t)$. We will call the difference between the two paths to be $\delta\mathbf{x}(t)$, i.e.

$$\bar{\mathbf{x}}(t) - \mathbf{x}(t) \equiv \delta\mathbf{x}(t) \ . \tag{2.14}$$

We say that the path $\bar{\mathbf{x}}(t)$ is *slightly perturbed* from the original path $\mathbf{x}(t)$. Since the start and end points are the same, $\delta\mathbf{x}(t_1) = \delta\mathbf{x}(t_2) = 0$. We can calculate the actions for both these paths

$$S[\mathbf{x}(t)] = \int_{t_1}^{t_2} dt \left(\frac{1}{2}m|\dot{\mathbf{x}}|^2 - V(\mathbf{x})\right) \ , \tag{2.15}$$

and for the perturbed path

$$S[\mathbf{x}(t) + \delta\mathbf{x}(t)] = \int_{t_1}^{t_2} dt \left(\frac{1}{2}m(\dot{\mathbf{x}}^2 + 2\dot{\mathbf{x}}\cdot\delta\dot{\mathbf{x}} + \delta\dot{\mathbf{x}}^2) - V(\mathbf{x}+\delta\mathbf{x})\right) \ . \tag{2.16}$$

Notice that in Eq. (2.15) and Eq. (2.16), we have not explicitly written the argument t e.g. $\mathbf{x}$ instead of $\mathbf{x}(t)$ etc. for simplicity, but it is important to keep in mind that the argument is there. The difference between the two action is

$$\delta S[\mathbf{x}(t)] = S[\mathbf{x}(t) + \delta\mathbf{x}(t)] - S[\mathbf{x}(t)] \ . \tag{2.17}$$

We now claim : the path in which $\delta S = 0$ is the extremum of the action. We can motivate it as follows – to find the extremum of a function, as we mentioned above, we take the derivative and set it to zero $df/dx = 0$. But recall the definition of the derivative

$$\frac{df}{dx} = \lim_{\delta x \to 0} \frac{f(x) - f(x+\delta x)}{\delta x} \ , \tag{2.18}$$

so setting the derivative to be zero is the same as setting $\delta f = f(x) - f(x+\delta x) = 0$. Comparing this equation to Eq. (2.17), we see that in the case of the functional, the analogy is $\delta f \to \delta S$, and $\delta x \to \delta\mathbf{x}(t)$. So to find the path $\mathbf{x}$ which extremizes S, we want to solve the equation $\delta S = 0$.

Let's calculate δS. Eq. (2.15) is easy – just leave it as it is. On the other hand, Eq. (2.16) is a bit more tricky – we have to Taylor expand the potential term

$$V(\mathbf{x}+\delta\mathbf{x}) = V(\mathbf{x}) + \nabla V\cdot\delta\mathbf{x} + \mathcal{O}(\delta\mathbf{x})^2. \tag{2.19}$$

For the last term $\mathcal{O}(\delta\mathbf{x})^2$, since we have assumed that $\delta\mathbf{x}$ is small, then $\delta\mathbf{x}^2$ is *even smaller*, thus we can neglect it. Putting all the equations together, we see that the only terms that survived all the subtraction are

$$\delta S[\mathbf{x}(t)] = \int_{t_1}^{t_2} dt\,(m\delta\dot{\mathbf{x}}\cdot\dot{\mathbf{x}} - \nabla V(\mathbf{x})\cdot\delta\mathbf{x}) \ . \tag{2.20}$$

The second term we can just leave it as it is, but the first term we can *integrate by parts*, i.e.

$$\delta\dot{\mathbf{x}}\cdot\dot{\mathbf{x}} = \frac{d}{dt}(\dot{\mathbf{x}}\cdot\delta\mathbf{x}) - \ddot{\mathbf{x}}\cdot\delta\mathbf{x} \ , \tag{2.21}$$

and then plug this back into Eq. (2.20)

$$\delta S[\mathbf{x}(t)] = \int_{t_1}^{t_2} dt\,(-m\ddot{\mathbf{x}} - \nabla V(\mathbf{x}))\cdot\delta\mathbf{x} + [\dot{\mathbf{x}}\cdot\delta\mathbf{x}]_{t_1}^{t_2} \ . \tag{2.22}$$

But since $\delta\mathbf{x}(t_1) = \delta\mathbf{x}(t_2) = 0$ as the starting and ending points are the same, we are left with our final answer

$$\delta S[\mathbf{x}(t)] = \int_{t_1}^{t_2} dt\,(-m\ddot{\mathbf{x}} - \nabla V(\mathbf{x}))\cdot\delta\mathbf{x} \ , \tag{2.23}$$

which must be set to zero. The only way the integral Eq. (2.23) can be zero is that *the integrand* is zero for the entire path $\mathbf{x}(t)$ which means that the following equation must be satisfied along the entire path

$$m\ddot{\mathbf{x}} = -\nabla V(\mathbf{x}) \ . \tag{2.24}$$

But this is literally Newton's 2nd law of Motion Eq. (2.4) as we have just described earlier! The requirement that particle travel on a trajectory that extremizes the action is completely equivalent to saying that it must obey Newton's 2nd law of Motion! It's like witchcraft.

This principle – that the path of the particle will be the one that extremizes the action – is called **The Principle of Least Action**, although it is named slightly wrong as it doesn't have to be "least action", it could be "maximum action". So sometimes it is just called **the action principle**. It is probably the most boring name for the most important principle in physics that you would ever learn.

How does the action principle "work"? Let's consider a simple example of our cannonball again. The cannonball is shot towards the sky at some angle with some initial velocity $\dot{\mathbf{x}}(t_0)$ into the air at some angle. The cannonball feels the gravity of earth, so as it goes into the sky it will gain potential energy $V(\mathbf{x}) = mgz$ as we have discussed just now. What should the particle do according to the action principle? The principle says that we should try to minimize the difference between the kinetic energy (KE) and the potential energy (PE). So since the particle started with some KE, it wants to go up to gain some PE, but it doesn't want to go straight up because eventually the PE will become more than the KE. So it will turn around at the point when PE=KE, i.e. when the difference is zero, and then head back down to earth. Thus the action principle suggests that the particle would want to trace a parabola in the sky.

The action principle changes the way we look at how particles move. In the old Newtonian way, we "follow" the particle around. At every time t, the particle wants to figure out "where to next", and it looks at the potential and say "aha! the potential is steep here so I should accelerate towards that point". But the action principle says that, before the particle even moved an inch, it "sniffs out" all possible paths, and then once it has done that, chose the one that extremizes the action. It seems to "know everything".

2.1.2 The Euler-Lagrange Equation

The integrand of the action Eq. (2.13), is called its **Lagrangian**

$$L[\mathbf{x}(t), \dot{\mathbf{x}}(t), t] = \frac{1}{2}m|\dot{\mathbf{x}}(t)|^2 - V(\mathbf{x}(t)) \ , \tag{2.25}$$

named after the Italian-French mathematician **Joseph-Louis Lagrange** (or his old Italian name Giuseppe-Luigi Lagrangia). It can be shown, although we won't show it here, that if we apply the action principle, we will get the following **Euler-Lagrange Equation**

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right) - \frac{\partial L}{\partial \mathbf{x}} = 0 \ . \tag{2.26}$$

Let's check for the action of the single particle Eq. (2.25) above. The first term of Eq. (2.26) is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right) = \frac{d}{dt}(m\dot{\mathbf{x}}) = m\ddot{\mathbf{x}} \ , \tag{2.27}$$

while the second term is

$$\frac{\partial L}{\partial \mathbf{x}} = \nabla V(\mathbf{x}) \tag{2.28}$$

and hence we recover the Newton's 2nd Law of Motion as promised. The really powerful thing about the Euler-Lagrange equation is that it automatically generalizes to any case where there is more than one particle. So consider a system with n particles moving under some very complicated potential $V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n)$, the Lagrangian for this system is still the sum of all the kinetic energy minus the total potential energy

$$L = \sum_{i=1}^{i=n} \left(\frac{1}{2} m |\dot{\mathbf{x}}_i(t)|^2 \right) - V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n) \ . \tag{2.29}$$

The equation of motion for each particle is then the solution to its individual Euler-Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}_i}\right) - \frac{\partial L}{\partial \mathbf{x}_i} = 0 \ . \tag{2.30}$$

2.2 Is this really a better way?

Of course we have not really “derived” the Newton’s 2nd Law of motion (for that, we would need to understand General Relativity), so the action principle is simply an equivalent formulation of the same thing. So why do we care? Practically, there is a nice reason – super complicated systems can actually be easily solved using the action principle. For example, consider the problem of the coupled pendulum in Fig. 2.2. Now, if you try to set up force balance diagrams as you were taught in Mechanics, it will be a rather tough problem. The action principle, however, tell us to write down the Lagrangian, which is the KE minus the PE. The KE of this system is simply

$$KE = \frac{1}{2} m \dot{x}_1^2 + \frac{1}{2} m \dot{x}_2^2 \ . \tag{2.31}$$

Meanwhile the PE is

$$V(x_1, x_2) = mgl(1 - \cos\theta_1) + mgl(1 - \cos\theta_2) + \frac{1}{2} k (x_1 - x_2)^2 \ , \tag{2.32}$$

where the first two terms are the gravitational potential energy of the two pendula, while the 3rd term is the PE stored in the spring. Assuming that the pendulum motion is small, we can use the small angle formula $\cos\theta = 1 - \theta^2/2 + \ldots$, and then $l\theta_i^2/2 = x_i^2/l$. Plugging all these in, we get the Lagrangian

$$L = KE - PE = \frac{1}{2} m \dot{x}_1^2 + \frac{1}{2} m \dot{x}_2^2 - \frac{mg}{l}(x_1^2 - x_2^2) + \frac{1}{2} k (x_1 - x_2)^2 \ . \tag{2.33}$$

To find the equations of motion of the two pendula, we simply calculate their individual Euler-Lagrange equation Eq. (2.30). E.g. for pendulum 1, we have

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}_1}\right) - \frac{\partial L}{\partial \mathbf{x}_1} = m\ddot{x}_1 + \frac{mgx_1}{l} - k(x_2 - x_1) = 0 \ , \tag{2.34}$$

and similarly for pendulum 2, its equation of motion is

$$m\ddot{x}_2 + \frac{mgx_2}{l} + k(x_2 - x_1) = 0 \ . \tag{2.35}$$

I challenge you to try to solve this equation using your old Mechanics force balance diagrams and see how much harder it would be!

However, beyond this simple practical reason, there are three deeper reasons why the action principle is so much better than the old way. Let’s look at two of them here, and we’ll save the 3rd for its own section 2.3

2.2.1 A Lagrangian for Everything

The first reason is that it provides a powerful way of looking at very different physical laws using the same formalism. We discussed earlier that how things move in physics are expressed in terms of their

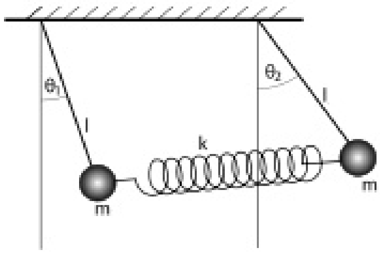


Figure 2.2: A coupled pendulum system.

dynamical equations such as Maxwell equations or Einstein's theory of General Relativity. In fact, *all known physical laws can be recast in terms of an action principle* – in fact the entire known laws of physics can be written as the following Lagrangian[2]

$$\mathcal{L} = \underbrace{\frac{R}{16\pi G}}_{\text{Gravity}} - \underbrace{\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{Yang}-\text{Mill}} + \underbrace{i\bar{\psi}\gamma^{\mu}D_{\mu}\psi}_{\text{Dirac}} + \underbrace{|D_{\mu}h|^2 - V(|h|)}_{\text{Higgs}} + \underbrace{h\bar{\psi}\psi}_{\text{Yukawa}} \quad , \tag{2.36}$$

which is known as the **Standard Model Lagrangian with Gravity**. The symbols are all complicated and you don't have to worry about it, but they briefly describe the following. The first "Gravity" term is Einstein's Theory of General Relativity – all of gravity is encoded in the (so-called "Ricci") term. The 2nd "Yang-Mill" term describes all the known forces other than gravity: the electromagnetic (including the Maxwell Equations), the weak and the strong nuclear forces. The 3rd "Dirac" term describes all the known particles – electrons, neutrinos, quarks etc. Finally, the last two "Higgs" and "Yukawa" terms describe the Higgs particle and its interactions with the rest of the particles which results in those particles gaining mass.

The fact that we can write down such an action or Lagrangian tells us that for each of these seemingly different things, nature wants to extremize something. Indeed, nowadays, when theoretical physicists try to discover a new force or understand some dynamics, they don't really try to write the equation of motion; instead they start with trying to figure out what is being extremized, and then write down the Lagrangian for that.

2.2.2 Quantum Sniffing

While we have not explicitly said it, what we have been discussing earlier are all in the classical viewpoint – the particles obey Newton's equation, we talked about the path of the particle as completely determined, "the particle picks the path that extremizes the action". However, as we have already discussed in the previous Chapter 1, the world is quantum, so how does this work in the quantum world?

The answer is : the whole thing works even more crazily and more amazingly. In fact, in the quantum world, *the particle takes all possible paths*. And by "all possible", I mean, it can also do completely strange things like go back in time, or go to the Sun and back, or do loops around the Earth! But the *probability* of it doing crazy things are small, while the chance of it doing something like what we would expect (e.g. near its "classical" path) would be big. From our discussion in Chapter 1, we learned that according to Born's rule, the probability of an even occurring is given by

$$\text{Prob} = |\chi^{*T}\psi|^2 \tag{2.37}$$

[2] Here in this case, it is really the *Lagrangian density* hence the scripted $\mathcal{L}$ instead of plain L, but it will take us too far afield, so don't worry about it!

where χ is the event you want to see occurring, and ψ is the starting state of the system. In the case of a particle traveling from $\mathbf{x}_1$ to $\mathbf{x}_2$, then χ is the state where the particle is in $\mathbf{x}_2$, while ψ is the state where the particle in $\mathbf{x}_1$. We can express this probability as

$$\text{Prob} = |(\text{particle in } \mathbf{x}_2)^*(\text{particle in } \mathbf{x}_1)|^2 \equiv |A|^2 \tag{2.38}$$

where A is then the probability amplitude as we have discussed before. Now, as it turns out, A is given by the following formula

$$A = \sum_{\text{all paths}} e^{iS/\hbar} \ , \tag{2.39}$$

where S is the action for each possible path. Each of the paths is called a **history**, and Eq. (2.39) is known as a **sum over histories**, or a **Feynman Path Integral**.

This crazy formula, as the name implied, is discovered by **Richard Feynman** following a suggestion by **Paul Dirac**. Not only does a particle "sniffs out" all the paths and then choose the one that extremizes the action, every path has a non-zero probability of being actually traveled by the particle! When we make an observation, the particle "chooses" one of the paths to travel according to Rule 3 of quantum mechanics, i.e. the particle's quantum state "collapses" into the observed path.

Now, why does the particle like to "collapse" close to the "classical path" – remember the uncertainty in the path as given by the Uncertainty Principle is very small? We can see this fact from Eq. (2.39). Suppose S_{cl} is the action of the classical path, then paths far away from it will wildly vary and cancel each other (remember that the value of S gives us the phase of the exponent, not its amplitude, in the sum) while paths close to the classical path will reinforce the sum.

Why does nature work this way? No one really knows. We will see that this fact is going to play an incredibly crucial role in our final Chapter 5

2.3 Conservation Laws and Symmetries

Finally, the third reason why the action principle is so powerful is that it gives us a way to understand **conservation laws**. You must have learned some conservation laws before – conservation of momentum, conservation of energy, conservation of angular momentum, conservation of electric charges etc. You are probably not really told *why* these things are conserved (or are they really conserved?) – there is a deep underlying principle before these conservation laws which we will now discuss.

We will begin with a very simple case. Consider a particle of mass m moving freely, i.e. the potential $V(\mathbf{x}) = 0$. The Lagrangian for this particle is then simply

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 \ . \tag{2.40}$$

Using the Euler-Lagrange equation Eq. (2.26), we can calculate

$$\frac{d}{dt}(m\dot{\mathbf{x}}) = 0 \ . \tag{2.41}$$

When we see an equation that looks like this, i.e.

$$\frac{d}{dt}(\text{something}) = 0 \ , \tag{2.42}$$

this means that the "something" is not changing with time, hence it is a *conserved quantity.* In the simple example above, we can integrate Eq. (2.41) to get $m\dot{\mathbf{x}} = \text{const}$. In other words, the **momentum** is conserved. If we just think a little bit about it, this is of course not a surprise – a free particle feels no potential, so there is no force acting on it, and hence we have learned from Newton's 1st law that the momentum must then stay constant.

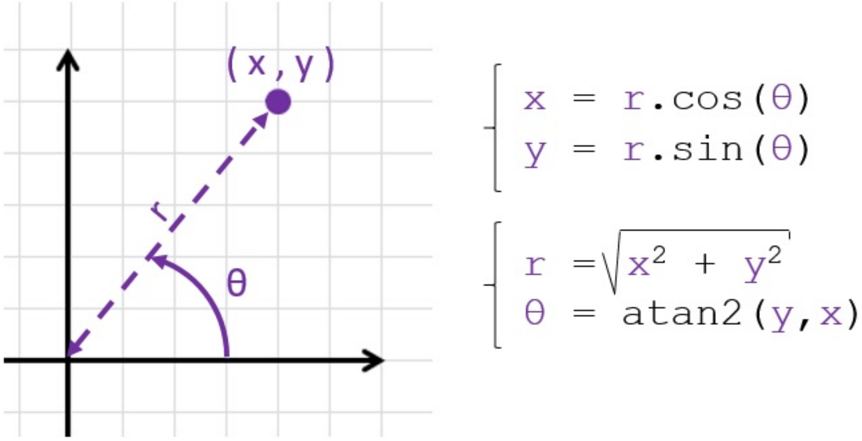


Figure 2.3: Polar Coordinates

Before we move on, let's define something that will be very useful. We can see from Eq. (2.41) that, if we write the momentum $\mathbf{p} = m\dot{\mathbf{x}}$, then Eq. (2.41) is

$$\frac{d\mathbf{p}}{dt} = 0 \tag{2.43}$$

which means that

$$\frac{\partial L}{\partial \dot{\mathbf{x}}} = \mathbf{p} \ . \tag{2.44}$$

In this case, $\partial L/\partial \dot{\mathbf{x}}$ is actually our good old standard linear momentum. However, this is not always the case. Nevertheless, we can *define* whatever is given by the partial derivative $\partial L/\partial \dot{\mathbf{x}}$ to be a "generalized momentum", i.e.

$$\mathbf{P} \equiv \frac{\partial L}{\partial \dot{\mathbf{x}}} \ , \tag{2.45}$$

where we have used a capital $\mathbf{P}$ to distinguish it from the our good old momentum. This generalized momentum is called **canonical momentum**, which is a completely unhelpful sounding name.

Let's now look at a slightly more complicated example. Consider a particle on a 2D plane moving under the influence of a potential $V(r)$ which is only a function the distance $r \equiv |\mathbf{x}|$ from the origin $r = 0$, and not dependent on the direction. Thus the potential is the same whichever direction you look at it – we say that the potential is **rotationally symmetric**.

The Lagrangian of this system, in 2D polar coordinates (r, θ) (Fig. 2.3), is

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V(r) \ . \tag{2.46}$$

In the polar coordinate system, we can calculate its canonical momenta associated with r and θ, e.g.

$$P_\theta = \frac{\partial L}{\partial \dot{\theta}} \ , \ P_r = \frac{\partial L}{\partial \dot{r}} \ , \tag{2.47}$$

The Euler-Lagrange equation for the θ coordinate is then

$$\frac{d}{dt}P_\theta = \frac{dV}{d\theta} = 0 \ , \tag{2.48}$$

which imply that $P_\theta = mr^2\dot{\theta} =$ const is a conserved quantity. But P_θ is nothing but the particle's **angular momentum**, which is conserved because the potential $V(r)$ does not depend on θ. Indeed, this is explicit in Eq. (2.48), since $V(r)$ is not a function of θ, $dV/d\theta = 0$.

These two examples suggest that perhaps conservation laws are consequences of the form of the Lagrangian. For example, in the Lagrangian of the free particle Eq. (2.40), if we *shift* the location of the particle by a small constant ϵ, $\mathbf{x} \to \mathbf{x}' + \epsilon$, then the kinetic term becomes $\dot{\mathbf{x}} \to \dot{\mathbf{x}}'$, and thus the Lagrangian

$$L[\mathbf{x}, \dot{\mathbf{x}}] = \frac{1}{2}m\dot{\mathbf{x}}^2 \to L[\mathbf{x}', \dot{\mathbf{x}}']] = \frac{1}{2}m\dot{\mathbf{x}}'^2 \tag{2.49}$$

remains *functionally* the same. A shift in the $\mathbf{x}$ coordinate such as this is called **a translation**. In words, we say that the "Lagrangian remains **invariant** under the translation in $\mathbf{x}$", or that "the Lagrangian possesses a **translation symmetry**".

For the second example, you might have guessed that the rotationally symmetric potential will now come into play, and you would be right. So if we now rotate the particle by a small constant, i.e. we perform the following transform

$$\theta \to \theta' + \epsilon_\theta \ , r \to r' \tag{2.50}$$

then

$$\dot{\theta} \to \dot{\theta}' \ , \dot{r} \to \dot{r}'. \tag{2.51}$$

The Lagrangian is then clearly invariant under this transformation

$$L[\theta, \dot{\theta}, r, \dot{r}] = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V(r) \to L[\theta', \dot{\theta}', r', \dot{r}'] = \frac{1}{2}m\left(\dot{r'}^2 + r'^2\dot{\theta'}^2\right) - V(r') \tag{2.52}$$

and we say that the "Lagrangian has a **rotational symmetry**".

As it turns out, conservation laws are intimately tied to the fact that Lagrangian possesses certain symmetries! This is proven by **Emmy Noether** in 1915, a remarkable woman who overcame the huge gender bias in academia at that time to become one of the most influential mathematical physicist of all time. We will now demonstrate the proof in the next section 2.3.1.

2.3.1 Noether's Theorem

Emmy Noether proved the following theorem :

For every continuous symmetry in the Lagrangian, there is a conservation law

We begin with a general Lagrangian with a set of coordinates q_i and its associated canonical momenta $P_i = \partial L/\partial \dot{q}$. Let's now shift each of the coordinates q_i by a small amount,

$$q_i \to q_i' + \delta q_i \tag{2.53}$$

where the shift

$$\delta q_i \equiv f_i(\mathbf{q})\epsilon \ , \tag{2.54}$$

may depend on the other coordinates through the arbitrary functions f_i (i.e. there is a f_i function for each variable q_i). This is a *continuous* transformation – you can make ϵ as small as you want until the transformation is infinitisimally small. So in our examples in the previous section, $f_i = 1$ (note that this is a vector, i.e. $\mathbf{f} = f_i = \mathbf{1}$), the first example, and $f_r = 0$ and $f_\theta = 1$ for the second example[3]. The change in the Lagrangian, δL due to these transformation can be calculated easily using the chain rule, and it is (we have dropped the primes from q_i for simplicity)

$$\delta L = \sum_i \left(\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i + \frac{\partial L}{\partial q_i}\delta q_i \right) \ . \tag{2.55}$$

Now the first term of Eq. (2.55), using our definition of the canonical momentum Eq. (2.45), we get

$$\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i = P_i\delta\dot{q}_i \ . \tag{2.56}$$

[3] You can do the 2nd example with cartesian coordinates (x, y) instead of polar coordinates (r, θ), and you will find in that case that the rotation transform imply that $f_x = y$ and $f_y = -x$.

For the 2nd term of Eq. (2.55), we use the Euler-Lagrange equation Eq. (2.26) to write $\partial L/\partial q_i = dP_i/dt$ to get

$$\frac{\partial L}{\partial q_i}\delta q_i = \dot{P}_i \delta q_i \ . \tag{2.57}$$

Inserting Eq. (2.56) and Eq. (2.57) back into Eq. (2.55), we then have

$$\begin{aligned} \delta L &= \sum_i P_i \delta\dot{q}_i + \dot{P}_i \delta q_i \\ &= \frac{d}{dt}\sum_i P_i \delta q_i \ . \end{aligned} \tag{2.58}$$

Using Eq. (2.54), we can then write this as

$$\frac{d}{dt}\sum_i P_i f_i(\mathbf{q}) = \delta L \ , \tag{2.59}$$

where we have canceled the ϵ as it is a constant. Now Noether told us that if the Lagrangian is invariant under the transformations, then $\delta L = 0$, and hence we get

$$\frac{d}{dt}\sum_i P_i f_i(\mathbf{q}) = 0 \ , \tag{2.60}$$

but this equation is simply $d/dt(\text{something}) = 0$, which we already learned that this means that the "something" is a conserved quantity! In other words, if the Lagrangian is invariant under the transformations Eq. (2.53), then there is a conservation law that says that the quantity

$$Q = \sum_i P_i f_i(\mathbf{q}) \ , \tag{2.61}$$

is always constant in time!

Let's see how this theorem stacks up against our examples above! In the first Lagrangian Eq. (2.40), there is only one variable $\mathbf{x}$ with its canonical momentum $\mathbf{P}$ so the conserved quantity $Q = \mathbf{P}$, as expected the linear momentum is conserved as we have already shown above. In the second Lagrangian Eq. (2.47), we have two variables $q_1 = \theta$ and $q_2 = r$, so $f_1 = P_\theta$ and $f_2 = 0$. Thus, the conserved quantity is then $Q = P_\theta$ which is the angular momentum as we have discussed earlier.

These are very simple examples of course – in general the symmetries can be a lot more complicated with highly non-trivial $f_i(\mathbf{q})$ – we will leave these examples for you to study when you take a full course!

2.3.2 Conservation of Energy

Finally, we will discuss the conservation of energy – what is the symmetry associated with it? As it turns out, the conservation of energy is associated with the **time translation symmetry**. To be precise, the Lagrangian must be invariant under *explicit* time translation symmetry $t \to t + f(\mathbf{q})\epsilon$. This needs a bit of care to explain. There are two ways the Lagrangian can depend on time – **implicit** and **explicit**. The variables in the Lagrangian are all functions of time – so as time changes, the variables $q_i(t)$ and $\dot{q}_i(t)$ also changes – in this case we say that the Lagrangian is *implicitly* depending on time through the functions $q_i(t)$ and $\dot{q}_i(t)$. In our two examples above, the Lagrangians are implicit functions of time.

On the other hand, the Lagrangian can also depend *explicitly* on time, if the variable t explicitly appear in the Lagrangian. For example, consider the case of our coupled pendulum in section 2.2. There the spring constant k does not depend on time. Suppose however, if we heat up the spring during the experiment, and due to the expansion of the spring, k changes as a function of time, i.e. $k \to k(t)$. Then its Lagrangian Eq. (2.33) would become

$$L = KE - PE = \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2 - \frac{mg}{l}(x_1^2 - x_2^2) + \frac{1}{2}k(t)(x_1 - x_2)^2 \ . \tag{2.62}$$

where k is now a function of time. This means that the Lagrangian now *explicitly* depends on time through $k(t)$ in addition to being *implicitly* dependent on time through its other variables. Noether's theorem now state that the energy of the system is conserved if the Lagrangian is invariant under an *explicit time translation.* Let's see how this works.

In general, the Lagrangian can now depend on its variables and also the explicit time variable t, so we write the functional as

$$L[\mathbf{q}, \dot{\mathbf{q}}, t] \ . \tag{2.63}$$

Taking the *total derivative* of Eq. (2.63), we get

$$\frac{dL}{dt} = \sum_i \left[\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i \right] + \frac{\partial L}{\partial t} \ , \tag{2.64}$$

noting that the last term is a *partial derivative* on L with respect to t *which will vanish unless the Lagrangian has an explicit dependence on time.* Using the Euler-Lagrange Eq. (2.26), we can rewrite each term in the sum as

$$\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i = \dot{P}_i\dot{q}_i + P_i\ddot{q}_i = \frac{d}{dt}\left(P_i\dot{q}_i\right) \ , \tag{2.65}$$

which we can insert back into Eq. (2.64) to get

$$\frac{dL}{dt} = \frac{d}{dt}\sum_i \left(P_i\dot{q}_i\right) + \frac{\partial L}{\partial t} \ . \tag{2.66}$$

Now, we define a quantity, called the **Hamiltonian** H as

$$H \equiv \sum_i \left(P_i\dot{q}_i\right) - L \ , \tag{2.67}$$

then Eq. (2.66) becomes

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} \ . \tag{2.68}$$

This Eq. (2.68) tells us that, if L has no explicit dependence on time, then $dH/dt = 0$, and hence the Hamiltonian is a conserved quantity. The Hamiltonian is actually an exact and mathematical precise definition of "energy" in a dynamical system, and thus Noether's theorem says that if the Lagrangian is invariant under time translations – i.e. it has no explicit dependence on time – then the energy of the system is conserved.

To check again, in example of the free particle Eq. (2.40), its Hamiltonian is $H = \mathbf{P}\cdot\dot{\mathbf{x}} - L = m(\dot{\mathbf{x}})^2 - (1/2)m(\dot{\mathbf{x}})^2 = (1/2)m(\dot{\mathbf{x}})^2$ which says that the total energy of the system is given by the kinetic energy of the particle as we expected since there is no potential. Noether's theorem then tells us that since the Lagrangian Eq. (2.40) is not explicitly dependent on t, this energy must be conserved – just like your high school teachers have told you they would.

This symmetry principle is an extremely powerful tool – every conservation law that we know off has an underlying symmetry associated with it. Beyond momentum and energy, conservation of electric charges, neutrons, protons and other more esoteric things like quarks or lepton number. Indeed, just like the how physicists now think about the dynamics of physics in terms of the action principle, they now think of the content of physics in terms of symmetry principles.

2.4 Assignment Topics

For this lecture, here are some topics you can choose to write your 2000 word essay and presentation on. * denote a challenging topic! You can also suggest topics to me, and we can discuss whether it will be appropriate!

- *The Principle of Least Time*: A precursor to the Principle of Least Action is Fermat's Principle of Least Time, which states that "a light ray propagates between two points so as to minimise its travel time". Describe this principle, and use it to demonstrate the phenomenon of *light refraction* and *Snell's Law*.

- *Feymann's Sum over Histories*: In our lecture, we briefly touched upon that in quantum mechanics, paths which are very far away from the "classical" path contribute very little to its probability amplitude while those that are close to the "classical" paths do. Explain this process in detail. Hint : You might find the book by Feynmann, *The Character of Physical Law*, very useful.

- *Derive the Euler-Lagrange Equations**: As we have discussed in the lecture, the Euler-Lagrange Equation is equivalent to using the action principle. A good exercise is to prove this equivalence by deriving the Euler-Lagrange equation from using the action principle in general.

Chapter 3

Special Relativity, General Relativity and Black Holes

Spacetime tells matter how to move. Matter tells spacetime how to curve.

John Wheeler

3.1 Special Relativity : A speed limit and its consequences

Just to demonstrate that modern theoretical physics is all about showing Newton was not quite correct, again we will start with Newton's 2nd Law of motion, and then argue that why it doesn't quite work. As usual, Newton's law of motion is given by

$$m\ddot{\mathbf{x}} = \mathbf{F} \ . \tag{3.1}$$

The equation above describes the dynamics of the particle – the forces acting on it are assumed to be instantaneous. If **F** is the gravitational force of the Sun on the Earth, then if we remove the Sun, the Earth "immediately" feels the loss of the gravitational force.

Newton's Laws are not just a bunch of equations where you plug in some initial conditions to compute the trajectory of the particle in question, they actively promote the idea that space and time are separate entities, to be treated differently. In the Newtonian view of the Universe, we live in a 3-dimensional world, which we call "space". This 3D world then dynamically evolves, with the evolution govern by some quantity we call "time". Every being in this Universe, must agree on this time. Furthermore, all forces are instantaneous across infinite distances on each slice of time. The Laws, and their underlying ideas about how space and time are subdivided into their separate domains, are usually known collectively as **Newtonian Mechanics**.

Now, there is nothing philosophically wrong with this picture – but is this how Nature works? For the longest time, scientists thought so. Until cracks start to appear, and no crack is bigger than the slow realization throughout the 19th century that light, which was thought to propagate instantaneously, actually has a finite speed. Indeed, **Leon Foucault** has measured in 1862 that it was $298,000,000$ m/s, which is impressively close to its presently measured value of $299,792,458$ m/s. In fact, **James Maxwell** in the same year wrote down the Maxwell equations we briefly discussed in Chapter 2 which describe the dynamics of light and used it to show that the speed of light is finite.

Albert Einstein's question at that time was not why is the speed of light finite, but how to reconcile the so-called **Newtonian Worldview** to the fact that the speed of light is finite. Here is an experiment

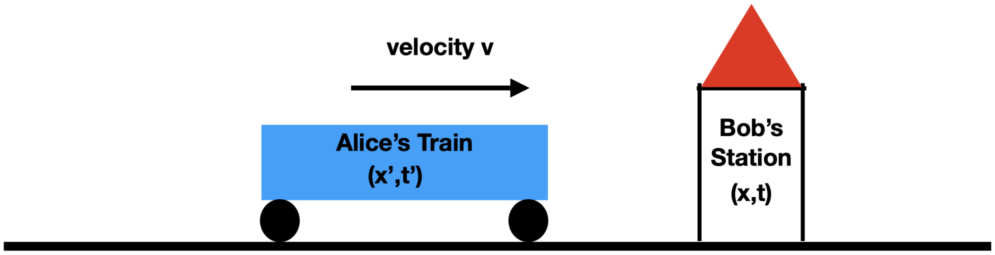


Figure 3.1: Alice's frame with coordinates (x', t') is moving at velocity v with respect to Bob's station platform with coordinates (x, t).

– let Alice be on a train moving at some velocity $\mathbf{v} = \mathbf{dx}/\mathbf{dt}$ with respect to the ground while Bob is on a stationary train platform at some point $x_0 = 0$. Both Alice and Bob shoot a laser beam at the same time when Alice's train passes Bob's train platform at a target a kilometer away at $x = 1$ km. If the Newton's worldview is correct, then we expect Alice's laser beam to hit the target before Bob's, since Alice "should" add the speed of the train to the speed of light. This is because the Newtonian Worldview suggests that there is an absolute "correct frame of refernce" and all equations are only correct in this frame. In particular, *everybody must agree on the time coordinate* t (up to a constant).

As you all know, experimentally, it turns out that both beams hit the target at the same time. But perhaps more surprisingly, it was realized at that time that this fact can actually be *predicted using the Maxwell equations*. Indeed, further to that, it can be shown[1] that if Alice uses coordinates $\mathbf{x}'$ and has a watch that measures in time t', and Bob uses coordinates $\mathbf{x}$ with a watch measuring time t, then Maxwell equations look exactly the same for both Alice and Bob as long as the two coordinate systems are related by

$$\begin{aligned} t' &= \gamma\left(t - \frac{vx}{c^2}\right) , \\ x' &= \gamma(x - vt) . \end{aligned} \tag{3.2}$$

where γ, the **Lorentz factor**, is given by

$$\gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} . \tag{3.3}$$

Notice that since $v < c$, $\gamma > 1$. This fact is discovered by **Henrikh Lorentz** in 1892, and is now called the **Lorentz Transformation**. What this suggest is that there are some special sets of coordinate systems where the equations "look the same". But more importantly – notice that t' depends on t, x and v. In other words, Alice and Bob do not agree on the time, and this disagreement depends on v. This is not compatible with the Newtonian Worldview that there is an absolute time that everyone must agree with. This incompatibility is what Einstein was trying to understand – how do we "fix" the Newtonian Worldview such that it is compatible with our experimental results?

Suppose we are in some frame, where there is a standard Cartesian coordinate x and some time t (we ignore the 2 other space dimensions y and z for now). To measure the velocity of any particle, you can measure its movement Δx over some time Δt, and then take the difference to get

$$v = \frac{\Delta x}{\Delta t} , \tag{3.4}$$

which you have done many times. In the limit when Δx and Δt is very small (i.e. **infinitisimally small**), then we get the usual $v = dx/dt$ which is a formula you know already. Suppose now that $v = c$,

[1] Unfortunately showing this will take us too far afield.

i.e. the speed of light, we can square Eq. (3.4) to get

$$0 = c^2 dt^2 + dx^2 \ . \tag{3.5}$$

As we have learned from the case of light, the speed of light c must be the regardless of the coordinate system – Alice on the train and Bob on the platform have different space coordinates and different watches, but you both agree that the speed of light is the same. In terms of Eq. (3.5), this means that *you can disagree on what x and t are but you must agree that c is the same.* In other words, suppose (x,t) is Bob's coordinates and and (x',t') is Alice's coordinates, then

$$0 = -c^2 dt^2 + dx^2 = -c^2 dt'^2 + dx'^2 \ . \tag{3.6}$$

So, the question is now : how is (x,t) related to (x',t') such that Eq. (3.6) is always true? Let's try the Newtonian Worldview suggestion, which is

$$x' = x + vt \ , \ t' = t \ , \tag{3.7}$$

which is that the difference between Alice and Bob's coordinates is simply that Alice's coordinate x' is moving at v with respect to Bob's, and that both must agree on the time[2] so $t = t'$. We can then calculate $dx' = dx + vdt$ and $dt' = dt$, but then you can check that

$$-c^2 dt'^2 + dx'^2 = (-c^2 + v^2)dt^2 + dx^2 + 2vdxdt \neq -c^2 dt^2 + dx^2 \ , \tag{3.8}$$

and thus it doesn't quite work. As you might have already guessed, what actually works is to use the Lorentz transformation Eq. (3.2) which we have introduced without much motivation above. Let's check :

$$dt' = \gamma\left(dt - \frac{v}{c^2}dx\right) \ , \ dx' = \gamma(dx - vdt) \ , \tag{3.9}$$

and plugging these, and after some algebra, you can show that Eq. (3.6) is indeed obeyed. In fact, notice that since the "0" of the equation $0 = -c^2dt^2 + dx^2$ is untouched by this transformation, we can generalize it

$$ds^2 = -c^2 dt^2 + dx^2 = -c^2 dt'^2 + dx'^2 \ , \tag{3.10}$$

where ds is a quantity called **proper length**. We can now restore the other 2 spatial dimensions y and z into our Eq. (3.10)

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \ , \tag{3.11}$$

and then you can show to yourself that Eq. (3.11) remains invariant under Lorentz transformations in the y and z directions – just replace $x \to y$ and $x \to z$ in Eq. (3.2). These transformations are named **Lorentz boosts** in the x, y and z directions respectively[3].

To see why ds is called a "length", notice that if we drop $-c^2dt^2$ term then $ds^2 = dx^2 + dy^2 + dz^2$, which is just the formula for length squared. The equation Eq. (3.11) is called a **metric** – which literally means "length" in French. The fact that the time coordinate dt is now part of a generalized length ds^2 tells you that space and time are not separate like the Newtonian Worldview asserts, but is part of **Spacetime**. So the proper length defines a "length" in spacetime. Nevertheless, the time coordinate $-c^2dt^2$ has a minus sign – which means that ds^2 can be zero or even negative – time has a different character to space.

Thus while observers can disagree on what dx, dy, dz and dt are, they have to agree on ds^2. Einstein was actually motivated by Lorentz's observation that Maxwell equations are **invariant**[4] under the

[2] This coordinate transformation is known as the **Galilean Transformation**.

[3] You might also notice that Eq. (3.11) is also invariant if we rotate the spatial dimensions around the x, y and z axes. To be precise, the set of 3 boosts and 3 rotations combined are properly called Lorentz Transformations.

[4] This is a *symmetry*, and indeed there is a conserved quantity associated with it as we have studied in Chapter 2. It is not very interesting though – the conserved quantity is the center of mass at $t = 0$ multiplied by γ, $Q = \gamma m(vt + x)$.

transformation Eq. (3.2) – indeed his famous paper on special relativity is titled *On the Electrodynamics of Moving Bodies*. He made the postulate that *not only Maxwell equations are invariant under Lorentz transformation, but all physical theories must be invariant under Lorentz transformation.* In other words, all equations of dynamics for everything in the entire universe must be invariant under Lorentz transformation. As you can easily check yourself, Newton's law of motion does not obey this postulate, and Newton's law of motion is wrong.

I have cheated a bit here – the postulate that I said Einstein made above is actually a "modern" take on things. What Einstein postulated were two things:

- **Postulate 1** : The speed of light c in vacuum is the same for all observers.

- **Postulate 2** : The laws of physics are invariant in all **inertial frames** of reference.

The term **inertial frame** roughly means that a coordinate frame that is not accelerating. A regular Cartesian (x, y, z, t) coordinate system that you are familiar with is an inertial frame, and so are all other Cartesian frames that are moving at constant velocity with respect to it. On the other hand, a rotating frame is not an inertial frame – we will have a bit more to say about non-inertial frames when we discuss general relativity in section 3.2. Combined, these two postulates imply that all laws of physics must be invariant under Lorentz Transformations.

The weird nature of the Lorentz transformation Eq. (3.2) gives rise to some interesting predictions, which have all being verified. Let's look at them.

- **Length Contraction**: Alice holds a ruler of length L', and Bob wants to measure Alice's ruler as she passes by in her train. Bob's coordinates (x, t) with respect to Alice's coordinate (x', t') is given by the Lorentz transformation Eq. (3.2). Say x'_1 is one end of Alice's ruler, and x'_2 is the other end, such that $L' = x'_2 - x'_1$, then according to Bob $x'_1 = \gamma(x_1 + vt_1)$ and $x'_2 = \gamma(x_2 + vt_2)$. Since Bob wants to make the measurements of the ruler simultaneously at his time, thus $t_1 = t_2$, subtracting the two equations we get $x'_2 - x'_1 = L' = \gamma(x_2 - x_1)$. So Bob measures the length $L = L'/\gamma$. But since $\gamma > 1$, he sees a *contracted* length of the ruler.

- **Time Dilation**: Alice's watch will tick at time t' while Bob's watch will tick at time t. Suppose Bob hummed a tune that lasted T measured by his watch $T = t_2 - t_1$, how long has Alice's watch passed at this time? Using Lorentz transformation $t = \gamma(t' - vx'/c^2)$, the two times then correspond to $t_1 = \gamma(t'_1 - vx'_1/c^2)$ and $t_2 = \gamma(t'_2 - vx'_2/c^2)$. But since Alice is standing still on her train, $x'_1 = x'_2$, subtracting we get $t_2 - t_1 = \gamma(t'_2 - t'_1)$, and thus Alice's watch $T' = T/\gamma$ – Alice's time passes *slower* with respect to Bob's watch. Now, at very relativistic speeds, where $v \to c$ (imagine a very fast train), then $\gamma \gg 1$ can become very big. This means that Alice's time will be much slower than Bob's. This leads to the so-called **Twin Paradox** – suppose Alice is taking a spaceship traveling at $v = 0.9999c$ to the next star Alpha Centauri about 4 light years away, and then back. Then her trip will last about 8 years. But as $\gamma = 1/\sqrt{1 - v^2/c^2} = 70$, so Bob would have aged $70 \times 8 = 560$ years – he would be long dead before Alice returns from her trip.

- $E = mc^2$: Finally, we come to Einstein's famous formula. We start by asking the question *what is the kinetic of an object traveling at* v? In the Newtonian picture when $v \gg c$, the answer is $1/2mv^2$. But obviously, this is not correct anymore. Let's compute it, the kinetic energy of an object is the total work done on it by a force F, i.e.

$$\mathrm{KE} = \int_{x_1}^{x_2} F dx \ . \tag{3.12}$$

  From Newton's second law, we have $F = dp/dt$. Now what is p? Imagine the particle is at rest, then $p = mv = mdx/dt'$ where t' is the time coordinate where the particle is at rest, and m is known as its **rest mass**. When the particle is moving, we know from our time dilation calculation above that $dt/dt' = \gamma$ and thus $p = mdx/dt' = m\gamma dx/dt$, remembering that γ itself has a $v = dx/dt$ inside. Plugging all these back into the equation Eq. (3.12), we got

$$\mathrm{KE} = \int_{x_1}^{x_2} m\frac{d}{dt}(m\gamma v)\,dx = \int_{x_1}^{x_2} m\gamma^3\frac{dv}{dt}dx\ . \tag{3.13}$$

You can do the integral Eq. (3.13) by using a change of variables from dx to dv (go on, try it!) to get the final answer

$$\mathrm{KE} = mc^2(\gamma - 1)\ . \tag{3.14}$$

It is easier to check that when $v = 0$, then $\gamma = 1$ and hence the kinetic energy is zero. But the true physical meaning of Eq. (3.14) comes into play when we do a simple rearrangement

$$\text{Total Energy } E \equiv \gamma mc^2 = \mathrm{KE} + mc^2\ , \tag{3.15}$$

where we E is known as the **total energy** of the particle, which is the sum of its KE and a **rest mass energy** mc^2. So if the particle is not moving then $KE = 0$, the energy of a particle at rest with mass m is the famous equation

$$E = mc^2\ . \tag{3.16}$$

Even though this is famous, the real power lies in the relation $E = \gamma mc^2$. The energy of a particle scales like γ. But you can check yourself that to get a particle to go from $v = 0.8$ to $v = 0.9$ requires much more energy than to get a particle from $v = 0.1$ to $v = 0.2$ due to the non-linear scaling of γ – as we increase the velocity of the particle, it becomes more and more expensive to make it go faster energetically speaking. Indeed, at $v = c$, $\gamma = \infty$, so in principle it is impossible to accelerate a massive particle to the speed of light. The only kind of particle that can go at the speed of light are massless particles, like the photon. In this special case, the energy of the particle is $E = pc$, where p is the momentum of the photon.

3.2 General Relativity

Einstein's true *magnum opus* of course is not special relativity (even though this would have on its own cemented his legacy), but the theory of General Relativity, which not only generalizes Special Relativity but go way beyond it to describe gravity. While, like all true stories about the development of physics, Einstein did not invent General Relativity on his own – many workers in the field like Marcel Grossman, David Hilbert and Gunnar Nordström were zooming in on the same ideas as he was, he definitely made the biggest conceptual leap, and more importantly – he got it right.

3.2.1 The Equivalence Principle

General Relativity is a beautiful mathematical theory that sometimes can be very intimidating to learn at first, and has a reputation of being hard to understand. However, I think that is completely wrong – the mathematics are not easy certainly, but the principles behind it is actually not hard. Indeed, in this lecture, we will show you how, from a very simple basic principle which guided Einstein will predict almost all of the important consequences of general relativity, without using too much math!

Again, let's go back to Newtonian physics. Recall that the Newton's 2nd law of motion is given by

$$\mathbf{F} = m_{\mathrm{inertial}}\ddot{\mathbf{x}}\ , \tag{3.17}$$

where we have explicitly labeled the mass m_{inertial} to be its **inertial mass**. On the other hand, as we have learned in chapter 2, the gravitational potential energy between particles of masses m and M is given by

$$V = -\frac{GMm_{\rm grav}}{r} \ , \tag{3.18}$$

Notice that we have explicitly labeled $m_{\rm grav}$ to be its **gravitational mass** – since Newton's Law of motion and Newton's law of gravity are two *separate laws*, there is no *a priori* reason for now that they are the same! In general, it is much more convenient to think of a "gravitational field" of a particle, instead of the potential energy between two particles. To do that, we can define a "potential energy per unit mass $m_{\rm grav}$" $\Phi \equiv V/m_{\rm grav}$, such that

$$\Phi = -\frac{GM}{r} \ , \tag{3.19}$$

which is then the **gravitational field** of a point particle of mass M. If we now take the derivative of Φ, we get the **gravitational acceleration g**

$$\mathbf{g} = -\nabla\Phi \ . \tag{3.20}$$

The force acting on a particle with gravitational mass $m_{\rm grav}$ is then

$$\mathbf{F} = m_{\rm grav}\mathbf{g} \ . \tag{3.21}$$

Now, as you have done many times in your calculations, you set Eq. (3.17) equal to Eq. (3.21) to get

$$m_{\rm inertial}\ddot{\mathbf{x}} = m_{\rm grav}\mathbf{g} \ \Rightarrow \ddot{\mathbf{x}} \overset{?}{=} \mathbf{g} \tag{3.22}$$

in other words is the inertial mass equal to the gravitational mass $m_{\rm inertial} \overset{?}{=} m_{\rm grav}$? In fact, this is *an experimentally verified fact* – the *MICROSCOPE* mini-satellite experiment has measured the equivalence up to 1 part in 10^{15}. But the point is that this fact *cannot be derived*, but must be experimentally checked. The assumption that $m_{\rm inertial} = m_{\rm grav}$ is called **The Equivalence Principle**.

The remarkable thing about the equation $\ddot{\mathbf{x}} = \mathbf{g}$ is that there is no information about what the particle is – the only information is its position $\mathbf{x}$ implying that *gravity is universal* – it affects *everything*. This universality is what makes gravity different from other forces (e.g. electromagnetic forces only act on things with electric charges) and as we will return later, also makes it hard to come up with a theory that describe it. The fact that we can identify $\ddot{\mathbf{x}}$ with $\mathbf{g}$ is usually formulated in the following way

Equivalent Principle : In a local patch, there is no way of distinguishing between acceleration and gravitational force.

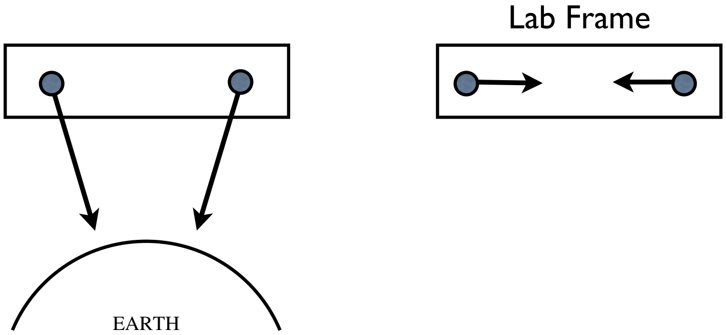


Figure 3.2: Tidal forces in a non-uniform gravitational field.

The key word **local** is very important – it means that Principle refers to the two forces *within an infinitisimally small area.* To illustrate why this is important, let's consider the idea of **Tidal Forces** (see Figure 3.2). Suppose we lock you in a windowless lab, and drop the lab towards the Earth from some distance. Even though the lab is windowless, you can devise an experiment to check that you are not in a uniform gravitational field by checking that two bodies at each end of the lab are accelerating towards each other. These forces are known as tidal forces – which arises when there is a non-uniform gravitational field. Suppose now that, due to budget cuts of the government and without EU funding, you are now given a smaller and smaller lab, then this experiment becomes increasingly hard to do – sadly not a *gedenkenexperiment* in real world. Indeed, in the limit where the Lab is now *infinitisimally small*, then you will not be able to do this experiment at all. Such an infinitisimally small patch of spacetime is called a **local** patch. Inside such a tiny lab, you *would not be able to tell that you are moving through a gravitational field*. Indeed, as far as you are concerned, you will be "free-falling" and weightless. Of course, since your lab is windowless, you won't even know that you are "falling" towards the Earth. So you can happily define an inertial frame around yourself, confident that you are non-accelerating. Such a free falling inetial frame is called a **local inertial frame**.

Let's consider some of the direct consequences of the Equivalent Principle.

3.2.2 Gravitational Lensing

Think of the windowless lab again. If I fire a beam of light horizontally from one end to the other end of the lab, then if the lab is accelerating, the light beam will clearly bend downwards. But since the Equivalence Principle tells us that this must be equivalent to the case where the lab is in a gravitational field. Thus, *gravitational field will bend the paths of light.* This is historically one of the first experiment to verify Einstein's theory.

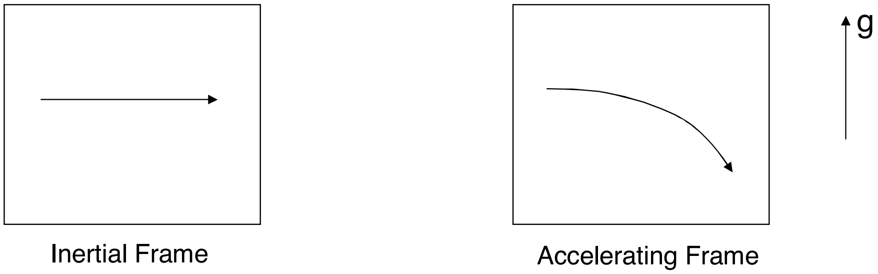


Figure 3.3: Gravitational Lensing

3.2.3 Gravitational Redshift

Now instead of shining the light horizontally, we shine the light vertically from the top of the lab to the bottom. Suppose the lab starts at rest when the light is emitted, and accelerates upwards with acceleration g, and the height of the box is h, then the velocity of the lab when the light hits the bottom would be

$$v = gt = \frac{gh}{c} \ . \tag{3.23}$$

A detector at the bottom of the lab can then measure the frequency ν' of the light – but since the lab velocity of the emission is zero while the lab velocity at measurement is v, this results in the **Doppler effect** (neglecting relativistic effects)

$$\nu' = \nu\left(1 + \frac{v}{c}\right) = \nu\left(1 + \frac{gh}{c^2}\right) \ , \tag{3.24}$$

where ν is the frequency at emission. Since $\nu' > \nu$, the frequency of light has **blue shifted**. If instead we have shot the light from bottom to the top, it would have **red shifted**. But the Equivalent principle tells us that this effect must be present if the lab is instead under the influence of a gravitational field $\mathbf{g}$. This is known as **gravitational redshifting effect**.

Since $\mathbf{g} = -\nabla\Phi$ using Eq. (3.20), and if the gravitational acceleration $\mathbf{g}$ is constant (e.g. near the surface of the Earth), then $\Phi = -gh$, and Eq. (3.24) becomes

$$\nu' = \nu\left(1 - \frac{\Phi}{c^2}\right) \ . \tag{3.25}$$

This effect was first observed in the **Pound-Rebka Experiment** in 1959.

3.2.4 Gravitational Time Dilation

Since frequency is the inverse of time $\nu = 1/T$, where T is the period, we can invert Eq. (3.25)

$$T = T'\left(1 - \frac{\Phi}{c^2}\right) \ , \tag{3.26}$$

which tells you that if time T has passed on top of the lab, then time $T' < T$ would have passed at the bottom of the lab (remember that $\Phi < 0$). Thus somebody who is on the surface of the Earth (and thus deeper in the potential well) would experience time *slower* than someone who lives in a tall building. So if you want to live longer, don't live in a high rise. This effect, called **gravitational time dilation**, has been observed using very accurate atomic clocks. The GPS sat-nav constellation actively corrects for this effect as the time dilation effect is about 45μs a day which is not small – a real world use of general relativity. The idea was used to good dramatic effect in the Christopher Nolan sci-fi movie *Interstellar.*

3.3 Curved Spacetime

As we mentioned in section 3.2.1, gravity is universal and hence everything is affected by it with no exception. Combined with the fact that accelerating frames are not inertial lead to a very difficult problem.

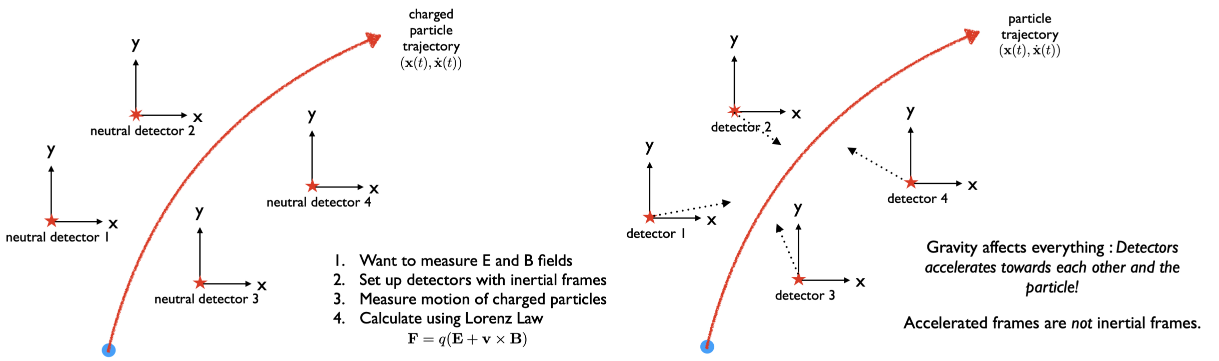
To see why this is so, consider how you would measure the electromagnetic fields : you set up a group of electrically neutral detectors resting (i.e. unaccelerated) in their inertial frames (these are known as **inertial observers**), and then you release a charged test particle. The detectors will then measure motion of the particle, and from there we are can compute the electric (and magnetic) fields (say using the Lorenz force law). The detectors, since they are electrically neutral, will *remain inertial* and hence is not affected by the charged test particle. Once you make all the measurements, since the detectors are all in inertial frames, you know how to reconstruct the entire "global" electromagnetic field as we discussed in the previous section – pick some favorite inertial frame and then Lorentz transform every other frame to it. However, if you try to do the same thing with the gravitational field – release a test particle and measure its motion to compute the gravitational field, the Equivalent principle ensures that the *detectors themselves affect and will be affected by the gravitational field*, so they no longer stay inertial – once you release them they will accelerate towards each other and as we have studied before, they no longer stay inertial.

Einstein's solution to his conundrum is to embrace this "problem" as a feature, and he proposed that

Gravity is not a force, but a property of the fact that spacetime is curved, and all free-falling (unaccelerated by any other means) objects simply follow the shortest path (called a *geodesic*) between two points.

To define the "length", we hark back to the idea of the *metric* we discussed earlier. There we said that the metric defines a notion of "length" in spacetime,

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 \ . \tag{3.27}$$


charged particle trajectory
$(\mathbf{x}(t), \dot{\mathbf{x}}(t))$
y
x
neutral detector 2
neutral detector 1
neutral detector 4
neutral detector 3
1. Want to measure E and B fields
2. Set up detectors with inertial frames
3. Measure motion of charged particles
4. Calculate using Lorenz Law
$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$
particle trajectory
$(\mathbf{x}(t), \dot{\mathbf{x}}(t))$
detector 2
detector 1
detector 4
detector 3
Gravity affects everything : Detectors accelerates towards each other and the particle!
Accelerated frames are not inertial frames.

(a) Measuring the electric and magnetic fields using inertial frames.

(b) Since gravity affect the detectors, the detectors' frames cannot stay inertial.

The $dx^2 + dy^2 + dz^2$ part of the metric Eq. (3.27) describes an **Euclidean flat** space – Euclidean in the sense that parallel lines stay parallel – as you have learned such a space the length is simply $L = \sqrt{x^2 + y^2 + z^2}$ so an infinitisimally small length would be $dl = \sqrt{dx^2 + dy^2 + dz^2}$. The addition of the "special" time coordinate $-c^2 dt^2$ means that it is a spacetime length, and such a spacetime is called **Minkowski flat space**. What about curved spacetime? Remember from our formula of time dilation Eq. (3.26), we can take the very small T limit, so $T \to dt$, and hence

$$dt'^2 = dt^2 \left(1 - \frac{\Phi}{c^2}\right)^{-2} \approx \left(1 + \frac{2\Phi}{c^2}\right) dt^2. \tag{3.28}$$

But now, as Einstein said, all theories must obey Lorentz transformation, and this transformation "mix" up all the t and $\mathbf{x}$ terms. In general, a curved metric looks like

$$ds^2 = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu dx^\nu \ , \tag{3.29}$$

where $\mu, \nu = (t, x, y, z)$ and $g_{\mu\nu}$ is a 4×4 matrix of functions. Thus Minkowski flat space $g_{\mu\nu}$ is just $\mathrm{diag}(-1, 1, 1, 1)$, but in general curved space the metric can be very complicated. And it's clear that the actual measured length ds will depend on the functions $g_{\mu\nu}$. We will talk about the how to find the "shortest path" in section 3.3.2 later, but next let's talk about black holes.

3.3.1 Black Holes

Going back to Eq. (3.28). Recall that Φ is the gravitational potential, so for a point particle of mass M

$$\Phi(r) = -\frac{GM}{r} \ . \tag{3.30}$$

Very far from $r = 0$, i.e $r \gg GM$, then $\Phi = 0$, so dt is then the time measured by an observer, say Alice, far away from this gravitational source. Now if Bob jumps in towards the source, as he approaches the center, he will instead measure the time dt' on his watch. So if he sends a signal out to Alice, for every dt' tick according to his watch, Alice will receive $(1 - 2GM/rc^2)^{-1/2} dt'$ ticks – as Bob gets closer to the center, she would notice that Bob's ticks taking longer and longer to arrive. Bob is *redshifting* from her point of view. At the point

$$r_{sch} = \frac{2GM}{c^2} \ , \tag{3.31}$$

Alice will have to wait an infinity between Bob's ticks. In other words, Alice no longer sees Bob. This special radius, r_{sch} is called the **Schwarzschild radius**, and anything that passes it will be lost forever – not even light can escape it. This point of no return is called the **Event Horizon**, and is what define a **black hole**. More than a hundred black holes have already been discovered, mostly through their gravitational waves emission when they collide against each other.

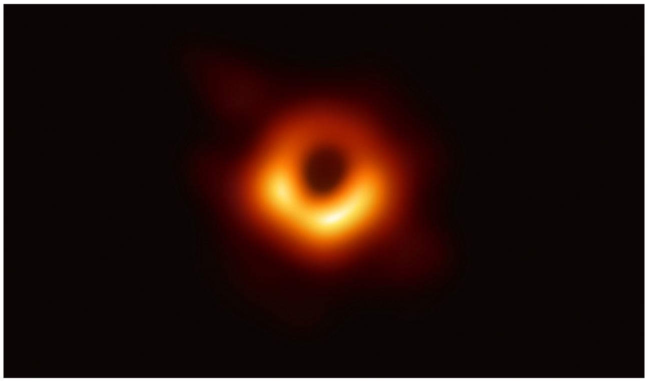


Figure 3.5: An image of the super massive black hole at the center of the M87 galaxy, taken by the Event Horizon Telescope.

3.3.2 Finding the Shortest Paths

In the epigraph of this chapter, John Wheeler succintly described General Relativity. Let's talk about the first part of this epigraph *spacetime tells matter how to move*. We have already stated in the section 3.3 above that objects in a curved spacetime follows the shortest path. Indeed, we have already learned how to calculate this in the previous Chapter 2! This problem can be solved for any general metric $g_{\mu\nu}$ (which will result in the **Geodesic Equation**), but for this introductory lecture, we'll do a simple example to illustrate the point.

Now ds^2 is the length (squared), so what we want is to find the path which minimizes ds. The action is then

$$S = -mc\int_1^2 |ds| \ , \tag{3.32}$$

where we have added in a factor of mc to keep the units right. Also, we have sneakily used $|ds|$ instead of ds since[5] $ds^2 < 0$. Let's redo the problem of gravitational redshift we have discussed in section 3.2.3, which has the metric

$$ds^2 = -\left(1+\frac{2\Phi(\mathbf{x})}{c^2}\right)c^2dt^2 + d\mathbf{x}^2 \ , \tag{3.33}$$

which we can plug into the action Eq. (3.32) to get

$$S = mc\int_1^2 dt\sqrt{\left(1+\frac{2\Phi}{c^2}\right) - \frac{\dot{\mathbf{x}}^2}{c^2}} \ , \tag{3.34}$$

where we have used $d\mathbf{x}/dt$ in the last term to pull out the dt. Assuming that the potential is small $\Phi \ll c^2$ and the velocity is small (i.e. non-relativistic) $\dot{\mathbf{x}} \ll c$, we get

$$S \approx \int_1^2 dt\left(\frac{1}{2}m\dot{\mathbf{x}}^2 - m\Phi + \ldots\right) \ , \tag{3.35}$$

where $\ldots$ mean terms which are small or constant so we can neglect them.

But look! This is exactly the form of "Kinetic Energy minus Potential Energy" which we discussed in Chapter 2, where the potential energy is now $V(\mathbf{x}) = m\Phi(\mathbf{x})$ and the kinetic energy is $(1/2)m\dot{\mathbf{x}}^2$. This is just Newtonian gravity and dynamics – and you can now see that the "Newtonian gravity" is actually induced by gravitational time dilation! In full general relativity, the $\ldots$ terms will become important – thus when Φ and $\dot{\mathbf{x}}$ are no longer small, we have to take into account the general relativistic effects.

[5] This is not as terrible as it seems. You can swap the signs of dt and $d\mathbf{x}$ in the metric and everything would still work – whichever you used is a choice of convention, called the "metric signature". Indeed, it is a constant amusement between physicists to argue about which is better – although obviously the one we used in this lecture is better.

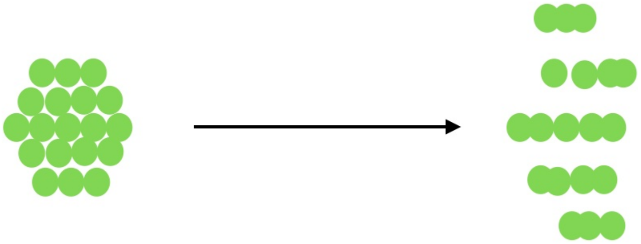


Figure 3.6: The volume and shape of a cloud of particles changes under the presence of matter field affecting the underlying spacetime curvature.

3.3.3 The Einstein Equation

Finally, we talk about the second part of Wheeler's epigraph *matter tells spacetime how to curve.* Unfortunately, this is when our ability to use simple physics come to an end, and there is really no easy way to describe it. Indeed, it is the genius of Einstein to be able to find how this actually works. Roughly, the statement ments

$$\text{Curvature} = \text{Matter}\ . \tag{3.36}$$

In equation form this is,

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}\ , \tag{3.37}$$

which probably means nothing to you.

However, there is a nice way to at least describe to you the content of Eq. (3.37). Suppose we have some matter which has some energy density ρ and some pressure P, which is not rotating or have little vortices[6]. Then if we consider a volume V initially (you can imagine a cloud of massless particles inside this volume which will follow the shortest paths described by the spacetime), then Einstein equation says that spacetime will curve in such a way that the volume and shape of the cloud evolves as[7]

$$\frac{\ddot{V}}{V} = -4\pi G\left(\rho + 3\frac{P}{c^2}\right)\ . \tag{3.38}$$

This is illustrated in Fig. 3.6. The point is that how the spacetime curves depend on the properties of the underlying matter described by its pressure and energy density (and usually a few other things).

Let's consider the case where the universe is filled with a lot of particles of mass m which are either not moving or moving very slowly. The energy density is then $\rho = nmc^2/V$, where n is the number of particles in the volume V and we have used $E = mc^2$ to account for the energy of the particles. If the particles are not moving, then it exerts no pressure, hence $P = 0$. Then Eq. (3.38) becomes

$$\ddot{V} = -4\pi G nmc^2\ , \tag{3.39}$$

which shows that $\ddot{V} < 0$ – i.e. *gravity is attractive* for "normal" stuff like particles with masses. We can easily integrate Eq. (3.39) once to find

$$\dot{V} = -4\pi G nmc^2 t + C, \tag{3.40}$$

where C is a constant which depends on the initial conditions. What Eq. (3.40) tells us is as follows, if $C \gg 4\pi G nmc^2 t$ at t today, then $\dot{V} > 0$ today. If the universe is filled with such particles, then it says that if the Universe started off with a big C, then the *universe is expanding.* But we have measured that the universe is still expanding today! So the universe expansion rate C initially must be very big

[6] To be precise, we have a **perfect fluid**.

[7] This equation is a simplified version of the **Raychaudhuri equation**.

– hence **The Big Bang**. However, if all that is in the universe are such particles, eventually the first term in Eq. (3.40) will eventually become bigger than C, and the universe will start to contract until a **Big Crunch** is reached. We can actually measure $\dot{V}$ of the universe today – it is known as the **Hubble Parameter**, and it presently is between the value of 67 and 73 Mpc/km/s. But more importantly, we can also measure $\ddot{V}$, and instead of $\ddot{V} < 0$, we found that $\ddot{V} > 0$!! This means that $\rho + 3\frac{P}{c^2} < 0$ – since $\rho > 0$ always then $P < 0$. This weird stuff with negative pressure is called **Dark Energy**, and it is a great mystery what it actually is made out of.

Beyond this simple example, I am afraid you will have to take an actual course on General Relativity.

3.4 Assignment Topics

- *Precession of the Perihelion of Mercury*: Historically, the orbit of the planet Mercury deviates from the predictions of Newtonian gravity, and is one of the motivations for general relativity. Describe the issue, and how Einstein solve the problem using the theory of general relativity.

- *Detection of Gravitational Waves by LIGO* : Gravitational waves is a prediction of General relativity. Describe qualitatively what they are, and the black hole collision event GW150914 which produce the gravitational waves observed by LIGO observatory.

- *Astrophysical black holes*: Presently, the only known sure way of making black holes is through the collapse of a massive stars. Such black holes are called "astrophysical black holes". Describe why a massive star can collapse into Black holes.

Chapter 4

Statistical Mechanics and the 2nd Law of Thermodynamics

More is different

Phillip W. Anderson

4.1 Things always gets worse

You room always get messier. An egg break but never “unbreak”. The milk mix with your cup of tea, but never “unmix”. These are your everyday experiences. If I play a movie backwards, it would not take you very long to realize that it was going backwards.

Wait! Is the last sentence true? What if the movie I played to you is actually a movie of a planet orbiting around a star – then you would not have known that I’ve pulled a prank on you. Why is that? The reason is simple – if we look at Newton’s law of motion (again!)

$$\mathbf{F} = m\frac{d^2\mathbf{x}}{dt^2} \ , \tag{4.1}$$

and now I **reverse** time from $t \to -t$, then $dt \to -dt$ but $dt^2 \to dt^2$, and Eq. (4.1) remains the same.

What this tells us is that Newton’s 2nd law of motion is **time reversal invariant** – a solution going forward is time is also a solution going backwards in time. This is why if I play the movie of a planet orbiting a star backwards, you would not have noticed it. In fact, **all the laws of physics that we**

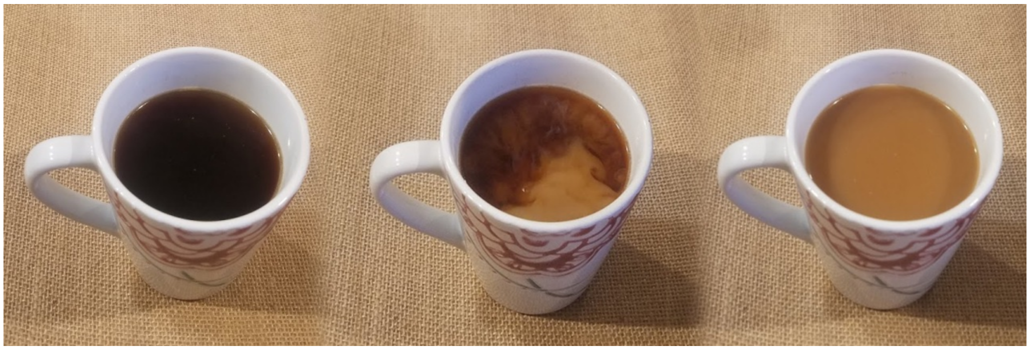


Figure 4.1: Adding milk to a cup of tea.

currently know off are time reversal invariant[1].

But then, why do you immediately realize it when I played a movie of a egg breaking backwards? This is a conundrum known as the **The Arrow of Time** problem. Nevertheless, if you think a bit more carefully about the difference between the egg-breaking movie and the orbiting-planet movie, you would realize that they are qualitatively different. In the orbiting-planet movie, the system is very simple – you have a planet which is basically a point particle, orbiting around a star which is another point particle. You can *model* the system with only 2 particles, even if the planet and star itself might be very complicated systems on their own – *you don't need to understand the details.* On the other hand, to model the egg breaking, you *do need to understand the details* of how the egg is made – how the shell is put together, how the egg white and yolk slosh around the inside and the spill out onto the floor, how the whole system interact all the forces needed to break it etc. In other words, to understand egg breaking, you need to understand how *many things are put together and how these things interact with each other.*

Another example is that of your tea mixing with milk – a cup of tea consists of about 10^{24} particles of water molecules, tea molecules and milk molecules. Each of these molecule obey a dynamical law – actually mostly Newton's law in the case of mixing – if I play the movie of a single molecule zipping about backwards you won't be able to notice that. But when we put together all 10^{24} molecules, you can tell if the movie of a tea mixing with milk is played backwards. In fact, even if we have a supercomputer (we don't) that can solve the 10^{24} equations (we didn't), what do we want to know about the system? It is useless to know the **microscopic** details like the positions and velocities of every single molecule – instead what we want to know are **macroscopic properties** like its temperature, its pressure, how fast the mixing occur etc.

The study of **Statistical Mechanics** is our very successful attempt to answer this question. From it, we developed a framework of understanding how, when we put many things together, new physical phenomena arises. These new physical laws have a very different character from the "laws" you have learned before like Newton's Law – they are **emergent** from the interactions of a large number of things interacting with each other in many complicated ways. More is indeed different. What we want to do is to show you how the basic principles of statistical mechanics will lead to an understanding of why egg break but not unbreak. But before we do that, let's take a history lesson!

4.2 The 2nd Law of Classical Thermodynamics

So this whole "things always get worse" issue did not escape the old scientists, especially in the 18th and 19th centuries. What they came up with was a set of empirical "laws", which are called the **Laws of Thermodynamics**. They mostly didn't know how to derive these laws – they literally made it up. But remarkably, it was very successful and *correct.*

There is the **First Law of Thermodynamics** (see Fig. 4.2). Consider a piston filled with gas in a volume V with some pressure P. This piston can do some work (or have work done to it) dW, and we can heat up the piston by transfering some quantity of heat dQ. The gas itself can hold some **internal energy** E, roughly the energy that all the molecules of the gas. So simple conservation of energy tells us that the change in the internal energy is

$$dE = dQ + dW \ . \tag{4.2}$$

The first "law" of thermodynamics is really just a restatement of the law of conservation of energy, so it's really nothing new.

[1] This is actually not 100% true – the weak nuclear force is ever so slightly non-time-reversal invariant, but the effect is so small that it takes very special experiments to detect it. Needless to say, it is something we never actually see in our every day experiences.

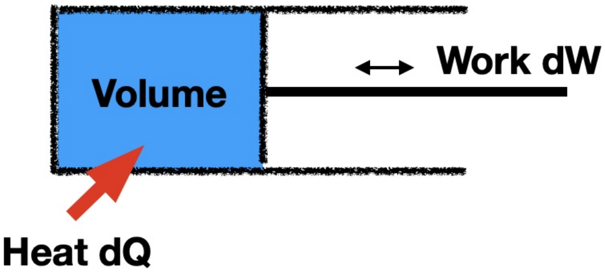


Figure 4.2: A piston with some gas it in.

What's new, and profound, is the **Second Law of Thermodynamics**. This is the law which try to understand the whole "why egg doesn't unbreak" business. It first makes the observation that a broken egg is "disordered" while an unbroken egg is "ordered". By "disorder" it vaguely means that there are so many ways that we can have a broken egg and there is an inherent "randomness" associated with it. The number of scare-quotes in the previous sentence tells you that the scientists were not quite sure how to describe it – it would take a genius to figure this out as we see later – but they could at least give it a name : **Entropy** S. The more "disordered" a system is, the higher the entropy of the system; so an unbroken egg has low entropy while a broken egg has high entropy. Then "things always get worse" can be expressed as

$$\text{2nd Law Part 1}: \ \frac{dS}{dt} \geq 0 \ . \tag{4.3}$$

To be precise, this is true only in **closed systems** where the system doesn't interact with anything external. This means that $dQ = dW = 0$ – since either of these processes (putting in/taking out heat and doing work on/by) will mean that the system has interacted with the external environment.

To quantify S, the old physicists need to associate it with something they can measure. So they proposed

$$\text{2nd Law Part 2}: \ dS = \frac{dQ}{T} \ , \tag{4.4}$$

where T is the temperature of the system (something eminently measurable). As it turns out, the physicists lucked out and this is exactly right[2]. Notice that you can *decrease* S if you are allowed to interact with the system like doing work or taking heat out – you can make your room less messy by cleaning up and doing work on it!

But can we actually *derive* the 2nd Law of Thermodynamics? Turns out that, yes we can! In this lecture, we will show you how.

4.3 The 2nd Law from Counting

Ludwig Boltzmann, who spent much of his life studying statistical mechanics, died in 1906 by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is your turn to study statistical mechanics. (David Goodstein)

4.3.1 Microstates and Macrostates

The aforementioned genius who figured it all out was **Ludwig Boltzmann**, who made the realization that the Laws of Thermodynamics are *probabilistic laws*[3]. He realized that the statement "things always

[2] To be specific, it is for a very special kind of process called **reversible** process.

[3] You might wonder whether this probability is due to the probablistic nature of quantum mechanics. The answer is no. On the other hand, the interplay between the probabilitistic mature of statistical mechanics and quantum mechanics has

get worse” must be modified into

Things ***probably*** *get worse.*

More precisely, he says that *entropy probably increases.* To illustrate that – suppose I give you a million coins, and ask you the arrange each coin by randomly flipping each of them. So you will flip a million coins and arrange them in a line. Let’s say 1 means head, and 0 means tail, then you would get something like

1000101001010010101011111100101010110...

But, very rarely, sometimes you get a long series of 11111111 just by chance. In fact, you *could* get a million 1 in a row which our physical laws allow, but the probability of that happening is $0.5^{1000000000}$, which is not very likely. This is the sense of what Boltzmann mean – entropy *can* increase naturally, but it is just not very likely – in fact, it is almost impossible to observe entropy decrease in the real world for a close system.

Note that this whole argument relies on a large number of coin-flips – it makes no sense if we just talk about a single coin flip (or even 3 coin flips). This is the core of Boltzmann’s argument – that the 2nd Law arises when there is a large number possibilities – either many events such as the coin-flip or that there are many objects. Let’s look at the coin-flip case again – there are many possibilities

110101010010100010101010001...
100100100111001010010010100...
001010101000111110010100101...
...
111111111111111111111111111...
...
001010010010110010100101001...
101011011111010000101001001...
000101011110001010010010010...
001010010111100100001010101...
...

But it’s clear that an event like 111111111111111111... is very special, while the other possibilities – even if they are unique – are not that “special”. Each of this possibility is called a **microstate**. For a million coins, the total number of possible microstates will be equal to $2^{1000000000}$, which is a mind-boggling big number. The set of all possible microstates is called **an ensemble**. So if you randomly pick a microstate out of this ensemble, your chance of getting a million 1 in a row is just impossibly small.

You can rightfully complain that “special” is not very well defined. We can define it as follows. Let N_0 be the total number of 0 and N_1 be the total number of 1, and then we will can define the following quantity

$$m = N_1 - N_0 \ , \tag{4.5}$$

to be its **macrostate**. We can now ask – how many microstates are there for each possible value of the macrostate m? Let’s put in some numbers to illustrate. Consider $N = 16$ coins (instead of a million) so that we don’t fill up all these pages with numbers. For $m = -16$, or $m = 16$, it’s clear that there is only one way : all 0 and all 1 respectively. If $m = -14$, then we have 15 0’s and a single 1, and since the 1 can be any of the 16 positions, we have 16 ways etc. We can write down a nice formula for this. Suppose

yielded a rich vein of new physical ideas.

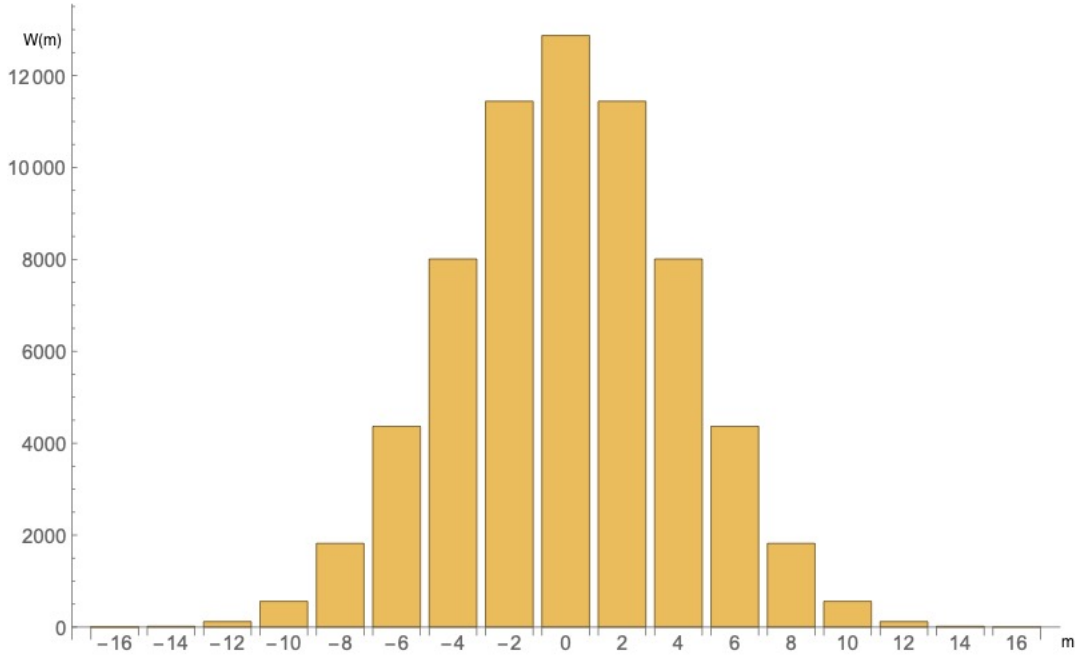


Figure 4.3: The number of possible microstates $W(m)$ for each microstate m.

$W(m)$ is the number of ways for the macrostate to be m, then

$$W(m) = \frac{N!}{N_0!N_1!} \ . \tag{4.6}$$

$W(m)$ is called the **statistical weight** of the system. As m gets closer to 0, which is when there are equal number of 1 and 0, $W(m)$ will increase until we reach a maximum when $m = 0$ where $W(m = 0) = 12870$ – see Fig. 4.3. Now, we add a new fundamental assumption

All microstates are equally probable.

This assumption means that the probability of picking a microstate with macrostate m is then

$$P(m) = \frac{W(m)}{\sum_m W(m)} \ . \tag{4.7}$$

Thus the bigger the $W(m)$, the more likely you end up picking it by random chance.

Put another way, for each macrostate m, there may be many possible microstates associated with it. But there are some macrostates where there are very few (or even just a single) microstate associated with it, like $m = -16$ or $m = 16$, and we can then call these macrostates "special". If I randomly pick a microstate from the ensemble, chances are you will pick where m is closer to 0 than not, since this is where most of the microstates are[4].

Boltzmann now makes the remarkable assertion : he stated that the statistical weight W is related to the entropy S by the **Boltzmann Entropy Formula**

$$S = k_b \ln W \ , \tag{4.9}$$

where $k_b = 1.38065\ \mathrm{m^2kgs^{-2}K^{-1}}$ is known as the **Boltzmann constant**. Thus a macrostate with many microstate, say $m = 0$ has a high entropy, while a "special" macrostate like $m = 16$ or $m = -16$ has a low

[4] For very large N, calculating factorials can kill your calculator. Fortunately, there is a nice approximate formula

$$W(m) \approx 2^N e^{-m^2/2N} \ , \tag{4.8}$$

which is a **Normal Distribution**.

Figure 4.4: Boltzmann's grave at Zentralfriedhoft in Vienna, with his eponymous formula engraved.

entropy. This gives meaning to the vague notion of "disorder" we talked about in the previous section. Boltzmann said that, don't think about "disorder" or "randomness", instead think about the number of microstates for each macrostate. He *quantified* entropy. This profound realization effectively started the whole study of statistical mechanics.

Unfortunately, Boltzmann had a difficult time trying to convince scientists of his time the profundity of his idea, and he was ostracised throughout his life. He got depressed, and took his own life in 1906. This formula is engraved on his grave stone.

4.3.2 Evolving the Ensemble

Let's see how this leads to the 2nd Law! Notice that the second law $dS/dt \geq 0$ is a *time-derivative* – it pertains to the *evolution* of the system, or its dynamics. Flipping coins has no dynamics, so let's invent a rule for it.

For every coin, there is a 1 in 6 chance (a die roll) that the coin will flip to 1 (if it is initially 0) or 0 (if it is initially 1) every small time period Δt.

So, if I have a single coin, then the "evolution" of the coin would go something like this –

$$\overset{\text{Time }\longrightarrow}{111111111100011110000011111111100011110000000} \tag{4.10}$$

Although this rule is probabilistic, if I flip the order above around

$$\overset{\text{Time }\longrightarrow}{000000011110001111111110000011110001111111111} \tag{4.11}$$

you can't really tell the difference, so the evolution is "time-reversal invariant[5]"!

Using this rule, we can then evolve our set of $N = 16$ coins. Let's start with all 1, i.e. the $m = 16$ macrostate. In the beginning, we will move rapidly towards $m = 0$ – since there are more 1's to flip, chances are that we will get driven towards smaller m. But as we collect more and more 0's, sometimes 0 will flip back to 1, and thus we can sometimes get m moving from a smaller back to bigger number.

[5] Since the probability for a flip at each time is uncorrelated with the previous one, it will also be uncorrelated with the future one.

For example, the following sequence of events may occur :

$$\begin{array}{ccccccccccc}
1111 & & 1111 & & 1101 & & 0101 & & 0000 & & 0010 \\
1111 & & 1011 & & 0011 & & 0001 & & 0011 & & 1001 \\
1111 & & 1111 & & 1010 & & 1110 & & 0010 & & 0110 \\
1111 & & 1110 & & 1111 & & 1101 & & 1111 & & 1110 \\
m=16 & \rightarrow & m=12 & \rightarrow & m=6 & \rightarrow & m=2 & \rightarrow & m=-2 & \rightarrow & m=0\ \ldots
\end{array} \tag{4.12}$$

As the macrostate reaches near $m = 0$, the number of 1's and 0's will then be approximately equal, and hence since the numbers of 1 and 0 flipping would be roughly the same at that stage, we expect the system to stabilize at around $m = 0$. Such a point is called **the equilibrium point** – generically a system left to its own devices would want to evolve towards this state. The reason is simple of course – there are just so many more microstates at equilibrium point than any other point – $W(m = 0)$ is maximum.

Thus, statistically, since m evolves from $m = 16$, which is a very special low entropy point using the Boltzmann formula Eq. (4.9), towards $m = 0$, the equilibrium point which is also the point of highest entropy, this means that

$$\frac{dS}{dt} \geq 0 \ , \tag{4.13}$$

which is (at least the first part of) the 2nd Law of Thermodynamics we discussed in Eq. (4.3)! Furthermore, it also predicts that the *equilibrium state is the final state of any closed system* – it doesn't matter where the system begin, it wants to go to the equilibrium state.

Despite its simplicity, this coin system does a decent job modeling the real world system of a **paramagnet** when heated to a constant and high temperature T. A paramagnet consists of a lattice of magnetic dipoles, which can have two possible states, ↑ and ↓ – you can think of these dipoles as tiny magnets with either N or S polarity. So its **magnetization** is then the difference between the number of ↑ dipole and ↓ dipole – this is $N_\uparrow - N_\downarrow$. So if you replace 1 with ↑ and 0 with ↓, the **macrostate** of the paramagnet is then simply the m we have defined in Eq. (4.5). As we heat up the magnet, the little dipoles gain energy and start to flip, and the whole system gets driven to an equilibrium state where there is no magnetization. You can try this experiment : just heat up a magnet and see what happens (this is known as **Curie Law**).

Finally, what about the 2nd part of the 2nd Law Eq. (4.4)? Here, the old physicists got it the wrong way round – that equation is supposed to define entropy, but Boltzmann already provided the actual formula for S which is Eq. (4.9). Instead, *temperature itself is defined by entropy!* In other words, Eq. (4.4) can be taken as a definition for temperature[6] in a reversible system.

4.4 Information is Physical

One of the most recent and exciting development in theoretical physics is the realization that the seemingly different field of **information theory**, started in the 1940s by the great **Claude Shannon**, is actually intimately related to statistical mechanics. To tell this story, we will start in the 19th century, where Maxwell in 1867 pointed out a thought experiment which seems to violate the 2nd Law of Thermodynamics. We will then see that the resolution of this paradox involves understanding how information is actually physical.

[6] To be precise, temperature is *defined* as $1/T = \partial S/\partial E$ where E is the internal energy.

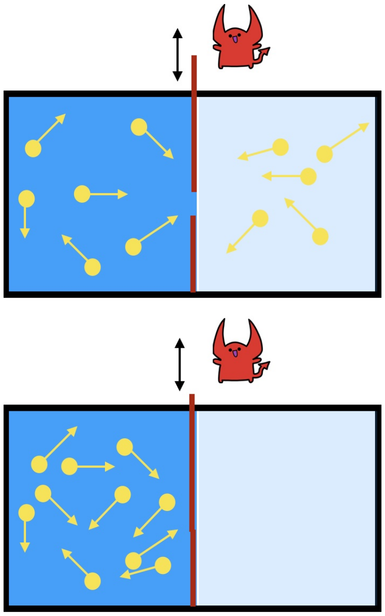


Figure 4.5: Maxwell Demon can open the gate strategically such that all the particles will eventually be trapped on one side while the other side is empty

.

4.4.1 Maxwell Demon and Szilard's Engine

Maxwell proposed the following experiment (Fig. 4.5). Imagine a box of gas particles which are moving around at some given temperature and some internal energy. We can partition the box into two halves with equal volumes, with a gate in between. There is a magical being, the **Maxwell Demon**, who can control the gate (which is assumed to be massless and hence takes no energy to open and close). The Demon observes each particle, and as a particle on the right side of the box approaches the gate, it will open the gate and let the particle to the left side, shutting the gate after. It can then do this for all the particles on the right side, until all of the particles have been sequestered on the left side. Given that we have the same number of particles, with the same kinetic energy, but with half the space, this means that the total number of possible microstates have gone down (by half actually), and hence the entropy must have gone down. Since the Demon apparently used no energy to open and shut the gate, it has done no work – so has the 2nd Law being violated?

This problem vexxed scientists for a long time. The first step to resolving the problem came when **Leo Szilard** proposed a simplified version of the Maxwell Demon problem, and showed that while the Demon has not used any energy to open or shut the gate, it has actually used something else – its knowledge! The **Szilard Engine** goes as follows. Imagine the same set up as the Maxwell Demon, but with only one particle. Instead of a box, it is a piston.

Now the Demon observes the particle. If it sees the particle on the right side, then it will open the gate to let the particle in to the left. If it sees the particle on the left side then shut the gate, then it will do nothing. In both cases, the particle is now trapped on the left side – the Demon now possess a *bit* of information which is that the particle is on the left side. Why is it a bit? Remember a bit is either 1 or 0 – so whether a particle is on the left or right side corresponds to a bit.

Once the particle is trapped on the left side, the Demon pushed the piston in, and then open the gate, Fig. 4.7. The "pressure" of the gas of single particle will push the piston out with some pressure P, allowing the Demon to extract *work* W out of the system. If we assume the piston is in temperature

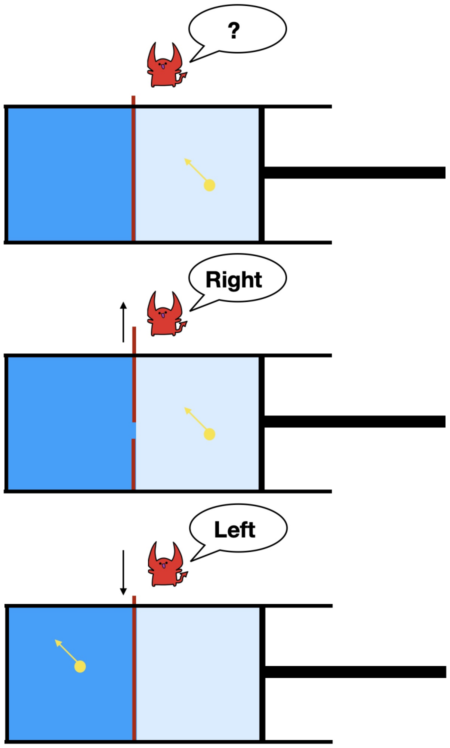


Figure 4.6: Szilard's Engine Part 1 : The Demon obtains knowledge about the location of the particle, and use it to set up the system such that the particle is on the left side.

T and the volumes V of the two sides are equal, then this work is given by the standard formula

$$W = \int_{V}^{2V} PdV \ , \tag{4.14}$$

and then using the ideal gas law $PV = k_b T$, we have

$$W = k_b T \log 2 \ . \tag{4.15}$$

At the end of the experiment, the particle could be anywhere in the piston again, so the Demon's knowledge of the particle is outdated (his brain is still stuck on "Left" – this will be important later). In other words, *the Demon has traded one bit of useful information to do $k_b T \log 2$ of work!*. Szilard's engine beautifully illustrates the deep connection between information theory and thermodynamics. Knowledge, literally, is power.

4.4.2 Landauer's Principle

Nevertheless, Szilard's engine just emphasised the Maxwell Demon paradox – the Demon can use this work to reduce the entropy of the system. Thus what is the true resolution of this paradox? It turns out that it is only 1982 that this paradox was resolved. In 1960, **Rolf Landauer** pointed out that while the Demon can make measurement of where the particle is without increasing the entropy of the system[7], the Demon still needs to store the information somewhere. In our figure, this storage is illustrated in the thought bubble of the Demon. At the end of the engine cycle in Fig. 4.7, the Demon is now stuck with an outdated information "Left", when in the start of the cycle in Fig. 4.6, it has started with an empty "?". Thus the cycle is actually not closed! To completely close the cycle, this piece of outdated

[7] Technically, this means that measurement processes can be **reversible**.

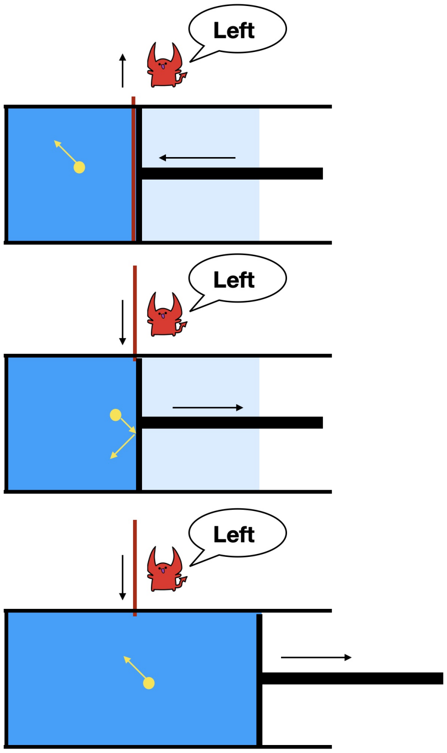


Figure 4.7: Szilard's Engine Part 2 : The Demon uses its one bit of information to extract work out of the system, but ended up with outdated information ("Left") when the particle location is unknown.

information "Left" *needs to be erased*, and Landauer showed that to erase this bit of information actually raises the entropy of the system[8] (which include the piston *and* the Demon itself) by $k_b T \log 2$! Now you can ask, well, what happens if the Demon actually has more than one bit of storage – so why can't the Demon keep just adding information. This question was answered by **Chuck Bennett** in 1982, when he pointed out that adding more memory doesn't actually close the cycle. Each cycle, the Demon acquire more and more information, but the information it acquired is useless (since it has been used to do work), and thus this increase the entropy of the whole system. In other words, the Demon's memory entropy increased to compensate for the loss of entropy of the environment. Thus the 2nd Law was finally saved from Maxwell Demon in 1982, 115 years after it was introduced.

4.5 The Beginning and the End of the Universe

Since the 2nd Law of Thermodynamics says that entropy must always increased, what happens when we apply the principle to the entire universe? If we take a look at the cosmos today, one of the weird and strange things today we see is that it is actually very ordered – the universe looks more or less the same regardless of where you are, in every single direction. This is known as the **Cosmological Copernican Principle**. This means that the entropy of the universe today is actually quite low.

[8] Here is a way to see why. Imagine that the memory bank of the Demon is a box and a single token. The Demon has a very simple system to keep track of Left and Right – put the token on the left side of its memory box if the particle in the piston in the Left, and right side if the particle is on the Right side. Since the token is not moving, the Demon's memory box allows it to keep track of where the particle is. How does the Demon *erase* this knowledge? It must make the system such that the memory box does not allow him to keep the definiteness of the position of the token. A simple way to do this is simply to heat up the token, and make it move around! Once it is freely flying around, the Demon has lost knowledge, and the memory box no longer keep the information. Heating up such a system requires $k_b \ln 2$ energy.

But since the 2nd Law says that the entropy of the universe must always increase, and we are at a very low entropy state today, then this means that **the entropy of the universe must be incredibly low at the beginning**. You ask : *so what*? Since there is no physical law that prevents the universe from starting at a very low entropy state, there is no problem. However, since the flow of entropy defines the **Arrow of Time**, and as we discussed, all of the known physical laws (except the weak nuclear force, in a very small way) is time-reversal invariant, the question of *why does time always flow forward* is completely equivalent to the question *why did the universe began at a low entropy state.* This is known as the **cosmological initial state problem**, and presently, it is unsolved.

Another way to put it is as follows. How does one *choose* which state would be the initial state of the universe? Absent any idea or theory, then it seems like the only choice is to choose one *randomly.* But as we have learned, a random state would not be a special low entropy state, and we are very likely to end up choosing a very high entropy state since there are so many more of those. So, this means that either we got very lucky, as a universe, or there is some deep yet unknown "theory of initial conditions" which has yet to be discovered.

4.5.1 Black Hole Thermodynamics

So much about the beginning of the universe. What about the end of the universe? Today, we see around us galaxies of stars mixed with hot and cold gas in the universe. Gravity is universal, thus all this matter will eventually attract each other, and they will collapse onto each other to form black holes. If you wait for a really long time, the universe will consists of mostly black holes.

This leads immediately to a problem. Every black hole is characterized by three numbers : its **mass** M, it's **spin** J, and it's **charges** Q. It doesn't matter what you used to make the black hole, at the end every black hole can be characterized by these three numbers. This is neatly summarized by the phrase *black holes have no hair.* The problem is now as follows : what happens to all the information you have used to make the black hole? For example, suppose you make a black hole with a set of the universe's most valuable encyclopedia – and the books all go into the black hole and never come out again. Is all this knowledge lost? This seems to violate the 2nd Law of Thermodynamics : you can just discard all the your high entropy things into the black hole (for example, imagine throwing the memory bank of our Maxwell Demon in the previous section into the black hole), and suddenly the universe's entropy has been lowered! Since it seems like all the matter in the universe will eventually become black holes, does this mean that the somehow the universe violated the 2nd Law?

This paradox inspired **Jacob Bekenstein**, to conjecture in 1973 (as a PhD student!) that perhaps black holes are not "low entropy" objects as seemingly suggested by the no-hair theorem, but they are objects with very high entropy. He put this fact together with **Stephen Hawking**'s **Black Hole Area Theorem** which says that the black hole horizon's area can only increase[9], and proposed that the area of the black hole $A_{\rm BH}$ is proportional to the entropy it must carry, i.e.

$$S_{\rm BH} \propto A_{\rm BH} \ . \tag{4.16}$$

This rather audacious proposal was eventually proven correct, by Hawking himself (who initially thought Bekenstein was wrong). In fact Hawking showed that

$$S_{\rm BH} = \frac{k_b A}{4 l_p^2} \ , \tag{4.17}$$

where l_p is the **Planck Length** given by

$$l_p = \sqrt{\frac{\hbar G}{c^3}} \approx 1.6 \times 10^{-35} \text{ m.} \tag{4.18}$$

[9] Roughly speaking, since $R \propto M$, then $A = 4\pi R^2 \propto M^2$, and since black hole can absorb mass but not emit mass *classically*, M must always increase, and hence A always increase.

This is a remarkable formula – it combine the all four fundamental constants: G for gravitation, $\hbar$ for quantum mechanics, c for special relativity and the Boltzmann constant k_b into a single formula. It implies that somehow, black holes are key to understanding the unification of gravity and quantum mechanics. The relationship between entropy and area of the black holes have inspired physicists to propose the idea of **Black Hole Thermodynamics**, which suggests that black holes obey the following "1st Law" of Black Hole Thermodynamics

$$dE = \frac{\kappa}{8\pi} dA + \Omega dJ + \Phi dQ \ , \tag{4.19}$$

where κ is the surface gravity, Ω is the angular velocity and Φ is the electrostatic potential. The "2nd Law" of Black Hole Thermodynamics is the Hawking Area Theorem $dA/dt \geq 0$ as mentioned above. We will discuss how Hawking came about this conclusion in the next chapter 5.

This is not the end of the story though! Given that we (mostly) know what the laws of physics are, we can calculate what would be a state of maximum entropy of the universe. Turns out that this maximum entropy state is actually just a universe filled with a homogeneous bath of radiation at the same temperature everywhere – the so-called **Heat Death** of the universe. If we believe in the 2nd Law, the universe always must evolve towards Heat Death, but this is not the same as what we logically expect from evolving the matter, which is a universe filled with black holes. What gives? The answer, as some of you might already know, is that black hole actually is not the final state of matter, but instead it will still emit **Hawking Radiation**. We will leave the discussion of this till next chapter 5.

4.5.2 Recurrence time

Finally, we will talk about an attempt to understand the origins of the universe using statistical mechanics. We argue that a system always wants to move towards equilibrium simply because that there are simply so many more accessible microstates at that point. However, there is no microphysical reason that, statistically, the system cannot evolve into a microstate that is very far away from the equilibrium point, just by random chance. In other words, if we wait long enough, we will come back to the same configuration of particles[10].

This is known as the **Poincaré Recurrence** time, and roughly it is

$$t_p \sim e^{e^S} \tag{4.20}$$

where S is the maximum entropy (at equilibrium), which is a very long time indeed, If we consider the entire observable universe as our combined system, then we can compute the recurrence time as

$$t_{rec} \sim 10^{10^{10^{10^2}}} \text{ years.} \tag{4.21}$$

This has been suggested as a solution to the problem of cosmological origins as follows. The universe is eternal, and spends most of its time in thermal equilibrium and hence is at the Heat Death. However, once in a Poincaré time (i.e. a really long while), the universe randomly fluctuates to a state of lower entropy. It will then statistically evolve towards the state of maximum entropy following the 2nd law, and we are simply living in this very special moment where the universe is still evolving to its equilibrium. In this universe, there is infinite time, so everything and anything can happen an infinite number of times. There are many issues with this model, and one of them is helpfully pointed out by Utahraptor in Fig. 4.8.

[10] This is not a guarantee – it requires the system to obey **Liouville's theorem and ergodicity** – if the orbits of the trajectory in phase space is bounded (which it usually is), then given infinite time to evolve, the trajectory will intersect every phase point an infinite amount of times, including those very rare phase points that are far away from equilibrium.

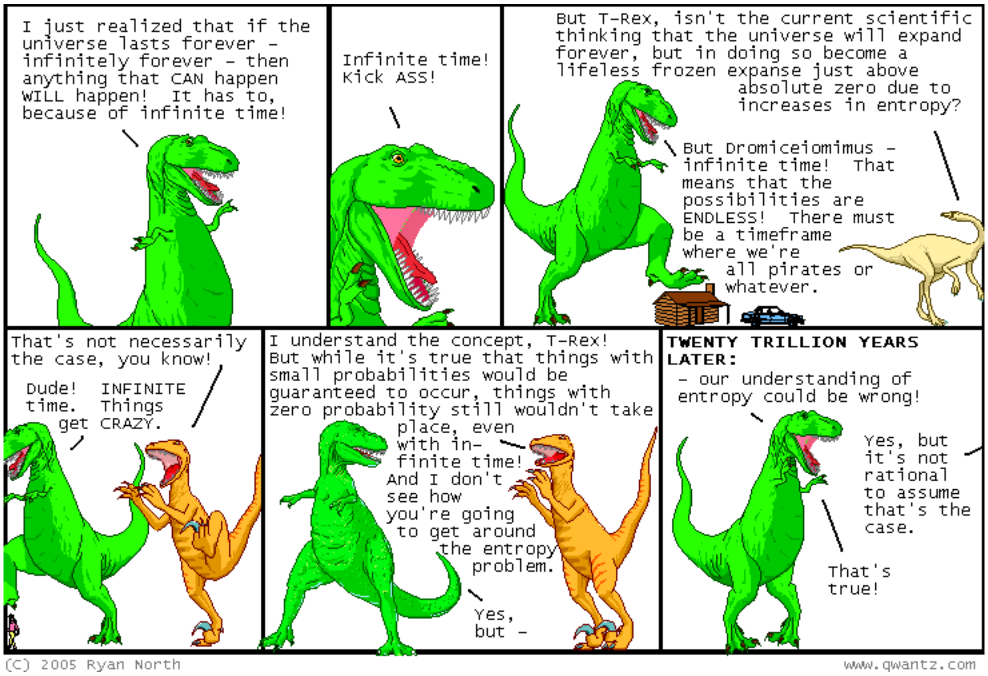

I just realized that if the universe lasts forever - infinitely forever - then anything that CAN happen WILL happen! It has to, because of infinite time!
Infinite time! Kick ASS!
But T-Rex, isn't the current scientific thinking that the universe will expand forever, but in doing so become a lifeless frozen expanse just above absolute zero due to increases in entropy?
But Dromiceiomimus - infinite time! That means that the possibilities are ENDLESS! There must be a timeframe where we're all pirates or whatever.
That's not necessarily the case, you know!
Dude! INFINITE time. Things get CRAZY.
I understand the concept, T-Rex! But while it's true that things with small probabilities would be guaranteed to occur, things with zero probability still wouldn't take place, even with in-finite time! And I don't see how you're going to get around the entropy problem.
Yes, but -
TWENTY TRILLION YEARS LATER:
- our understanding of entropy could be wrong!
Yes, but it's not rational to assume that's the case.
That's true!
(C) 2005 Ryan North
www.qwantz.com

Figure 4.8: T-Rex contemplating Poincaré recurrence, though Utahraptor insightfully points out that the recurrence requires the universe to be ergodic, which is not necessarily true. Credit : `www.qwantz.com`.

4.6 Assignment Topics

- *Charles Bennett and his resolution to the Maxwell Demon paradox*: Describe how Benette uses Landauer's Principle to resolve the Maxwell Demon paradox. There is a good *Scientific American* article in 1987 which he wrote that will make a good reference.

- *The Boltzmann Brains paradox*: One of the problems of using the Poincaré Recurrence time to resolve the cosmological initial state problem is that it gives rise to the so-called "Boltzmann Brains". Describe what they are, and why they are a problem.

- *The Inflationary Cosmology*: Another attempt at solving the cosmological initial state problem is a theoretical idea called "cosmic inflation". Describe what it is, and how it proposes to solve the problem.

Chapter 5

Quantum Gravity and the Black Hole Information Paradox

> *I was taught that the way of progress was neither swift nor easy.*
>
> Marie Curie

5.1 What's keeping us in our jobs?

While there are many unsolved problems in theoretical physics today, and we have mentioned a fair few in these lectures so far, e.g. what is Dark Energy and Dark Matter, how did the universe begin and how would it end, why is there a limit to the speed of light, it is increasingly clear to us that perhaps our failure to answer these questions may lie in our failure to understand a long known problem – how do we unify quantum mechanics and gravity ? In other words, what is the right theory of **quantum gravity**?

In this lecture, we will explain to you *why* this is such a difficult problem. As you will soon see, the fact that we can actually talk about this in a meaningful way in an introductory lecture is in a way amazing – you have already learned most of the physics in all the earlier Chapters to (hopefully) understand the main cause! But it is also the source of its difficulty – the obstruction is extremely fundamental and hence it is very hard to work around it. So, now let's put together all the knowlege we have learned so far in these lectures to tell you this story.

5.2 The Black Hole Information Paradox

In section 4.5.1, we alluded to the fact that black holes must radiate **Hawking Radiation**. Let's see how this come about. To recap, Bekenstein conjectured that black holes must carry entropy proportional to their area,

$$S_{\rm BH} \propto A_{\rm BH} \ , \tag{5.1}$$

and that since $dA_{\rm BH}/dt \geq 0$ then $dS_{\rm BH}/dt \geq 0$, obeying the 2nd Law. But Hawking noted that this is not all of the 2nd Law – remember that there is a 2nd part of the 2nd Law, which is

$$dS = \frac{dQ}{T} \ , \tag{5.2}$$

which, as we argue, defines the temperature T. So, any system that has entropy must also possess a temperature T. And if it has a temperature T, then it must also emit **blackbody radiation**. But obviously, black holes are black, so there can be no radiation...no? Hawking initially thought so, and hence he thought Bekenstein was wrong. On the other hand, it is known that blackbody radiation can only be derived when we incorporate quantum mechanics into our study of statistical mechanics (something unfortunately we didn't have the time cover in these lectures). So perhaps the answer lies in somehow incorporating *some* quantum mechanics into the physics of black holes? This is exactly what Hawking did.

5.2.1 Virtual Particles from Quantum Mechanics

Hawking's calculation in his famous 1975 paper *Particle Creating by Black Holes* was a *tour de force* – a masterpiece in clear cut calculation and explanation that I think any other explanation of that paper is just a cheap imitation. Sadly for us, the paper is very technical, so we would not be able to discuss it in our introductory lecture. What I will tell you now though, is a heuristic version (actually proposed by Hawking himself). So let's begin.

The main ingredient we need from quantum mechanics is the Heisenberg Uncertainty Principle Eq. (1.2),

$$\Delta\mathbf{x}\Delta\mathbf{p} \geq \frac{\hbar}{2} . \tag{5.3}$$

However, this version is a *non-relativistic* version – we know from our study of special relativity that velocities, and hence momenta, is a frame dependent quantity. When we incorporate the physics of special relativity into quantum mechanics, what we get is the following

$$\Delta E\Delta t \geq \frac{\hbar}{2} , \tag{5.4}$$

which is what I like to call the "interest-free loan version of quantum mechanics". The equation roughly states that

Any system can borrow free energy from the vacuum of up to ΔE for a period of time Δt, as long as there is nobody making an observation during this period.

To see how this works, let's consider the case of an electron moving along some velocity v. This electron can be relativistic, so its total energy is given by the formula Eq. (3.15) $E = \gamma m_e c^2$ where m_e is the rest mass of the electron as we have studied in Chapter 3. Now quantum mechanics says that, at *any point in time*, there is some probability of the electron "borrowing" some quantity of energy ΔE, which it then proceed to use to make a photon with energy $\Delta = \hbar\omega$, where ω is the frequency of the photon. But quantum mechanics says that you also have to pay it back within the time $\Delta t < \hbar/2\Delta E$ – so *the more you borrow, the less time you have to enjoy it* – and the photon gets "reabsorbed" by the electron and returned to the bank. Since you pay back as much as you borrow, this loan is interest free. We illustrate this in the Fig. 5.1 below.

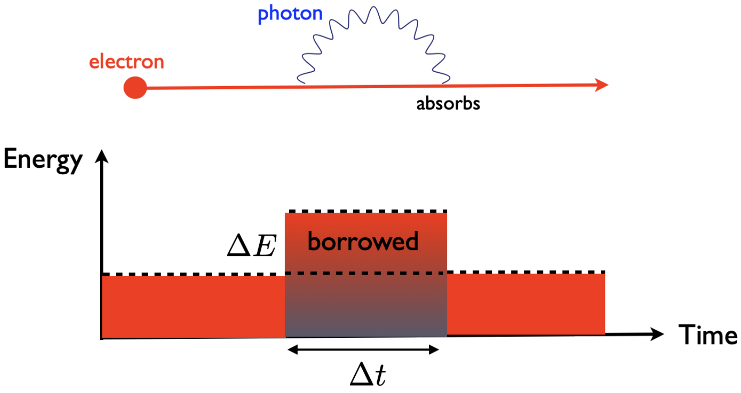


Figure 5.1: An electron borrowed energy ΔE to make a photon, which it then absorbs back after time Δt, with $\Delta E \Delta t \geq \hbar/2$.

Does this violate energy conservation? Yes! But that's ok, because you will never actually observe it in real life anyhow! The "borrowed" photon is known as a **virtual particle**, and such an event is called a **quantum fluctuation**. This is because even "empty space" is subject to quantum uncertainty. As it turns out, we don't even need the electron to borrow an electron – empty space itself can borrow energy to make virtual particles! For example, it can borrow energy to make a virtual particle and a virtual anti-particle, which fly around for a while before coming back together to annihilate each other[1], releasing energy which is then used to pay back the "loan", Fig. 5.2.

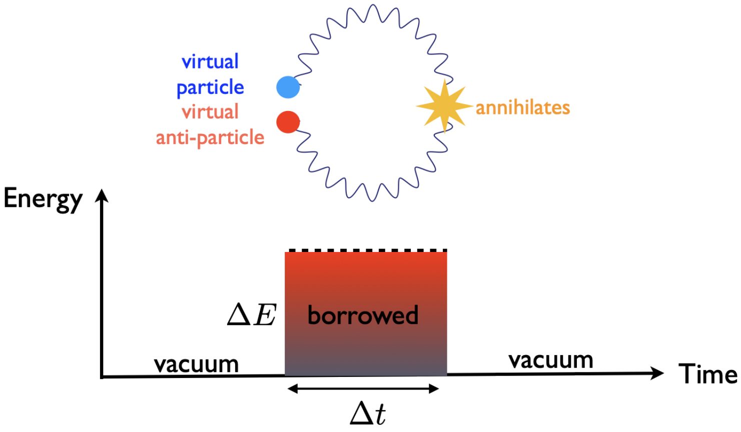


Figure 5.2: Empty space itself borrowed energy ΔE to make a pair of particle and anti-particle, which then self-annhilate after time Δt, with $\Delta E \Delta t \geq \hbar/2$.

Nevertheless, while energy doesn't have to be conserved, *momentum must be conserved* – the interest-free loan version of the Heisenberg Uncertainty Principle does not say there is uncertainty in the momentum. This will be crucial later.

[1] The reason why it has to be a pair of particle and its anti-particle is because they have to find a way to "pay back" the loan somehow, and the only way for two particles to return the energy is to self-annihilate. Free energy, like free money, is great and you can do a lot of crazy things with it, but it still has its limitations.

5.2.2 Black Hole Thermodynamics Again and Hawking Radiation

Such a quantum fluctuation of empty space itself is known as **vacuum fluctuation**, and it is indeed all around us. You are surrounded by vacuum fluctuation, and this effect is physically measured! So far so good. But now consider what would happen if such a fluctuation occurs very near the event horizon of a black hole. A pair of particle and anti-particle appear, and suppose one of them appear *inside* the black hole, and the other appear *outside*, then the one inside will fall into the black hole, and the one outside will fall "away" from the black hole, see Fig. 5.3.

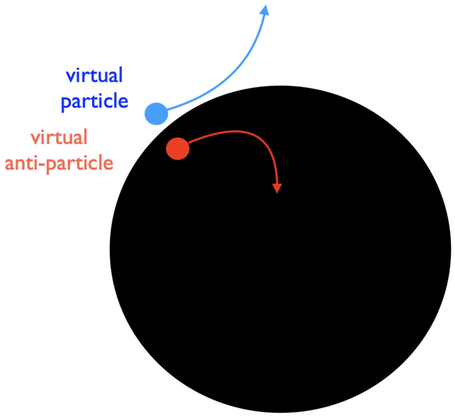


Figure 5.3: A pair of virtual particles were created near the black hole horizon, and one falls into the black hole, and the other falls away from the black hole to become Hawking radiation.

The virtual particle that falls away from the black hole will become **Hawking Radiation**, while the virtual particle that falls into the black hole will reach the singularity in the middle. Since the outgoing radiation is now "real" and observable, the loan has to be paid by someone – the black hole. Thus the black hole *loses mass in the process*. Hawking did this calculation in detail, and not only did he compute the fact that such radiation can occur, he showed that the radiation actually has a blackbody form with the **Hawking temperature**

$$T_{\rm BH} = \frac{\hbar c^3}{8\pi k_b GM} \ , \tag{5.5}$$

where M is the mass of the black hole. This remarkable formula includes the fundamental constants of all the physics we have studied: c for the speed of light and special relativity, G for gravity and general relativity, $\hbar$ for quantum mechanics and k_b for statistical mechanics! Thus a black hole is not black, but radiates with a temperature which is *inversely proportional* to its mass – the smaller the mass of the black hole, the more intense the radiation. For the black holes that we have already detected using the LIGO observatory that is about 20 times the mass of the sun, their temperature is roughly 3×10^{-7} K, which is very tiny and indeed much cooler than the present temperature of the universe[2] $T = 2.73$ K.

Hawking then took this result to the natural conclusion – given that we have a temperature, then we can use the 2nd law to calculate the entropy. Since the heat transferred out of the black hole must be

[2] This actually imply that these black holes are still absorbing more energy than radiating them, so they are getting even colder and radiate even less.

related to the mass loss, thus $dQ = c^2 dM$, using $E = mc^2$, we get

$$\begin{aligned} dS_{\rm BH} &= \frac{dQ}{T_{\rm BH}} \\ &= \frac{8\pi k_b G}{\hbar c} M dM \\ &= \frac{4\pi k_b G}{\hbar c} dM^2 \ , \end{aligned} \tag{5.6}$$

which we can then integrate to get

$$S_{\rm BH} = \frac{4\pi k_b G M^2}{\hbar c} \ . \tag{5.7}$$

Recall that the Schwarzschild radius Eq. (3.31) is $r_{sch} = 2GM/c^2$, and hence $M^2 = r_{sch}^2 c^4/4G^2$, so the above Eq. (5.7) the becomes

$$S_{\rm BH} = \frac{\pi k_b r_{sch}^2 c^3}{G\hbar} \ . \tag{5.8}$$

Using the fact that the area $A_{\rm BH} = 4\pi r_{sch}^2$, we finally get

$$S_{\rm BH} = \frac{k_b A_{\rm BH} c^3}{4G\hbar} \ , \tag{5.9}$$

which is exactly the **Bekenstein-Hawking Entropy** Eq. (4.17) that we have discussed in chapter[3] 4.

5.2.3 Where did all the information go?

Since the black hole radiates, and if we wait long enough, it will eventually radiate away all its mass, and *poofs*, disappear, having converted all its mass into radiation which is at a higher entropy than the black hole itself (and hence obey the 2nd Law). So, will the black hole radiates out all the information that has been thrown in? In other words, if I have made a black hole by throwing in a set of encylopaedia of important physics knowledge, then can I recover this knowledge by collecting all the radiation that comes out of the black hole and somehow reconstruct the books? Does the black hole radiate out what we used to make it?

The answer is a resounding **no**! The problem is as follows. Recall that each radiated particle is half of the virtual particle-anti-particle pair. Now, while it doesn't matter which of this pair of particles, i.e. either the particle or anti-particle, gets radiated away, the key point is that if the radiated particle is an anti-particle, then the particle that falls back into the black hole must be a particle, and *vice versa.* This means that the pair of particles are *entangled*, exactly the way we have described in section 1.2.2. Let's use ↑ to denote a particle and ↓ to denote an anti-particle, the quantum state of this pair of virtual particles is

$$\psi = \sqrt{\frac{1}{2}}\left(\uparrow\downarrow + \downarrow\uparrow\right) \ . \tag{5.10}$$

How much information is contained in such a pair? It turns out that it contains *exactly a single bit of information*[4]. As we have learned in chapter 4, each bit of information is physical – you can use it to do $k_b \ln 2$ of work. But the problem is that *we need to bring both pair of particles together to be able to reconstruct this single bit of information.* If you like, to be able to reconstruct the state Eq. (5.10), we need the data from the measurement of both particles[5]! But since the infalling particle goes back into

[3] Sometimes the Boltzmann constant k_b is dropped in the Bekenstein-Hawking formula Eq. (5.9). The Boltzmann constant strictly speaking is not a fundamental constant – it is simply a conversion factor that allows us to convert from temperature to energy and *vice versa.* In fact, statistical mechanics tells us that temperature and energy are closely related and we should have measured them using the same units. Of course we didn't know this fact until late on, and hence k_b was the result of this historical oddity.

[4] Technically, the **entanglement entropy** is $k_b \ln 2$, which is also the entropy of a single classical bit.

[5] Or more precisely, the statistics of many measurements of a pure ensemble of Eq. (5.10) states.

the black hole, and when the black hole eventually radiated until it *poofs*, we have no way of recovering the other part of the entangled state. In other words, Hawking radiation is entangled with the stuff in the black hole, but when the black hole evaporates, suddenly the radiation has nothing to be entangled with, and hence the entanglement information is lost. This is the **black hole information paradox**, and at present it is an unsolved problem. Physicists believe that a resolution of this problem will require a deep understanding of quantum gravity.

5.3 Why is it so hard to unify Quantum Mechanics and Gravity?

We have finally reached the last part of our short introduction to modern concepts in theoretical physics. In this section, we will put together all the knowledge we have learned to tell you why a theory of quantum gravity – a theory which unify quantum mechanics and gravity is so hard.

5.3.1 Unification of quantum mechanics and special relativity

"Unification" means that we want to find a way to write down the equations that describe the apparently different theories, which are seemingly inconsistent with another, in a consistent whole such that they are *predictive*. There are roughly two flavour of "unifications". There is the so-called **unification of forces**, where physicists have showed that the weak nuclear force and the electromagnetism can be unify into a single electroweak force – presently physicists are trying to unify the electroweak and the strong nuclear force into a **Grand Unified Theory**.

Today, we want to talk about the second kind of unification, which is the unification of the theoretical framework which underlies our physical laws themselves. A classic example is the unification of quantum mechanics and special relativity – we have alluded to the fact when we discussed the "interest free loan" version of the Heisenberg Uncertainty Principle in section 5.2.1 that quantum mechanics was initially developed for non-relativistic particles – the so-called **Schrödinger's Equation** that describe the quantum state of a particle was not invariant under Lorenz Transformation. When physicists tried to generalize quantum mechanics to relativistic particles in the early part of the 20th century, i.e. make quantum mechanics Lorentz invariant, they started to find all sorts of inconsistencies and unexplained phenomena – anti-particles, spins of particles etc, which at that time were undiscovered. By unifying special relativity and quantum mechanics, these "unexplained phenomena" became predictions, and when the first anti-particle, the **positron** (predicted by Paul Dirac) which is the positive charged version of the electron, was discovered in 1932 by **Anderson**, it was clear that physicists were on the right track. Nevertheless, it took a while to iron out all the issues with this unification, and the result – **quantum field theory** – is now one of the most remarkable achievement of theoretical physics. We have successfully unified c and $\hbar$.

5.3.2 What we know and how do we put them together?

The next goal is to unify c, $\hbar$ and G. We have all the ingredients – let's put them all in one place.

- **Quantum Mechanics** : The Heisenberg Uncertainty Principle tells us that empty space is full of quantum fluctuations, and you can borrow free energy ΔE if you pay it back without anyone observing it within Δt such that

$$\text{Rule QM1 :} \quad \Delta E \Delta t \geq \hbar/2 \ , \tag{5.11}$$

.

Furthermore, quantum mechanics tell us that the probability amplitude A of any event is given by a **sum over all histories**, i.e.

$$\text{Rule QM2}: \quad A = \sum_{\text{all paths}} e^{iS/\hbar} \ , \tag{5.12}$$

where S is the **action** which describe the physics of this event.

- **Special Relativity** : Energy and mass are equivalent, and are related by the formula

$$\text{Rule SR}: \quad E = mc^2 \ . \tag{5.13}$$

- **General Relativity** : General relativity tells us that if we compress a quantity of mass m smaller than its Schwarzschild radius r_{sch} given by

$$\text{Rule GR}: \quad r_{sch} = \frac{2GM}{c^2} \ , \tag{5.14}$$

then a black hole will form. The center of the black hole is a singularity, which is an infinity.

Recall from section 5.2.1 that the Rule QM1 (Heisenberg Uncertainty Principle) allows an electron to borrow free energy to make a photon. Since the more energy you borrowed – and hence the more energetic the photon is – the shorter time you can keep the photon, the photon can travel less distance before it was reabsorbed to pay back the loan. We can represent this by drawing smaller loops as shown in Fig. 5.4.

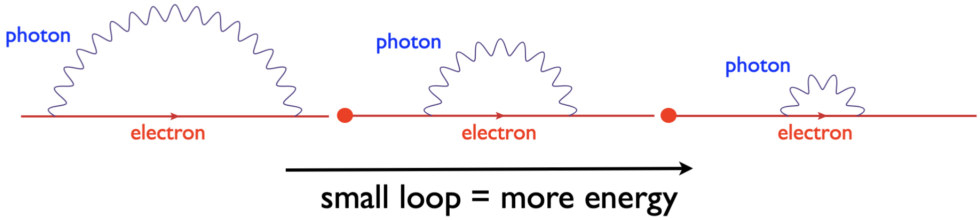


Figure 5.4: The more energy you borrow, the more energetic the photon, but the less time it lives, so it travel less distance.

In fact, you can take out more than one loan! And not only that, the *virtual particles themselves can also take out loans*, leading to all the following possibilities shown below in Fig. 5.5.

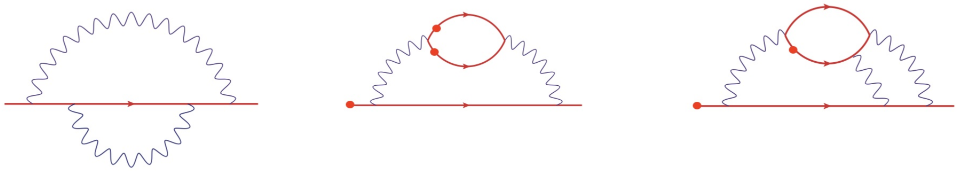


Figure 5.5: Other possibilities afforded to us by the Heisenberg Uncertainty Principle – there is an infinite number of ways the electron can move from left to right.

Each of these diagram is a possible history of the electron moving from left to right. Quantum mechanics (Rule QM2) now tells us that the probability amplitude A of an electron moving from left to right is the sum over all these possibilities (see Fig. 5.6).

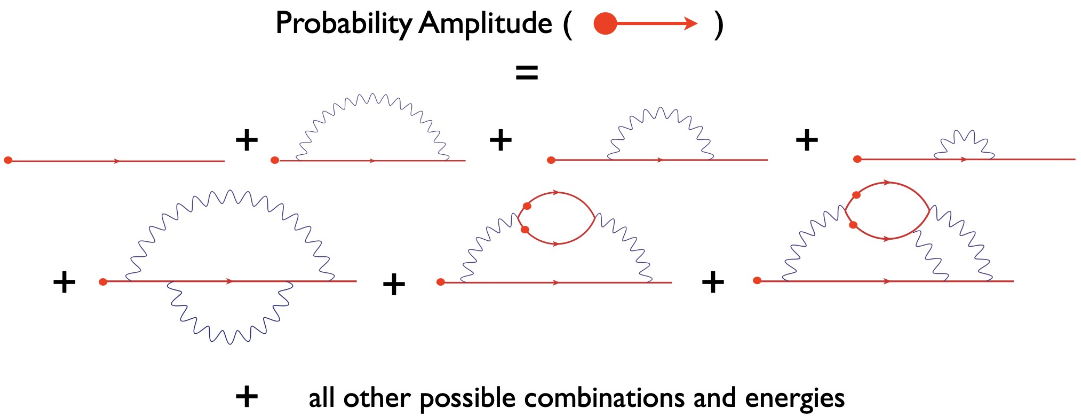

Probability Amplitude (   )
=
+
+
+
+
+
+
+ all other possible combinations and energies

Figure 5.6: The probability amplitude of an electron moving from left to right is the sum over all possible histories, taking into account the freedom afforded to us by the Heisenberg Uncertainty Principle.

The probability of an electron moving from left to right is then given by $|A|^2$ as usual. Such figures shown in Figs. 5.4, 5.5 and 5.6 are known as **Feynman diagrams** and is used by physicists to keep track of all the possibilities which they need to do when they calculate A. This calculation is actually non-trivial – indeed you will learn how to calculate such things when you take an advanced quantum field theory course and its knowledge separates the casuals from the hardcore. Nevertheless, it is the mathematics that's hard, not the idea behind it, which is as simple as we just described to you. When you sum over an infinite number of terms, you either get a finite number or an infinite number. If you get a finite number, that means that your theory actually makes sense – you get a finite number and give this prediction to your experimentalist friends who can then test your theory for you. Theories in which such a sum gives finite values are called **complete theories**[6]. In fact, the theory of the electron and photons which I just described to you is a complete theory known as **quantum electrodynamics**, which was formulated by Feynman, **Julian Schwinger** and **Shinichiro Tomonaga** who jointly won the Nobel prize for this work in 1965. It is verified experimentally up to 1 part in 10^{10} and is one of our most successful theory in terms of predictiveness.

5.3.3 Incomplete theories and how to fix it

What about theories that, after doing the sum over histories, give you an infinite number? Such theories are **incomplete**. As the name implies, it usually means that we are missing some crucial ingredient. Let's see how we can fix incomplete theories by considering a famous incomplete theory : the **Fermi Theory of Beta Decay**. A beta decay is when a neutron decays into a proton, emitting an electron in the process. **Enrico Fermi** realized that to conserve momentum, one would need to add a 4th particle which is very light, which he called the neutrino ("little one"), i.e. $n + \nu \rightarrow p + e$ (see Fig. 5.7). Such a process occurs, for example, when a Carbon-14 atom decays into a Nitrogen-14 atom.

[6] Actually, a naive summation of all the terms did lead to an infinite result, which was known as the **ultraviolet catastrophe**, named because the infinities seem to come from when we try to sum up terms which has very high energies (i.e. "in the ultraviolet"). Physicists realized that these infinities actually cancel via some somewhat dubious mathematical trick, and although eventually **Kenneth Wilson**, building on work done by **Leo Kadanoff** showed that it is consistent. Controversially, Wilson was awarded the Nobel prize in 1982 but not Kadanoff. This framework is known as **Renormalization**. Theories which can be "renormalized" are finite, while theories which are not "renormalizable" are infinite and incomplete.

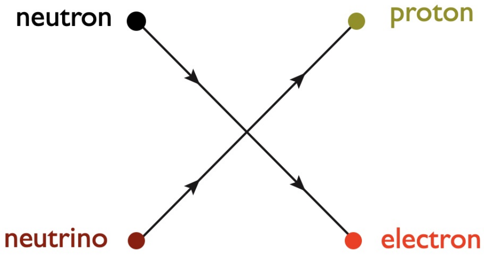


Figure 5.7: Fermi Beta Decay where a neutron interacted with a neutrino to make a proton and an electron.

Fermi's theory was experimentally verified in 1956 in the Cowan-Reines experiment, which detected the predicted neutrino – what they did was to look for the **inverse beta decay**, which is $\bar{\nu} + p \rightarrow n + e$. The theory actually works very up to a point – it will give very accurate predictions for low energy decays *if you ignore quantum mechanics* but even so its accuracy will begin to drop as the decay process become more and more energetic until it's completely wrong when the energies reach around a hundred times the mass of the proton. But worse still, if you now try to include quantum mechanics into the calculation, i.e. calculate the sum over histories allowing quantum fluctuations as shown in Fig. 5.8, we will find that the result is an infinity. Fermi Beta Decay is an incomplete theory.

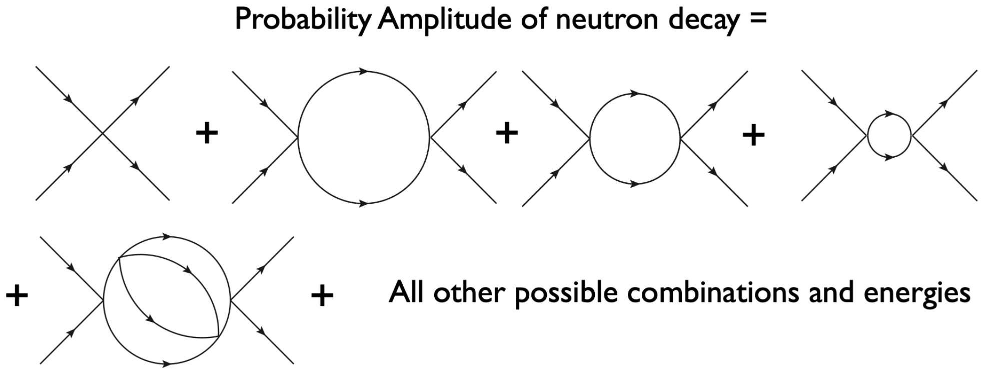

Probability Amplitude of neutron decay =
+
+
+
+
+
All other possible combinations and energies

Figure 5.8: The sum over histories of a Beta Decay process – we have dropped the particle names for simplicity. Each particle can borrow free energy to create virtual particles, which can also borrow energy to make virtual particles etc. This sum results in an infinity.

So how do we fix it? The fact that theory makes bad predictions at high energies suggests that perhaps there is a missing particle or force in the theory. Let's see how this might arise in the following way. Consider an electron borrowing some energy to make a virtual photon again, but instead of reabsorbing it when the time comes to pay it back, it was absorbed by *another nearby electron*, see Fig. 5.9.

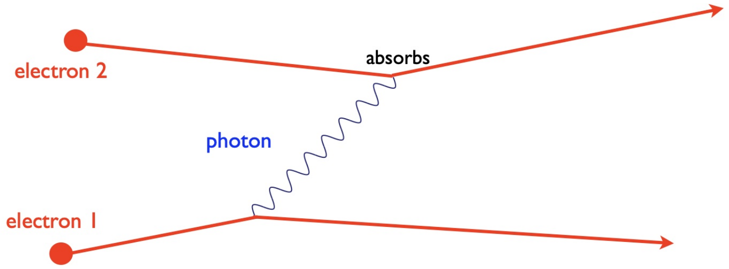


Figure 5.9: Two electrons can exchange a virtual photon, generating a force between them. This is actually how electrostatic force works!

While the energy is paid back, remember that since *momentum must be conserved*, the first electron has given some of its momentum to the photon, but hasn't gotten it back. Instead the momentum gets transferred to the second electron – thus momentum has been exchanged between the two electrons. In other words, applying Newton's 3rd Law, *there is a repulsive force between the two electrons*, exactly how you would expect when you try to put two electrons close together – an electrostatic force will repulse them. In fact, this is the origin of all forces – they are nothing but exchanges of particles.

Now, if we compare the diagrams in Fig. 5.7 and 5.9, suggests that if there is a missing particle or force in the Fermi Beta Decay process, perhaps it has the structure in Fig. 5.10, where we call this new particle or force W.

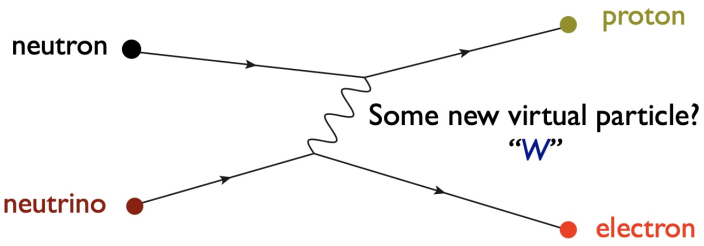


Figure 5.10: A hypothesized W particle that acts as a force between the two pair of particles.

What is the mass m_W of this particle? Special Relativity Rule SR says that if we borrow enough $E > m_W c^2$, then Rule QM2 says that it *must* be incorporated into our calculations. Also, we know that the mass of m_W must be big – if it is small we would have seen it already! In fact, since Fermi Beta Decay theory begins to give bad predictions around a hundred times the mass of the proton, we suspect that maybe this is the mass of m_W. Actually, this is exactly right! In fact, we are missing more than one particle, but instead a trio of particles W_+, W_- and Z plus a very special particle H, which all have more or less the mass of a hundred times the mass of the proton. If we incorporate all this into our theory, and recalculate the Feynman sum over histories, we find that it will give us finite numbers – the theory is completed by adding more massive particles into it. $W_\pm$ and Z was discovered in the Super Proton Collider in 1983 (1984 Nobel prize). This force carried by $W_\pm$ and Z is known as the **weak force**.

The last particle H is a very special particle – it is the **Higgs** particle, and it was needed to give the particles the right masses in a self-consistent way. But the key point is that when we discovered $W_\pm$ and Z, we *knew* that H must be there, because or else the theory is no longer complete! This sureness is what convinced us to spend so much money to build the Large Hadron Collider, which eventually detect the Higgs in 2012 (and yes, another Nobel prize). This complete theory which combined quantum electrodynamics and electroweak force is now known as the **Electroweak theory**.

This idea – an incomplete theory which breaks down at some energy level often means that there is a missing particle with masses around the energy is a powerful organizing principle in physics. When we encounter a theory which gives us some infinities, it is a way to tell us that "there are some new physics which we don't understand hiding somewhere when we go to higher energies"[7]. We can break down the life-cycle of a particle physicist's thinking process in Fig. 5.11.

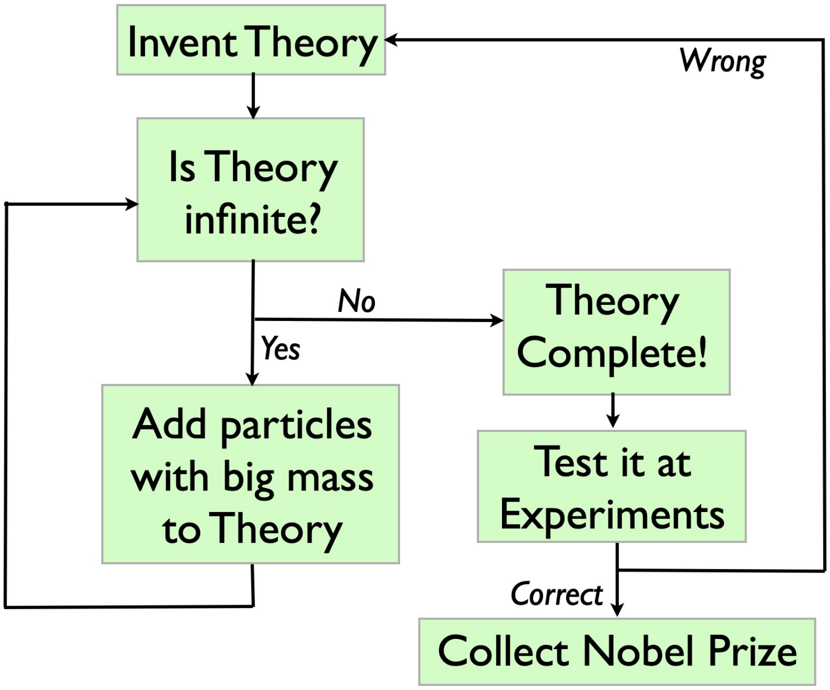


Figure 5.11: The idea that there are some new physics hiding in our non-renormalizable theories is a powerful organising principle – it allows us to systematically search for new physics.

5.3.4 Fixing Gravity?

What about gravity? Is it a finite complete theory? In other words, if we force gravity to obey the rules of quantum mechanics, then it must be subject to quantum fluctuations. Gravity is a force, so there must be a force carrier particle. The force carrier for quantum electrodynamics is the photon, which is also light waves. Einstein told us that a quanta of light have energies $E = \hbar\omega$ where ω is now the *wavelength* of the light waves. In other words, light can be considered as a particle (the photon) or waves. We know that gravity waves exists (we detected it in 2015!), so if we apply quantum mechanics to the gravity waves, then gravity waves can also be considered to be a particle called the **graviton**, which is then the force carrier for gravity.

In Einstein's General Relativity, the only possible vacuum wave solutions are actually gravity waves, so gravitons are the only "particle" in gravity. Nevertheless, we can apply quantum mechanics Rule QM2 (Heisenberg Uncertainty Principle) to it, and hence gravitons are now allowed to borrow energy to make virtual gravitons, which can also borrow energy to make more virtual gravitons etc. This means that the probability of a graviton moving from left to right, according to Rule QM2 (sum over histories), must look like Fig. 5.12.

[7] Nowadays, non-renormalizable theories are considered low energy **effective theories** of some more complete UV theory.

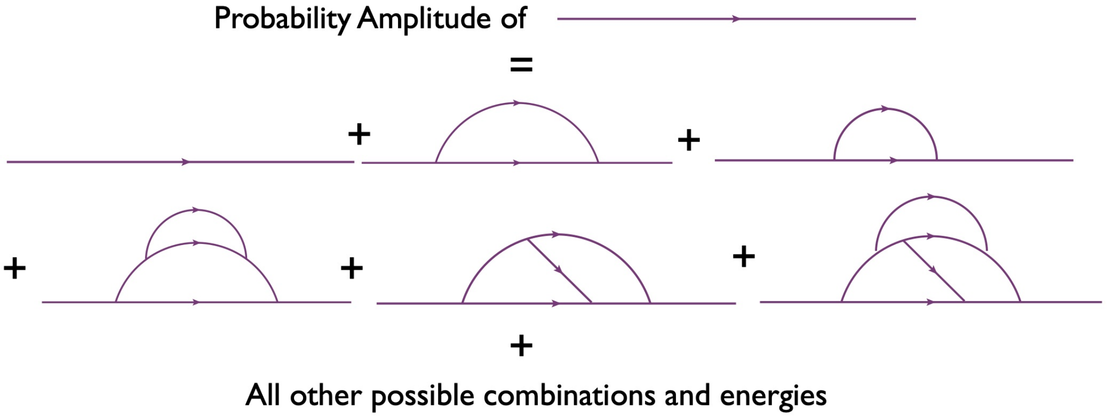


Figure 5.12: Applying quantum mechanics to gravity, the probability amplitude of a graviton moving from left to right is then the sum over all the histories allowed by the creation of virtual gravitons.

As you probably have guessed (else this lecture won't exists!), when we do the sum, we get an infinity. The quantum theory of general relativity is not a complete theory.

But wait! Didn't we just show you how to fix a theory? Maybe there is a missing force or particle that is massive? Let's call this particle P. What would be the mass of P, M_P? Well, we know that general relativity works really well, special relativity works really well, and quantum mechanics works really well. The fundamental constants associated with these theories are G, c and $\hbar$ respectively. So if the theory breaks down because we put them together, then perhaps we can guess that the M_P must be related to all three constants. In fact, there is only one way to put the constants together to make a mass, which is

$$M_P = \sqrt{\frac{\hbar c}{G}} = 2.18 \times 10^{-8} \text{ kg} , \tag{5.15}$$

which is known as the **Planck mass**. While it is small in terms of kilograms, it is actually ginormously huge in terms of particle mass – it is 10^{19} times the mass of the proton!

Suppose now a virtual P particle is emitted. Since it is so massive, it takes a lot of energy $\Delta E = M_P c^2$, which according to the Heisenberg Uncertainty Principle, it will only live for a very short time given by $\Delta t = \hbar/\Delta E = \hbar/(2M_p c^2)$. This means that, even if it travels at the speed of light, it will at most travel the distance

$$d = c\Delta t = \frac{1}{2}\frac{G}{c^2}\sqrt{\frac{\hbar c}{G}} = \frac{1}{2}\frac{GM_P}{c^2} = \frac{r_{sch}}{4}. \tag{5.16}$$

In other words, P's whole virtual life is spent within its Schwarzschild radius r_{sch}! General relativity Rule GR then tells us that P must collapse into a black hole. Inside the black hole is a singularity, and singularities are infinities. We tried to fix a infinity, and ended up with another infinity! Our attempt to unify quantum mechanics and gravity has failed.

The fact that I can explain this failure to you using the simple basic rules of special relativity, general relativity and quantum mechanics mean that it is a very deeply fundamental problem – it is not caused by some missing "trick" or difficult mathematics, but by just taking our very well known and tested theories and driving it to its natural conclusion. This makes the unification of quantum mechanics and gravity really hard – we have very little wiggle room! Something is deeply wrong with our understanding of the foundational theories.

5.4 A road towards quantum gravity : String Theory

So what next? There are many attempts to try to formulate a theory of quantum gravity. The most successful attempt at the moment is **String Theory** – by "successful" I mean it in comparison with other theories in terms of development status. It might be totally wrong still, but it has given us some tantalizing clues that perhaps we are on the right path.

The infinities of Feynman's sum over histories can be traced to the fact particles are assumed to interact at points in spacetime – and hence the Feynman diagrams look like lines. We can get rid of the infinities if "fatten" the lines. A fattened Feynman diagram then describes the interactions of *loops of strings* instead of particles as shown in Fig. 5.13.

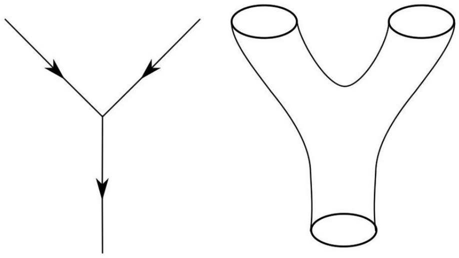


Figure 5.13: A fattened (right) Feynman diagram describes the interactions of closed loops of strings, instead of point particles. Sometimes the diagrams are called "pants" diagrams – very American so apologies to British people.

String theory has made two intriguing predictions which suggests that it is in the right path. First, *it automatically incorporate gravitons* – gravity is "built-in" and doesn't have to be added in by hand. Secondly, **Cumrum Vafa** and **Andrew Strominger**, using a particularly special version of string theory, successfully computed the Hawking-Bekenstein Entropy of a black hole. Unfortunately though, String Theory has made no explicit testable experimental predictions, and itself has a lot of theoretical inconsistencies which renders it at the moment more of a hopeful punt then a true theory of quantum gravity.

Happily though, this means that we all still have jobs.

5.5 Assignment Topics

- *The Large Hadron Collider and the Discovery of Higgs*: Describe briefly the Large Hadron Collider and how it was used to discover the Higgs particlea.
- *The Firewall*: Recently, physicists has suggested an alternative way to view the black hole information paradox problem known as the "the Firewall". Describe the Firewall, and explain why it is not a solution but an alternative formulation.
- *The Casimir Effect*: The vacuum fluctuation effect can be detected using the so-called Casimir Effect, which has been measured. Describe the experiment.

Epilogue

So this is the end of our short tour on the modern concepts of theoretical physics. They are deliberately made as simple as possible to illustrate the core ideas. Think of it as a terrible movie trailer which gives away all the plot points without telling you the details. Hopefully though, that these spoilers won't stop you from going out to actually learn the thing for real by pursuing a degree in physics!