4 The Weak Force

It is now time to turn to the weak force. For many years, there was just one manifestation of the weak force, namely beta decay

$$n \;\rightarrow\; p + e^- + \bar{\nu}_e$$

We now know that this can be understood in terms of the down quark decaying into an up quark, electron and anti-electron neutrino

$$d \;\rightarrow\; u + e^- + \bar{\nu}_e$$

There is nothing in the strong force or electromagnetism that would allow one type of quark to morph into another. We need to invoke something new.

That "something new" turns out to be something old and familiar. Nature has a tendency to re-use her good ideas over and over again, and the weak force is no exception. Like the strong force, it too is described by Yang-Mills theory. The difference is that the matrices are now 2×2 instead of 3×3. However, as we'll see in this section, it's not just the strengths of the forces that differ and the weak and strong forces manifest themselves in a very different manner.

The three forces of Nature together provide the foundation of the Standard Model. In mathematical language these forces are characterised by a group

$$G = SU(3) \times SU(2) \times U(1)$$

where the 3×3 matrix fields of $SU(3)$ describe the strong force, and the 2×2 matrix fields of $SU(2)$ describe the weak force. However, rather surprisingly the fields of $U(1)$ do *not* describe the force of electromagnetism! Instead, they describe an "electromagnetism-like" force that is called *hypercharge*. The combination of $SU(2)$ and $U(1)$ is sometimes referred to as *electroweak theory*. We will learn in Section 4.2 how electromagnetism itself lies within.

The weak force has few obvious manifestations in our everyday life and, in many ways, is the most intricate and subtle of all the forces. It is intimately tied to the Higgs boson and, through that, the way in which elementary particles get mass. Moreover, both the most beautiful parts of the Standard Model, and those aspects that we understand least, are to be found in the weak force.

Figure 30. Parity violation of Chien-Shiung Wu.

4.1 The Structure of the Standard Model

When describing the strong force, we saw that it affects some particles (we call these quarks) while leaving other untouched (we call these leptons). Our first task now should be to describe which particles are affected by the weak force.

You might think that we could simply list those particles that feel the weak force. But, as we will see, things are not quite so straightforward. It turns out that the weak force acts on all the particles in the universe. But it does so by acting on exactly half of each particle!

4.1.1 Parity Violation

There is one defining characteristic of the weak force and hypercharge that differentiates them from the strong force (and from electromagnetism). They do *not* respect the symmetry of *parity.*

This fact was discovered by Chien-Shiung Wu, on a cold winters day in New York City in December 1956. Wu's experiment was technically challenging but conceptually very simple. She placed a bunch of Cobalt atoms in a magnetic field and watched them die. Cobalt undergoes beta decay

$$ {}^{60}\text{Co} \ \rightarrow \ {}^{60}\text{Ni} + e^- + \bar{\nu}_e + 2\gamma $$

with a half-life of around 5.3 years. The two photons arise because cobalt first decays to an excited state of the nickel nucleus, which subsequently decays down to its ground state emitting two gamma rays. The whole point of the magnetic field was to make sure that the nucleon spins of the atoms were aligned. Wu discovered that the electrons were preferentially emitted in the opposite direction to the nucleon spin

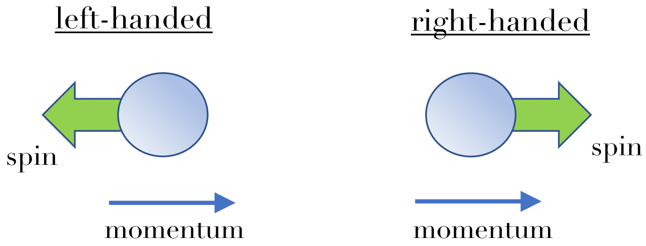


Figure 31. The handedness of a massless particle is determined by the relative direction of its spin and momentum.

This may sound innocuous but we now realise that it was one of the most significant discoveries in all of particle physics. The key point is that when we say the nuclei spins in a given direction, we mean in a *right-handed* sense. Take your right-hand and curl your fingers round in the way the nuclei are spinning: then your thumb points in the "direction of spin" which, as Wu observed, is opposite to the motion of the electrons.

Now suppose that, for some bizarre reasons, Wu looked at her experiment reflected in a mirror. The directions of the electrons would remain unchanged but the spin of the nuclei would be reversed, because right-hands are reflected into left-hands. This means that, viewed in a mirror, Wu would have come to the opposite conclusion: the electrons would be preferentially emitted in the *same* direction as the nuclei spin.

To put this in context, stare at the two photographs of Wu and her experiment shown in Figure 30. If you look closely you can tell which photograph is the original and which is flipped about an axis. (For example, one way to do this is to note that the writing is only legible on the left-hand picture.) Wu discovered that the same is true of the laws of physics at a fundamental level: you can tell if you're looking directly at sub-atomic particles or viewing them reflected in a mirror. There are things that can happen in a mirror that cannot happen in our world! This property is known as *parity violation*.

How can we write down theories which violate parity, meaning that they describe a world which looks different when reflected mirror? The key is something that we learned back in Section 2.1.4: any massless spin $\frac{1}{2}$ particle decomposes into two pieces, called *left-handed* and *right-handed*. Recall that a right-handed particle is one whose spin is aligned with its momentum, while a left-handed particle has spin and momentum anti-aligned. This distinction only makes sense for massless particles since they travel at the speed of light and so all observers, regardless of their own motion, agree on the direction of spin and momentum.

To write down a theory which violates parity is then straightforward: we simply need to ensure that the left-handed particles experience a different force from the right-handed particles.

The weak force accomplishes this in the most extreme way possible: only left-handed particles experience the weak force. Right-handed particles do not feel it at all. For reasons that we now explain, this is the key property of the weak force and one of the key properties of the Standard Model.

There are quite a few things that we will need to unpick regarding the weak force. Not least is the fact that, as stressed above, the distinction between left-handed and right-handed particles is only valid when the particles are massless. A remarkable and shocking consequence of parity violation is that, at the fundamental level, all elementary spin $\frac{1}{2}$ particles are indeed massless. The statement that elementary particles – like electrons, quarks and neutrinos – are fundamentally massless seems to be in sharp contradiction with what we know about these particles! We learn in school that electrons and quarks have mass. Indeed, in the introduction to these lecture notes we included a table with the masses of all elementary particles. How can this possibly be reconciled with the statement that they are, at heart, massless? Clearly we have a little work ahead of us to explain this. We'll do so in Section 4.2 where we introduce the Higgs boson.

4.1.2 A Weak Left-Hander

We're now in a position to explain how the three forces of the Standard Model act on the matter particles. The short-hand mathematical notation for the forces is

$$G = SU(3) \times SU(2) \times U(1)$$

Let's first recall some facts from the previous chapter. The strong force is associated to the "$SU(3)$" term in the equation above. As we explained in Chapter 3, the analog of the electric and magnetic fields for the strong force are called gluons, and are described by 3×3 matrices. (This is what the "3" in $SU(3)$ means.) Correspondingly, each quark carries an additional label, that we call *colour* that comes in one of three variants which we take to be red, green or blue.

While quarks come in three, colour-coded varieties, the leptons – i.e. the electron and neutrino – do not experience the strong force and hence they come in just a single, colourless variety. In the introduction, we said that each generation contains four particles: two quarks and two leptons. However, a better counting, including colour,

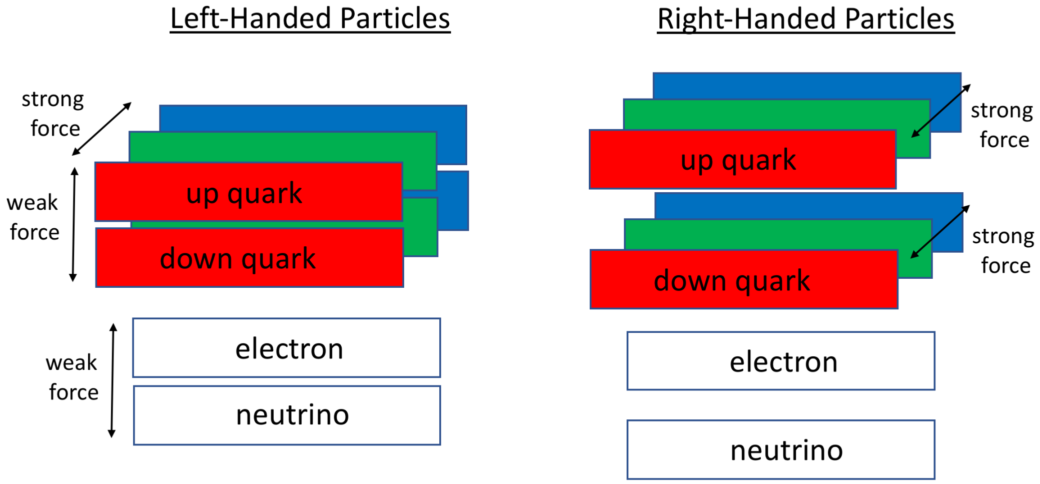


Figure 32. The jigsaw of the Standard Model. Only quarks experience the strong force. Only left-handed particles experience the weak.

shows that each generation really contains 8 particles (strictly 8 Dirac spinors). There are $3+3$ from the two quarks, and a further $1+1$ from the two leptons.

The next step in deconstructing the Standard Model is to note that each spin $\frac{1}{2}$ particle should really be decomposed into its left-handed and right-handed pieces. Only the left-handed pieces then experience the weak force $SU(2)$. If we were to follow the path of the strong force, you might think that we should introduce some new degree of freedom, analogous to colour, on which the weak force would act. A sort of weak colour. Like pastel. In fact, that's not necessary. The "weak colour" is already there in the particles we have.

This is illustrated in Figure 32. The right-handed particles are the collection of coloured quarks and colourless leptons. The left-handed particles are the same, except now the weak force acts between the up and down quark, and between the electron and neutrino. In other words, the names of distinct particles — up/down for quarks and electron/neutrino for leptons — are precisely the "weak colour" label we were looking for! We've denoted this in the figure by placing the weak doublets in closer proximity.

This should strike you as odd. For the strong force, the red, blue and green quarks all act in the same way. We say that there is a symmetry between them. However, it's very hard to make the same argument for the "weak colour" label. The electron and neutrino are very different beasts. If we're really introducing "weak colour" in analogy with actual colour, surely there should a symmetry between them. What's going on?

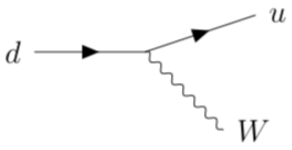
This brings us to our second striking fact: at the fundamental level, there is no distinction between a left-handed electron and left-handed neutrino! Nor is there a distinction between a left-handed up quark and left-handed down-quark. These particles share all their properties. However, as the story of the Standard Model plays out, the Higgs field intervenes. In addition to giving the elementary particles mass, the Higgs fields also leaves the electron/neutrino and up/down pairs with the distinctive characteristics that we observe. This aspect of the Standard Model is called *symmetry breaking* and will be described in Section 4.3.2.

Gauge Bosons

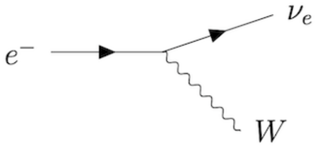
As with the other forces that we've met, there are spin 1 particles associated to the weak force. These are the analogs of the photon for electromagnetism and the gluon for the strong force. The spin 1 particles associated to the weak force are called *W-bosons* and *Z-bosons*. We'll see the difference between the W and Z later when we discuss the Higgs.

The type of spin 1 particles – like the photon, gluon, W and Z – that mediate forces are rather special and collectively go by the name of *gauge bosons*. (Here "gauge" is pronounced to rhyme with "wage".) The kind of Yang-Mills type theories that underly these are called *gauge theories*.

We can draw Feynman diagrams associated to the weak force. The W-boson interacts with the quarks, changing a down into an up like this

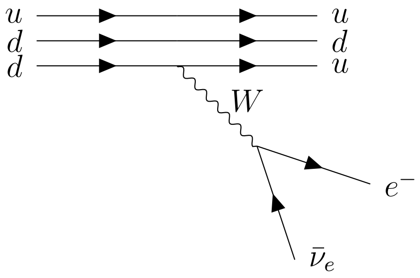


Note that the colour of the quark remains unchanged. For example, a red down-quark will turn into a red up-quark. Similarly, the W-boson interacts with the leptons like this.



In each case, the conservation of electric charge tells us that the W-boson must carry charge -1. We'll describe this in more detail later.

Already from these diagrams, we can start to understand how beta decay works. Recall that the quark content of the neutron is *udd*, while that of the proton is *uud*. One of the down quarks in a neutron decays into an up quark by emitting a W-boson. This is subsequently followed by the decay of the W-boson into an electron and anti-neutrino.



Note that we get an anti-neutrino rather than a neutrino because the arrow is now running backwards in time. We see that beta decay doesn't proceed through a direct interaction between quarks and leptons. It's mediated by the W-boson, in the same way that the electromagnetic force is mediated by the photon.

The fact that the down quark decays into an up quark, and not the other way round, can be traced to the fact that the down quark is (marginally) the heavier of the two. Moreover, the W-boson is heavier than the combination of the electron and neutrino (now the difference is not so marginal) so this decay too is allowed. These facts, like so many other things, come from the Higgs boson. By now, you should probably be getting the feeling that the Higgs plays an important part in all of this!

Hypercharge

Shortly we'll see how electric charge emerges in this story. But first, we need to understand its counterpart at a more fundamental level. This is something called *hypercharge* and is associated to the $U(1)$ factor in the Standard Model. It is a force that is closely related to electromagnetism. However, the various particles don't have the same charges under hypercharge as they do electric charge. Importantly, hypercharge does not respect parity and left-handed and right-handed particles carry different charges. These hypercharges are listed in Table 2, together with a summary of how the strong and weak forces act.

We've normalised the hypercharges in the standard way so that they come in units of 1/6. However, this is just convention and there's nothing deep in this choice. We could just as well multiply everything by a factor of six, so all hypercharges are integer. Notice that one advantage of the current normalisation is that the hypercharge for the

Particles		Strong	Weak	Hypercharge
Left-handed	quarks	yes	yes	+1/6
	leptons	no	yes	-1/2
Right-handed	up quark	yes	no	+2/3
	down quark	yes	no	-1/3
	electron	no	no	-1
	neutrino	no	no	0

Table 2. The Standard Model forces acting on each of the fermions in a single generation.

right-handed particles coincide with their electromagnetic charge. This, as we shall see, is no coincidence. However, the hypercharges for the left-handed particles are rather different.

Before we move, I should mention that there is one caveat. (Isn't there always!) We don't yet have direct evidence for the existence of the right-handed neutrino and there is a possibility that it doesn't exist! Indeed, many people would say that the right-handed neutrino should not be included in the list of particles in the Standard Model. From the table you can see that the right-handed neutrino is neutral under all three of the forces in the Standard Model and this makes it very challenging to detect. We'll see the indirect evidence for its existence in Section 4.4 where we describe more about neutrinos in general.

4.1.3 A Perfect Jigsaw

The particles and forces listed in Table 2 summarise 150 years of work (dated from Röntgen's discovery of X-rays), dedicated to understanding the structure of matter at the most fundamental level. The first thing that comes to mind when you see it is: what a mess! The individual elements comprise some of the most gorgeous objects in theoretical physics – the Dirac, Maxwell and Yang-Mills equations. And yet any semblance of elegance would seem to have been jettisoned at the last, with the different components thrown together in this strange higgledy-piggledy fashion. Why this collection of forces and particles? In particular, why this strange collection of hypercharges?

Happily, there is a wonderful and astonishing answer to these questions. The beautiful truth is simply: it could barely have been any other way.

The reason for this has its roots in Wu's observation that the world does not respect the symmetry of parity. As we explained above, we can account for this in the Standard Model by ensuring that left-handed fermions experience different forces from right-handed fermions. However, it turns out that this is not quite as straightforward to achieve as I've made out.

To explain why, I need to tell you a few further facts about the mathematics underlying quantum field theory. Quantum field theories in which left- and right-handed fermions experience different forces are called *chiral theories*. It turns out that chiral theories are particularly fragile objects, always teetering on the brink of mathematical inconsistency. Put bluntly, most chiral theories that you write down don't make any sense. If you write down a random collection of particles, with left- and right-handed components experiencing different forces, it is overwhelming likely that the equations will spit back stupid, and obviously wrong answers like "1=0".

If you want to write down a sensible chiral quantum field theory then there are a bunch of hoops that you have to jump through. These hoops are mathematical consistency conditions that the different forces must obey if the theory is to be sensible. One simple way of obeying these consistency condition is to ensure that left- and right-handed particles feel the same force but such theories don't exhibit parity violation. If you want to write down a theory with parity violation, you're obliged to work harder and find a delicate balance between the forces experienced by the left-handed particles and those experienced by the right-handed particles.

I won't describe all the consistency conditions, but here's a taster. Let's call the hypercharge of each fermion $\tilde{Q}_f$ where f labels the fermion. Then one consistency condition reads

$$\sum_{\text{left}-\text{handed}} \tilde{Q}_f^3 = \sum_{\text{right}-\text{handed}} \tilde{Q}_f^3 \tag{4.1}$$

To check that this works for one generation of the Standard Model particles listed in Table 2, you have to remember that each quark comes with three colours, while the left-handed fermions are really a pair under the weak force. We then have

$$3 \times 2 \times \left(\frac{1}{6}\right)^3 + 2 \times \left(-\frac{1}{2}\right)^3 = 3 \times \left(\frac{2}{3}\right)^3 + 3 \times \left(-\frac{1}{3}\right)^3 + (-1)^3 + 0^3$$

A successful solution to the mathematical consistency conditions, like the one above, is known technically as *quantum anomaly cancellation*. It's not a particularly enlightening name. For now, I can only tell you that underlying these conditions are some of the deepest and most powerful ideas from the mathematics of topology. Indeed, the subject of "quantum anomalies" is currently where the fields of mathematics and physics have their richest intersection. (You can read more about quantum anomalies, in their full technical glory, in chapter 3 of the lecture notes on Gauge Theory.)

Viewed through the lens of quantum anomalies, the Standard Model morphs from a seeming mess into a perfect jigsaw. The forces act on the various particles so that the mathematical consistency conditions are satisfied. If you try to change any small piece, the whole thing falls apart and ceases to be a sensible physical theory.

I should stress that the Standard Model is not the only chiral gauge theory. If you allow for entirely different forces, or different collections of particles, then you can write down other such theories. But the Standard Model is, arguably, the simplest chiral gauge theory. Although, at first glance, the Standard Model looks like a jumble of random forces and particles, it is instead a beautiful and surprising theoretical construct. You just have to look at it the right way.

Hypercharge and Fermat's Last Theorem

It's difficult to explain the quantum anomaly consistency conditions at a deeper level without getting into the full mathematics. But here's a quick calculation that I'm particularly fond of that will hopefully give a sense of what's going on. Take the set of particles listed in Table 2, and assign them the properties under the strong and weak force that are listed. But allow them to have almost arbitrary values $\tilde{Q}_f$ of hypercharge. The "almost" is there because I'll impose two restrictions. First, we'll take the right-handed neutrinos to be neutral. (This can be motivated on the grounds that we're not really sure that they exist!) Second, we'll take the hypercharges $\tilde{Q}_f$ to be rational numbers, of the form p/q where p and q are integers. There are good theoretical reasons to think that charges should be quantised in this way. Then we ask the question: what values of hypercharges satisfy the mathematical consistency conditions?

It turns out that some of the hypercharges can be immediately related to others through the consistency conditions. And some remain arbitrary. If you follow through the calculation, and do some fairly complicated change of variables, the equation (4.1) ends up turning into the equation

$$X^3 + Y^3 = Z^3$$

where X, Y and Z must all be integers. This is a very famous equation! Fermat's last theorem tells us that the equation has no non-trivial solutions. There are, however, trivial solutions like $1^3 + 0^3 = 1^3$. If you take this trivial solution and plug it back into the complicated change of variables, you'll discover the set of hypercharges listed in Table 2. These hypercharges may look random, but they're related to some very deep and beautiful mathematics. In particular, they're related to Fermat's last theorem!

Further Generations

As we explained in the introduction, the pattern of particles listed in Table 2 is repeated twice more. The second generation consists of the strange and charm quarks, together with the muon and muon-neutrino. The third generation consists of the bottom and top quarks, together with the tau and tau-neutrino.

We don't understand why there are three generations. However, the quantum consistency conditions tell us that each generation must come in a complete set. For example, there was a 20 year gap between the discovery of the tau lepton in 1975 and the discovery of the top quark in 1995. (The bottom quark was discovered in 1977). Yet no one doubted that the top quark was there because mathematical consistency required it. The whole theory doesn't make any sense without the top quark.

4.2 The Higgs Field

Finally, it's time to introduce the famous Higgs boson. This is, it turns out, the simplest particle in the Standard Model. But the way in which it interacts with other fields is, by far, the most intricate. And, as we shall see, it is ultimately the Higgs boson that ties everything else together.

The Higgs boson is the simplest particle because it has spin 0. Indeed, it is the only fundamental particle that we know of without spin. Fields without spin are also referred to as *scalar fields*.

Recall that spin $\frac{1}{2}$ fields are described by the Dirac equation and spin 1 fields by the Maxwell or Yang-Mills equations. Both were pretty enough to be put in picture frames in earlier chapters. Any spin 0 field, like the Higgs boson, is described by the *Klein-Gordon* equation. It's nowhere near as beautiful as earlier equations, largely because it is too simple: it lacks the subtleties and surprises that make the Dirac and Yang-Mills equations so special. Nonetheless, it would be slightly churlish to deny it a place on the wall so, for what it's worth, here is the Klein-Gordon equation

$$\boxed{\mathcal{D}_\mu \mathcal{D}^\mu \phi - V(\phi) = \lambda \psi\psi}$$

In this equation, $V(\phi)$ is the Higgs potential while the terms on the right-hand-side describe the couplings to the fermions in the theory. Both of these will be described in more detail below.

As we've alluded to earlier in these lectures, the Higgs field plays a number of roles and we'll elaborate on these as we go along. For now we mention only that the Higgs field does not experience the strong force, but it does feel both the weak force and hypercharge. We should therefore augment Table 2 listing the forces experienced by each particle with one further entry:

Particle	Strong	Weak	Hypercharge
Higgs	no	yes	+1/2

Table 3. The forces experienced by the Higgs boson

Note that, because the Higgs field has no spin, it doesn't decompose further into left- and right-handed pieces. It just is.

Like all other fields, ripples of the Higgs field give rise to particles. This is the Higgs boson, the last of the Standard Model particles to be discovered. It weighs in at a mass

$$m_H \approx 125 \text{ GeV}$$

making it the second heaviest particle in the Standard Model, after the top quark.

However, the real importance of the Higgs lies not in the particle (although that's certainly interesting!) but, as we'll now explain, more in a property of the field itself.

4.2.1 The Higgs Potential

Given that the Higgs field is simpler than all the others, why does it play such an important role? Well, there's something that a spin 0 field can do that higher spin fields cannot: they can "turn on" in the vacuum.

To understand what this means, recall that in the introduction we gave some intuition for a quantum field as an object that's constantly fluctuating. More precisely, the vacuum of space should be viewed as a quantum superposition of many different field configurations. An example of a typical configuration of the gluon field in the vacuum is reproduced in Figure 33. Given this, it would seem that all fields "turn on" in the vacuum. However, importantly, for any field with spin the *average* of all these field fluctuations always vanishes in the vacuum.

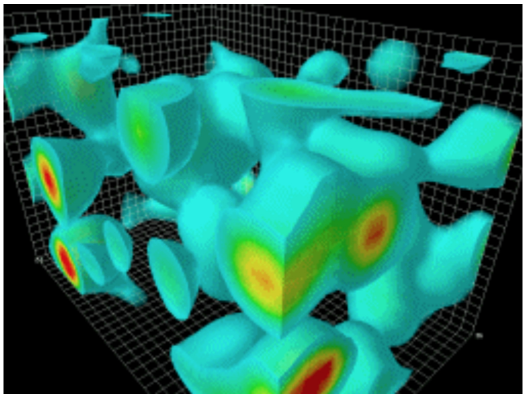


Figure 33. The fluctuations of a quantum field, taken from the simulations of Derek Leinweber.

The reason for this is simple: if a field with spin has a non-zero average, then it has to point in some direction. For example, if the average of the electric field **E** is non-vanishing then it picks out a direction in space. But the vacuum of space must look the same in every direction, so the average must be zero. (This statement is a little quick, but there are mathematical theorems that make it more precise.)

However, a scalar field is spinless, so doesn't point in any particular direction even when turned on. If we denote the field as $\phi(\mathbf{x}, t)$ then it's possible that, even in the vacuum, the average of all the fluctuations doesn't vanish, so

$$\langle \phi(\mathbf{x}, t) \rangle = \text{constant} \neq 0$$

where the angular brackets denote the average. In technical language, we say that the field *condenses*, a term that has its origin in the study of phase transitions and the condensation of water from vapour.

This new possibility brings up two immediate questions. What determines whether the field condenses? And what are the consequences if it does? Here we'll answer the first of these questions, postponing the second to Section 4.2.2.

The fate of a scalar field is not something that we get to chose. It is determined dynamically by the theory. Any scalar field experiences a potential energy that we call $V(\phi)$. This is a function that tells us how much energy it costs for the field to take certain values. Roughly speaking, there are two different shapes that these potentials take in theories of particle physics. These are shown in Figure 34.

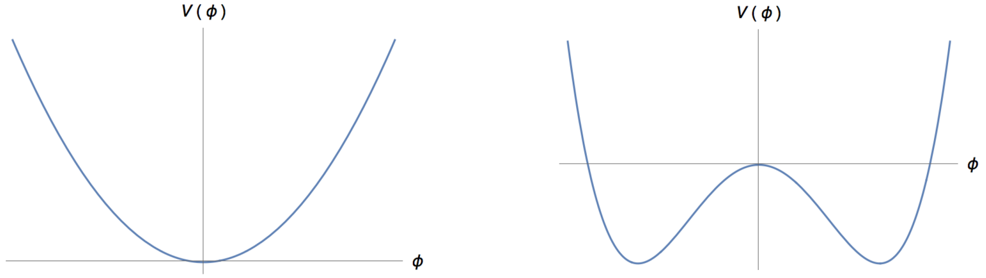


Figure 34. Two possible shapes for the potential for a scalar field. The Higgs field has a potential like that shown on the right, so that the field condenses, with $\langle\phi\rangle \neq 0$ in the vacuum.

In the vacuum, the scalar field sits at the minimum of the potential. If the potential has the shape shown on the left of Figure 34, then $\langle\phi\rangle = 0$ in the vacuum. However, if the potential has the shape shown on the right of Figure 34, then $\langle\phi\rangle \neq 0$ in the vacuum, and more interesting things happen. It turns out that the potential for the Higgs field in the Standard Model has the shape shown on the right, and this is what endows the Higgs with its power. (This statement is roughly true. A more accurate depiction of the Higgs potential will be given in Section 4.3.2.)

This, of course, brings up another interesting question: why does the Higgs potential in our world have the shape on the right, and not the shape on the left? We don't know the answer to that. At present, it is an input into the Standard Model and, hopefully, will be explained by some more complete theory in the future.

Here we are focussing on the question of whether the minimum of $V(\phi)$ sits at $\phi = 0$, or $\phi \neq 0$. But another question that we could ask is the value of $V(\phi)$ itself at the minimum. In the context of particle physics, this plays no role: it is just like any other potential energy, where only potential differences really matter and you can always add a constant to $V(\phi)$ without changing the physics. However, once we include gravity into the mix, the value of the potential energy becomes very important and contributes towards the cosmological constant. We'll say more about this in Section 5.3.1.

The Higgs Expectation Value

The upshot of the discussion above is that, even in the vacuum, the Higgs field averages to something non-vanishing. That something non-vanishing turns out to have the dimensions of energy. It is

$$\langle\phi\rangle \approx 246 \text{ GeV} \tag{4.2}$$

This is known as the Higgs *vacuum expectation value*. It is one of the key fundamental scales in the universe.

So what are the consequences of this? Well, they are pretty dramatic: the vacuum expectation value (4.2) turns out to give a mass to everything that it touches. This is known as the *Higgs mechanism* or the *Higgs effect*. The particles that get masses include both the W-bosons and Z-bosons that mediate the weak and hypercharge, together with all the spin 1/2 matter particles of the Standard Model. We'll postpone a discussion of fermion masses to Section 4.3, but the punchline is simply that the quarks, electrons and neutrinos get the masses that we advertised back in Section 1. Here, we will begin by focussing on the masses of the W-bosons and Z-bosons and some of their consequences. But first we will attempt, and largely fail, to give some intuition for why the Higgs field gives mass to particles at all.

Analogies for the Higgs Mechanism

The result of the Higgs expectation value $\langle \phi \rangle \neq 0$ is utterly startling. The Higgs field is like the ancient king Midas, but instead of turning everything to gold it makes everything massive. (Both make things heavier.) Why?

This is not an easy question to answer at the level of these lectures. What's perhaps unusual about this is that it's not at all difficult to understand the Higgs effect at the level of equations. Indeed, it's one of the simpler calculations in quantum field theory, but that doesn't change the fact that you do first need to learn quantum field theory. Largely, the difficulty in translating from equations to everyday language lies in the fact that the Higgs effect is really a phenomenon that is to do with fields rather than particles and we simply don't have much intuition for how these objects behave.

If we want to avoid the mathematics, we're obliged to rely on analogy. And, sadly, good analogies that relate the Higgs boson to more familiar, everyday phenomena are hard to come by. Here, for example, is a bad analogy. We could say that the Higgs field is like some kind of treacle. If you drag a spoon through treacle, you experience more resistance than if you drag it through water. Something similar happens with the elementary particles that interact with the Higgs. As we've seen, at the fundamental level all matter particles (as well as the W-boson and Z-boson) are actually massless and so should travel at the speed of light. But as they move in the vacuum, they have to plough through the Higgs field and this slows them down, effectively giving them a mass.

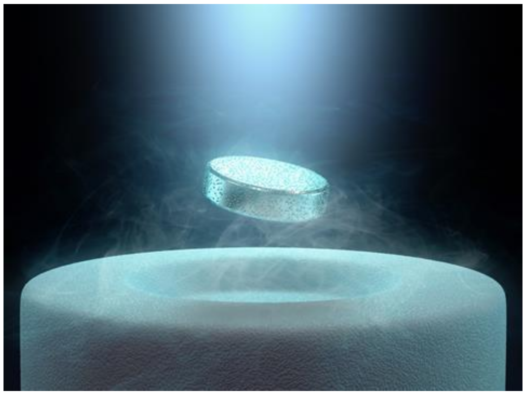


Figure 35. A superconductor and the Higgs boson are described by identical mathematics. (Image taken from the Royal Society of Chemistry).

In fact, this isn't a totally terrible analogy. But if you try to push it any further it will quickly break down. For example, the reason that the spoon slows down is because of friction forces, which means that it continually loses energy to the treacle. In contrast, there's no friction force going on with the Higgs, nor any constant transfer of energy from the particle to the Higgs field. The analogy just stops working at this point.

There is, however, one completely excellent analogy for the Higgs effect, although not one that's particularly everyday. This can be found in phenomenon of superconductivity. In certain metals (for example, aluminum or lead) the electrical resistivity plummets to zero as they are cooled to low temperatures. At the same time, the metal expels any magnetic field, giving rise to dramatic technological applications like levitating trains.

The theory behind superconductivity is well understood. First, the electrons bind together into pairs. This, already, is somewhat surprising since the electrostatic Coulomb force repels two electrons and so it seems unlikely that they would want to bind together. However, in a metal the sound waves (also known as phonons) give a second, attractive force between electrons and in favourable conditions this can win over, causing the electrons to bind. While the individual electrons have spin, the bound state can be spinless (rather like the mesons, formed of two quarks, that we met in Section 3.) It turns out that the potential for this bound state looks like the right-hand graph in Figure 34, and it condenses. The result is that, inside a superconductor, the photon gets a mass and this underlies behind all the subsequent phenomena, including the resistance free conduction and the expulsion of magnetic fields.

On a mathematical level, the analogy between the Higgs effect and superconductivity is exact: the key aspects of the equations describing the two are identical. It is a manifestation of the remarkable unity of physics, where the same ideas crop up in diverse situations.

4.2.2 W and Z Bosons

There are gauge bosons associated to all three forces in the Standard Model. Because the Higgs doesn't experience the strong force, the associated gauge bosons – which we named gluons in Section 3 – are unaffected by the Higgs boson. This is why we could discuss them earlier without dealing with all these subtleties.

In contrast, both the $SU(2)$ weak force and the $U(1)$ hypercharge interact with the Higgs. Which means that they get a mass. The result is three, massive spin 1 bosons:

$$\begin{aligned} W^\pm \text{ bosons:} \quad & M_W \approx 80 \text{ GeV} \\ Z \text{ boson:} \quad & M_Z \approx 91 \text{ GeV} \end{aligned}$$

Here the W^+ and W^- particles are distinguished by their ± electric charge. They are the anti-particles of each other. In contrast, the Z-boson is neutral. It is its own anti-particle. Note that both masses are a factor of 3 or so below the Higgs expectation value (4.2). This is not a coincidence: the expectation value sets the scale of the masses, with the reduction due to the strength with which the Higgs field interacts with these spin 1 bosons.

The fact that the force-carrying particles become massive greatly changes the properties of the force. Instead of the familiar $V \sim 1/r$ Coulomb potential of electromagnetism (or the less familiar $V \sim r$ confining force of QCD), the massive W- and Z-bosons give a potential energy between particles that takes the form

$$V(r) \sim \frac{e^{-Mr}}{r} \tag{4.3}$$

where r is the distance between two particles, while M is the mass of the W or Z boson. We've met a force of this kind before: it takes the same form as the Yukawa force (3.4) that is mediated by pions, and binds the protons and neutrons together in a nucleus. These forces have the characteristic feature that they become negligibly small for distances $r \gg 1/M$.

When, in Section 3, we discussed the force mediated by mesons, their mass was $m \approx 140$ MeV and the force extended over a distance $\approx 2 \times 10^{-15}$ m. Which, of course, is the size of the nucleus.

For the weak force, the masses of the force-carrying particles are almost 1000 times heavier than the meson, so the distance over which the weak force acts is almost a 1000 times smaller than the size of the nucleus. That's a pretty small distance scale! Indeed, it's so small that it means our old Newtonian way of thinking about forces as acting between particles really isn't particularly useful anymore. Unlike electromagnetism and the strong force, there are no examples where the weak force can be viewed of as an attractive force that make things stick together. There are no atoms or mesons of the weak force. Instead, as we've already seen, the primary role of the weak force is one of decay. Its job is to rent asunder that which the strong put together. This affects the neutron through beta decay and many other particles whose lives are cut tragically short by the weak force. We will describe this in more detail shortly.

The One That Got Away

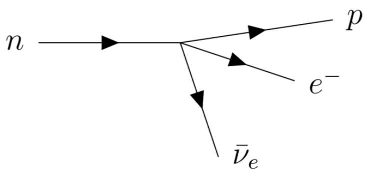
There's one final, very important twist to the story above. As we've seen, the gauge bosons associated to both the $SU(2)$ weak force and the $U(1)$ hypercharge interact with the Higgs. However, there's one special combination of these gauge bosons where the weak interaction is precisely cancelled by the hypercharge interaction. In other words, one combination of spin 1 bosons manages escape the attention of the Higgs and therefore remains massless. This is the photon.

This gives us a new perspective on the familiar electromagnetic force. Electromagnetism is not one of the three fundamental forces in the Standard Model, but instead is a combination of the weak force and hypercharge. It is special, because it is the only combination uncontaminated by the Higgs boson.

For this reason, the $SU(2) \times U(1)$ weak and hypercharge forces are referred to collectively as the *electroweak* theory. It is sometimes said that they are a unification of the weak and electromagnetic forces. While it's certainly true that the weak and electromagnetic forces are intricately interwoven, it's not quite correct to say that they're "unified". Indeed, it might be better to say that they're divided by the Higgs: the weak force became mired in the Higgs condensate, while the electromagnetic force is the one that got away.

4.2.3 Weak Decays

Enrico Fermi was the first person to understand beta decay. In 1934, just 18 months after the discovery of the neutron, he proposed a simple quantum field theory in which a neutron can decay into a proton, an electron and an anti-neutrino. The associated Feynman diagram is



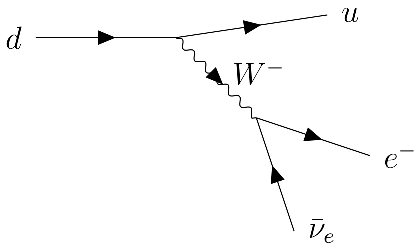
Thus, right from the beginning, the story of the weak force was one of decay.

Fermi's theory was one of the great breakthroughs of particle physics. Not only did it give a correct explanation for beta decay, but it was the first time that a quantum field theory was written down in which a particle of one type can transmute into something else. Any idea that the neutron was composed of a proton plus electron (or, as was also suggested, the proton was composed of a neutron + positron) were consigned to the waste bin.

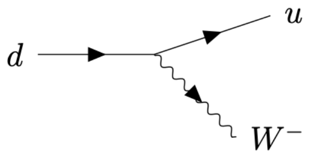
It took several more decades to understand what things look like if we zoom in a little further. The structure of the electroweak theory in its essentially complete form was first understood by Steven Weinberg, but many others got close including Sheldon Glashow, Abdus Salam and John Ward.

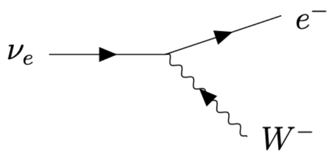
W-bosons

We know that if we zoom into the neutron and proton we find quarks. Beta decay occurs when a down quark changes into an up quark. However, this process doesn't happen through a direct interaction of four fermions: it is mediated by the W-boson, and looks like this.



Let's unpack this a little. There are two vertices that involve the W-bosons. The relevant sub-diagrams look like this:





We've met both of these before, except this time we've specified that it's the W^- boson involved in the process. This is needed to ensure the conservation of electric charge. As usual, diagrams with W^+ can be formed by flipping the arrows or, equivalently, replacing particles with anti-particles.

There are a few things to say about beta decay when mediated by a W-boson. First, the reason that down quarks decay into up quarks, rather than the other way around, is because the mass of the down quarks is heavier than the masses of the decay products

$$m_d > m_u + m_e + m_{\nu_e}$$

We'll explain more about how these masses arise from the Higgs field in Section 4.3, but at this stage I'll simply point out that we don't currently have any understanding of why the masses are ordered this way rather than some other way. Nonetheless, this is crucial for beta decay to take place while conserving energy. The remaining energy $\Delta E = m_d c^2 - m_u c^2 - m_e c^2 - m_\nu c^2$ goes into the kinetic energy of the decay products.

If the masses had been ordered in some other way, then we would have a world in which, say, the electron could decay into a down and anti-up quark, together with a neutrino. Indeed, you can easily write down a Feynman diagram for such a process, but it's forbidden in our world because of energy conservation.

However, this begs a new question. The first sub-diagram in beta decay looks like a down quark decaying to an up quark and a W^- boson. But the W^- boson is way heavier than the up and down quark. Hence, energy conservation would suggest that such a decay is impossible. However, internal particles in diagrams – sometimes called *virtual particles* – are not subject to the same strict rules of energy conservation as external legs. (We described a similar idea in Section 2 when we first introduced Feynman diagrams.) Although beta decay proceeds through the creation of a W-boson, the existence of this W-boson is fleeting. Heuristically, we sometimes say that the W-boson briefly borrows some energy from the vacuum, and this is allowed by the Heisenberg uncertainty principle as long as it is paid back in a suitably short period of time. A better explanation is to admit that all is fields, and the kind of random fluctuations needed to create the W-boson are part and parcel of quantum field theory.

Whatever words we choose to drape around this, the fact that the intermediate W-boson is much heavier than any of the incoming or outgoing particles has consequence: it reduces the probability for the decay to take place and this in turn means that the lifetime of particles that decay through the weak force can be much larger than other timescales in particle physics. In fact, the lifetime of particles depends both on the mass of the W-boson and the mass differences between the initial and final particles.

So, for example, the half-life of a neutron is around 10 minutes, a very human number, while the half-life of uranium is around 4.5 billion years. And, of course, most nuclei do not decay at all, with the strong force providing a safe haven that stabilises the neutron.

That Time, Before We Were Born, When We Nearly All Died

As something of an aside, there's a wonderful race-against-the-clock story from the early universe that someone should turn into a Hollywood movie.

A long long time ago – a few fractions of a second after the big bang to be precise – protons and neutrons lived together in happy equilibrium. Beta decay happened, but so too did inverse beta decay and the two reactions were in a delicate balance. However, this state of harmony could not last forever. At around 2 seconds after the big bang, an imbalance kicked in and the inverse beta decay no longer occurred.

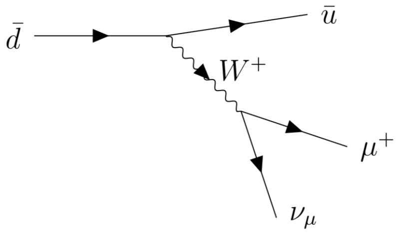
From then on, the neutrons were on their own. If any were to die, there was nothing to replace them. With a half-life of just 10 minutes, they needed to quickly find a sanctuary before they were all killed off by beta decay, a fate which would leave the later universe a very dull place to live.

The refuge was obvious: they should bind together with a proton to form a deuterium nucleus, or with a couple of protons to form a helium nucleus. Sadly the early universe back then was a hot and violent place, and every time the neutrons tried to bind together with a proton, the resulting nucleus was quickly smashed apart. For some time, the fate of neutrons – and all future life – hung in the balance. The neutrons needed to wait long enough for the universe to cool so that they could form nuclei, but time was not on their side. Eventually, around 6 minutes after the big bang, the temperature dropped sufficiently and the first stable deuterium nuclei formed, followed quickly by helium. The universe was saved, leaving a future that could be filled with atoms and stories.

The scary part about the above tale is that the 6 minutes needed for nuclei to form seems to have nothing to do with the 10 minutes needed for neutrons to decay. They come from entirely different pieces of physics. We should all feel very lucky to have survived this perilous time in history. You can learn more about the calculations underlying this in the lectures on Cosmology.

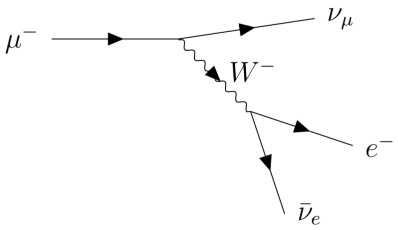
Pion Decay

Neutrons are not the only victim of the weak force. A world without the weak force would be awash with pions which, as we saw in Chapter 3, are the lightest of the particles formed from quarks. The vast majority of the time (something like 99.99%) charged pions $\pi^+ = u\bar{d}$ decay through the weak force to muons. This occurs through the Feynman diagram:



The resulting $\bar{u}$ anti-up quark then combines with the other up quark in the pion, and the two rapidly decay into photons. The lifetime of the charged pion is about 10^{-8} second.

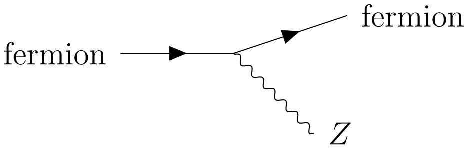
The resulting muons don't live too long either although, as we already saw in Interlude C, they hang around longer than the pions. Their demise is also due to the weak force and they decay to electrons and neutrinos through the process



The lifetime of the muon is around 2×10^{-6} seconds. All other particles involving quarks and leptons from the second and third generation have the same fate, decaying through the weak force to the more familiar particles from the first generation.

Z-Bosons

The Feynman diagrams involving Z-bosons are similar to those involving photons that we met in Section 2 in the sense that they don't change the type of fermion with which they interact:



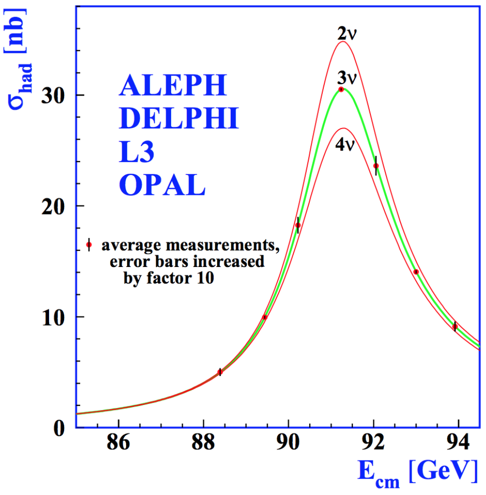


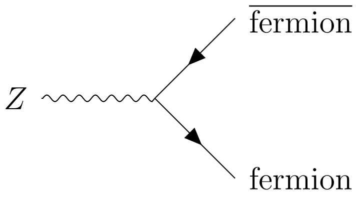
Figure 36. The Z-boson seen as a resonance in e^+e^- scattering. The three graphs show the shape of the graph that would be seen if there were 2, 3 or 4 species of neutrinos. The data points fall beautifully on the middle option. This plot was taken from a joint paper, combining many different experimental results.

The fermion on the two legs is the same, but now can be either u, d, ν_e or e^-. Similar Feynman diagrams involving higher generations of fermions also exist for both W and Z interactions.

The effects of the Z-boson are less immediately dramatic than those of the W. There are a number of processes that have been key for us in unravelling the structure of the weak force. In particular, Z-interactions allow neutrinos to scatter off any other lepton or quark, and this is one of the key ways in which these particles are detected.

Second, there is an important story involving the lifetime of the Z-boson. This gives our best current understanding to the question: how many generations of fermions are there? Of course, we've discovered three. But are there more to be found?

A very nice experiment involving the Z-boson strongly suggests that that we've found them all. This comes from looking at how the Z-boson decays. As we've seen, this can occur through a diagram of the form



where the fermion-anti-fermion pair can be either quarks or leptons. All that's needed is that the mass of the final two fermions is less than the 91 GeV mass of the Z.

About 20% of the time, the Z-boson decays to a pair of neutrinos. As we'll explain in more detail, the neutrinos are extremely light with a mass less than $\lesssim$ 1 eV. If there were more generations of fermions they would, as we've seen in Section 4.1.3, necessarily include further neutrinos. But, assuming that these additional neutrinos are not widly heavier than the first three, they would affect the lifetime of the Z-boson and so could be detected indirectly.

A careful experimental study of the lifetime of the Z-boson then gives a striking result. The number N_ν of neutrinos that the Z decays into is

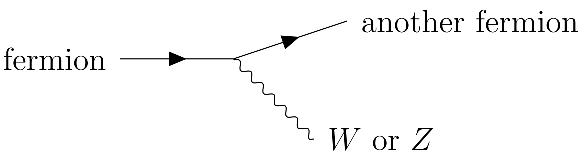
$$N_\nu = 2.994 \pm 0.008$$

If you're looking for an integer, this is the same thing as 3. If there are further generations to be found, then the neutrinos must be so heavy that they don't affect the lifetime of the Z-boson. In other words, any further neutrino must be something like 10^9 times heavier than the three that we know and love. Similar arguments come from the study of the Higgs decays.

The Fermi Constant

For both QED and QCD we made a big deal out of the coupling constants and the way they change as you look at different energy scales. But, so far, we've made no comment about the strength of the coupling for the weak force.

There is such dimensionless coupling which we write as α_W. It's analogous to the fine structure constant α in QED. Whenever you see a Feynman diagram like the one below, it contributes to a process with probability proportional to α_W:



Rather surprisingly, the value of α_W is not particularly small. Like all coupling constants, it changes with scale. At the scale of 100 GeV, it takes the value

$$\alpha_W \approx \frac{1}{30}$$

This is somewhat larger than $\alpha \approx 1/137$ of electromagnetism! So why is the weak force so weak?

The answer, like everything to do with the weak force, lies in the Higgs mechanism. The weak coupling constant runs under renormalisation group and, like the strong force, actually increases as we look at lower energies. However, the Higgs ruins all of this. At the scale of the Higgs condensate, the coupling freezes to roughly the value $\alpha_W \approx 1/30$. More importantly, the W and Z bosons get a mass at this scale and this greatly limits the range over which the weak force can operate to roughly distances $r \sim 1/M_W$. We saw this already in the effective Yukawa-type force (4.3) that arises from the exchange of massive W and Z bosons. Ultimately the reason that the weak force is so weak is because the distance over which it operates is so small, rather than the intrinsic weakness of the force itself.

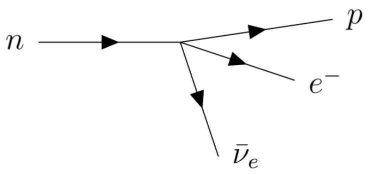
We can make this more quantitative. It's often more useful to characterise the strength of the weak force using a dimensionful coupling constant

$$G_F = \frac{\pi}{\sqrt{2}} \frac{\alpha_W}{M_W^2} \tag{4.4}$$

where the strange value of $\pi/\sqrt{2}$ is there for historical purposes. G_F is called the *Fermi constant* and takes the value

$$G_F \approx 1.166 \times 10^{-5} \text{ GeV}^{-2}$$

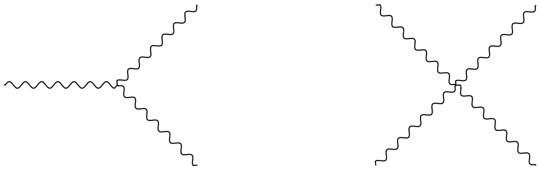
To understand the importance of this coupling, let's go back to our favourite beta decay. If we squint and ignore the W-boson, we return to Fermi's original theory where the process looks like a direct interaction between four fermions



The probability of such a decay is characterised by G_F which, as we see from the definition (4.4), incorporates both the original weak coupling constant α_W and the finite range of the W-boson due to its mass. More precisely, to get a probability we should construct a dimensionless object out of G_F so the probability of any low-energy event with $E^2 \ll G_F^{-1}$ is really $\sim G_F E^2$. Because G_F^{-1} is so big, any weak process that happens at low energy is suppressed. This is why the weak force is weak.

W and Z Self-Interactions

To complete our collection of Feynman diagrams, I'll finish by pointing out that there are interactions between the $W^\pm$ and Z bosons themselves. This follows because, just like for gluons, the underlying equations describing the weak force are the Yang-Mills equations, whose defining property is the presence of interaction terms between spin 1 particles. These Feynman diagrams take the form



where each leg can be any $W^\pm$ or Z boson, as long as charge conservation is obeyed at the vertex. These diagrams are mostly important in higher order corrections to the kinds of processes that I've described above.

4.3 Flavours of Fermions

In this section, we will describe how the quarks and leptons interact with the Higgs field. The six quarks are often referred to as *flavours*. As we will see, the structure of flavour is, in many ways, the most fiddly and poorly understood part of the Standard Model. It is certainly where the vast majority of the parameters in the Standard Model reside and, most likely, the best place to look for clues about what lies beyond.

4.3.1 Yukawa Interactions

The Higgs field talks to quarks and leptons through *Yukawa interactions*.

There is room for confusion in this name. The original interactions postulated by Yukawa were designed to explain how neutrons and protons bind together inside a nucleus, with a scalar field providing the mediating force. We now known that the scalar particle Yukawa had in mind is the pion and, as explained in Section 3, is composed of two quarks.

However, Yukawa's basic idea – that a scalar field can interact with two fermions – is reprised in the Standard Model with the Higgs field arising as the fundamental scalar, and the name "Yukawa interactions" has been co-opted in this new context. More precisely, the Yukawa interactions in the Standard Model couple the Higgs field to one left-handed fermion and one right-handed fermion.

Like all forces, there are dimensionless numbers that tell us the strength of the interaction between the Higgs field and the fermions. These dimensionless numbers are called *Yukawa couplings* and we will denote them by λ. The Higgs field then gives a mass to each fermion that is directly proportional to the strength of the Yukawa coupling,

$$\text{mass} = \frac{\lambda}{\sqrt{2}} \times 246 \text{ GeV} \tag{4.5}$$

where 246 GeV is the value of the Higgs expectation value (4.2) that we met earlier. The factor of $1/\sqrt{2}$ in this formula is just convention.

Each fermion has a different Yukawa coupling, and hence a different mass. The Yukawa couplings for the various quarks are:

$$\begin{array}{rccl}
\text{top :} & \lambda \approx 1 & \Rightarrow & m_t \approx 173 \text{ GeV} \\
\text{bottom :} & \lambda \approx 2.5 \times 10^{-2} & \Rightarrow & m_b \approx 4.2 \text{ GeV} \\
\text{charm :} & \lambda \approx 7.5 \times 10^{-3} & \Rightarrow & m_c \approx 1.3 \text{ GeV} \\
\text{strange :} & \lambda \approx 5.5 \times 10^{-4} & \Rightarrow & m_s \approx 96 \text{ MeV} \\
\text{up :} & \lambda \approx 1.3 \times 10^{-5} & \Rightarrow & m_u \approx 2.2 \text{ MeV} \\
\text{down :} & \lambda \approx 2.7 \times 10^{-5} & \Rightarrow & m_d \approx 4.7 \text{ MeV}
\end{array}$$

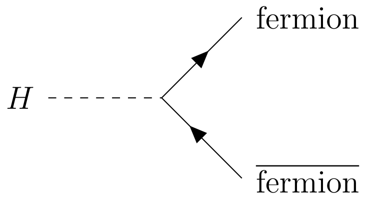
Meanwhile, the Yukawa couplings for the electron and its cousins are

$$\begin{array}{rccl}
\text{tau :} & \lambda \approx 1 \times 10^{-2} & \Rightarrow & m_\tau \approx 1.8 \text{ GeV} \\
\text{muon :} & \lambda \approx 6.1 \times 10^{-4} & \Rightarrow & m_\mu \approx 106 \text{ MeV} \\
\text{electron :} & \lambda \approx 2.9 \times 10^{-6} & \Rightarrow & m_e \approx 0.5 \text{ MeV}
\end{array}$$

Although we've reduced the masses of the various quarks and leptons to dimensionless coupling constants λ, we currently have no understanding of why the Yukawa couplings take these values. The Yukawa couplings span 5 orders of magnitude in order to explain the quark masses, and a further order of magnitude to explain the electron mass. We don't know why. Is it coincidence that the top Yukawa coupling is almost exactly one? Again, we simply don't know. All of these are, from the perspective of the Standard Model, fundamental constants and we have no deeper understanding of them.

You may have noticed that we haven't yet discussed neutrinos. These have masses at least six orders of magnitude smaller than the electron, or 10^{-12} times smaller than the top quark. Thankfully there is a plausible reason for this vast discrepancy in mass, but we postpone the discussion until Section 4.4.

The fact that the masses of all fermions derive from the Higgs field has a consequence that could only be tested after the Higgs boson was discovered. The Higgs boson couples to all fermions with the Feynman diagram interaction



The strength of this interaction is dictated by the same Yukawa coupling λ that determines the mass of the fermion. This means that once the Higgs boson is produced, we have an entirely different way of measuring the Yukawa couplings by determining the relative probability that the Higgs boson decays to various fermion-anti-fermion pairs. So far, decays to the top and bottom quarks and the tau have been measured, all in agreement with theoretical expectations.

4.3.2 Symmetry Breaking

Symmetry is one of the key concepts in particle physics. For example, the $SU(3)$, $SU(2)$ and $U(1)$ labels that we've been using to describe the three forces are really mathematical expressions of symmetry. Until now, we've somewhat underplayed the idea of symmetry in these lectures, largely because it's a fairly formal mathematical idea and analogies tend to get bogged down in useless diversions. However, in describing how fermions get a mass we have a chance to elaborate on this. At the same time, we'll also get a better understanding of how electromagnetism emerges from the electroweak force.

For an example of symmetry, we can first look to the strong force. As we've seen, each quark comes with a colour: red, green or blue. But there is a symmetry underlying the choice of colours. To see this, you take a collection of particles and swap all the colours around,

$$\text{red} \longrightarrow \text{green} \longrightarrow \text{blue} \longrightarrow \text{red}$$

The end result will then look identical to the set-up we started with. This happens for the proton because it has a collection of quarks with one colour of each and that doesn't change under a permutation of colours. Meanwhile, as we described in Section 3.2, a meson has a quantum superposition of colours $\bar{r}r + \bar{g}g + \bar{b}b$, so that too remains unchanged.

What's the analogous statement for the weak force? We've seen that "weak colour" is the up/down and electron/neutrino label for left-handed particles. This means that, at the fundamental level there is a symmetry that swaps

$$\text{up} \longleftrightarrow \text{down}$$

and

$$\text{electron} \longleftrightarrow \text{neutrino}$$

for left-handed particles. Under such an exchange, all physics should remain the same.

Now, although there is such a symmetry at the fundamental level, it certainly doesn't manifest itself in our world. If you were to exchange all the electrons in your body for neutrinos, bad things would happen. So what's going on? What happened to this fundamental symmetry of nature?

The answer is that the symmetry is broken by the Higgs boson. To understand this, we first note that the Higgs experiences the $SU(2)$ weak force. This means that the Higgs too must have a "weak colour" label, something that we've neglected to mention so far. The Higgs field is actually a pair of complex-valued Higgs fields

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

where the two fields ϕ_1 and ϕ_2 are mixed by the weak force.

Because there are actually two Higgs fields, the potential $V(\phi)$ should really be plotted in 3d. We have drawn this in Figure 37. Note that the shape of the potential doesn't change if you rotate it around the vertical axis, and this reflects the symmetry of the weak force. There's no sense in which the potential prefers the ϕ_1 direction to the ϕ_2 direction: both are on the same footing. If you slice this potential along any direction, then you get the 1d graph that we previously drew on the right-hand side of Figure 34.

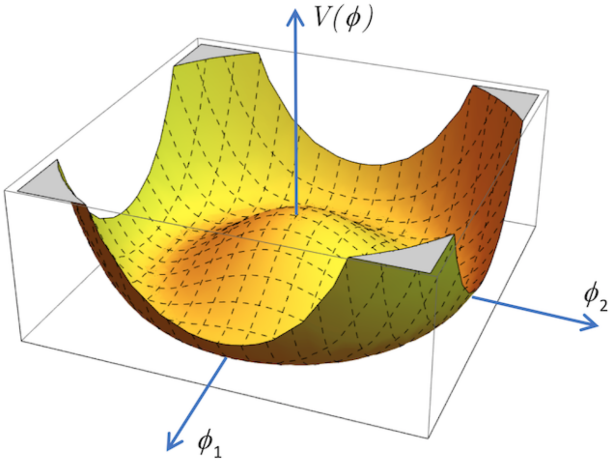


Figure 37. A better depiction of the Higgs potential. This is sometimes called the "mexican hat" potential for its sombrero-like quality.

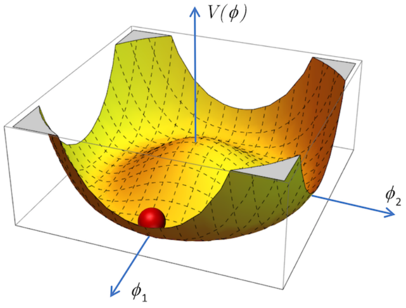


Figure 38. The choice of Higgs expectation value breaks the symmetry

This new 3d potential doesn't have isolated minima. Instead, there is a ring of minima, all lying at some fixed distance $\sqrt{\phi_1^2+\phi_2^2}$ from the origin. In the vacuum, the Higgs field must sit somewhere in this valley. But where? Nothing tells the Higgs field where to sit and, moreover, because of the symmetry, any choice is as good as any other. Nonetheless, like a ball on a roulette wheel, the Higgs must choose somewhere to sit. And choose it does. Because it doesn't matter where the Higgs field sits, we may as well decide that it lies along the ϕ_1 axis, although any other position would be just as good. This is shown in the figure to the right.

Once the Higgs field has nestled in place, it's no longer true that the ϕ_1 direction is the same as the ϕ_2 direction, because the Higgs field sits in one of these directions and not the other. We say that the symmetry has been broken. More precisely, we say that the symmetry has been *spontaneously broken*.

Ultimately, this is the reason why electrons and neutrinos (and up and down quarks) behave so very differently in our world. The laws of physics endow the left-handed versions of the particles with identical properties. But then the Higgs field comes along and spoils it, choosing to sit in one place and rather than another. The ϕ_1 direction in which the Higgs field sits coincides with the "electron direction" for leptons, and the "down quark" direction for quarks. The ϕ_2 direction is where the neutrino and up quark live.

Finally, we can complete the story of how electromagnetism emerges from the electroweak force. The photon is the combination of the weak and hypercharge forces that does not fall into the clutches of the Higgs boson. The right-handed particles know nothing about the weak force, so for them the coupling to the photon (which we call electric charge) is identical to the coupling to hypercharge. Indeed, if you look at Table 2 you'll see that the hypercharge for the right-handed particles is the same as their electric charge.

However, the left-handed particles feel both the weak and hypercharge and the resulting electric charge is a combination of both. How that combination plays out depends on whether the particle is aligned with the Higgs field, or orthogonal to the Higgs field. For those left-handed particles that are aligned with the Higgs field, the resulting electric charge is

$$\text{electric charge} = \text{hypercharge} - \frac{1}{2}$$

while for those that lie orthogonal to the Higgs field, the electric charge is

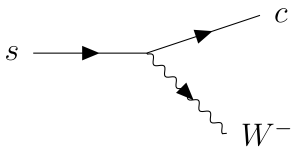
$$\text{electric charge} = \text{hypercharge} + \frac{1}{2}$$

We can now check what this means for the left-handed quarks and leptons by looking back at Table 2. After symmetry breaking, the left-handed quark splits into two particles with electric charges $\frac{1}{6} \mp \frac{1}{2} = -\frac{1}{3}$ and $+\frac{2}{3}$. These, of course, are the charges of the down and up quark respectively. Meanwhile, the left-handed leptons also split into two, now with charges $-\frac{1}{2} \mp \frac{1}{2} = -1$ and 0. These are the electric charges of the electron and neutrino.

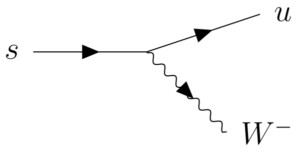
4.3.3 Quark Mixing

We've now described all the forces of the Standard Model. However, within the structures that we've outlined above, there are certain particle interactions that we cannot explain. We can't, for example, explain the decay of hadrons that contain quarks from the higher generations.

To see the problem, consider the kaon K^-, whose quark content is $\bar{u}s$. We stated in Section 3.4 that kaons decay with a lifetime of around 10^{-8} seconds. But where do they go? Within the strong and electromagnetic forces, there's no way for quarks to change their type, so it must be a weak interaction that does the job. But the weak interaction that we've described above doesn't allow quarks to change generation. For a strange quark, we have the Feynman diagram



But then we're left with a charmed quark and there's nowhere for that to go. Instead, for a kaon to decay we would need an interaction that mixes the generations, like this:



If such a decay was possible then the resulting up quark could annihilate with the $\bar{u}$ in the kaon, while the W^- can decay into, say, an electron and anti-neutrino in the usual way.

So what are we missing? Is it possible for the weak force to mediate decays that change quarks from one generation into another? The answer to the second question is: yes. The answer to the first question is that we're missing something rather subtle about the meaning of the word "particle"!

So far, we have been using two, somewhat different meanings of the word "particle" and tacitly assuming that they coincide. These are:

- A particle is an excitation of the field that has a fixed energy. Or, because $E = mc^2$, an equivalent way of saying this is that we can assign a specific mass to the particle. In the language of quantum mechanics, we say that it is an *energy eigenstate.*
- A particle is the object that interacts with a particular force. This is really pertinent only for the weak force which, as we've seen, turns one particle into a different particle: say the down quark into an up quark.

The subtlety comes about because, for the weak force, these two ideas of what it means to be a particle don't quite agree. The excitations of the field with a fixed energy aren't the same thing as the excitations of the field that have a specific interaction with the weak force. Another way of saying this is to recall that the mass of the particle comes from the Higgs field. So what's really going on here is a mismatch between the way the Higgs interacts with fermions and the way the W-bosons interact with fermions. The two interactions are not quite aligned.

Let's keep our original names for quarks – u, d, s, etc – for the particles that have a definite mass. We'll then denote the particles that experience the weak force with primes – u', d', s' and so on. It turns out that we can always simply define the up-sector quarks to be aligned,

$$u = u' \quad \text{and} \quad c = c' \quad \text{and} \quad t = t' \tag{4.6}$$

But the down-sector quarks are then misaligned. It's simplest to explain what's going on if we first ignore the bottom quark. The misalignment is then given by

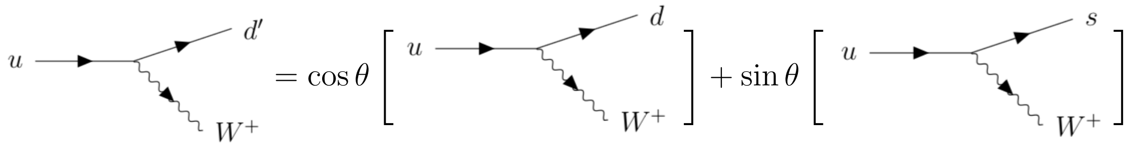
$$\begin{aligned} d' &= d\,\cos\theta + s\,\sin\theta \\ s' &= s\cos\theta - d\,\sin\theta \end{aligned} \tag{4.7}$$

Here θ is known as the *Cabbibo angle*. It is a fundamental parameter of Nature, an example of what's called a mixing angle. We'll see many more of these shortly. The Cabbibo angle is measured experimentally to be

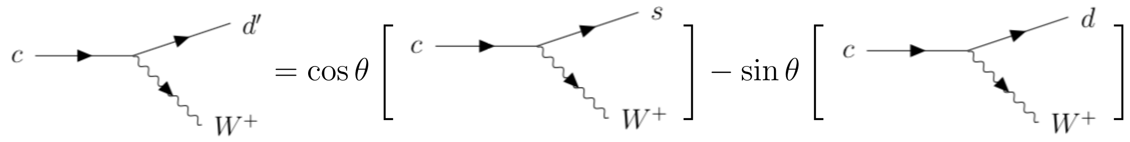
$$\sin\theta \approx 0.22 \quad \Rightarrow \quad \theta_c \approx \frac{\pi}{14} \approx 13^\circ$$

Why this number and no other? We don't know! We don't currently have any deeper explanation for this.

The formulae (4.7) might remind you of the equations for a rotation, and that's exactly the right way to think about it. The (d', s') quarks that feel the weak force are rotated relative to the (d, s) quarks that interact with the Higgs. This phenomenon is called *quark mixing*. It means that the Feynman diagrams that we previously wrote down for the weak force should be amended. The correct Feynman diagram involving the up quark is

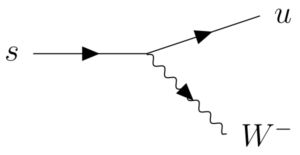


and similarly for the charm quark



How should we think about this? It seems to say that if, for example, a charm quark decays by emitting a W-boson, then the end product is both a strange quark and a down quark, in some combination. But, of course, we're in a quantum world here. And just as a particle can be in two places at the same time, or a cat both dead and alive, the decay product of a charm is indeed a quantum superposition of a down quark and a strange. As usual, this manifests itself in our experiments as probability. We get probabilities by taking the square of a Feynman diagram: so the probability of $c \to s+W^+$ is proportional to $\cos^2\theta$ while the probability of $c \to d+W^+$ is proportional to $\sin^2\theta$.

The phenomenon of quark mixing resolves our earlier puzzle: it's now quite possible for a meson like the kaon to decay, because there is an escape route for the strange quark, with Feynman diagrams like this now allowed:



The only price we pay is that the probability for such events to happen is reduced by $\sin^2\theta \approx 0.05$. This results in an increased lifetime for mesons containing strange quarks.

This story repeats with the addition of an extra generation. Now there is mixing between the down, strange and bottom quarks, and the simple rotation (4.7) is replaced by a more complicated matrix equation

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{4.8}$$

The 3×3 matrix is known as the Cabbibo-Kobayashi-Maskawa (or CKM) matrix. The upper-left 2×2 sub-matrix agrees with the Cabbibo mixing (4.7), but now we have the possibility of mixing between all three quarks.

You might reasonably ask: what made the up-sector special? Why is the up-sector aligned, as in (4.6), while the down-sector has the complicated CKM matrix? The answer is that this is just a choice. There's some freedom in the equations to guarantee a partial alignment between the weak and Higgs forces. The convention is to pick the up-sector to be aligned because then the misalignment looks simple for the relatively light strange mesons, with no need to invoke the charm quark in the argument. But if you were feeling a little perverse, there's nothing to stop you redefining everything with the down-sector aligned and the up-sector askew, or even some combination of the two.

The components of the CKM matrix have been accurately measured experimentally. It turns out that some of the elements can be complex numbers and we'll explain the significance of this in Section 4.3.4. For now, we give just the absolute values of each element which are roughly

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.97 & 0.22 & 0.004 \\ 0.22 & 0.97 & 0.04 \\ 0.009 & 0.04 & 0.999 \end{pmatrix}$$

You can see the Cabbibo angle sitting there in $V_{us} = \sin\theta \approx 0.22$. The full CKM matrix extends the Cabbibo angle to a 3×3 matrix. Not all the elements of this matrix are independent. (The matrix can be shown to be "unitary".) It turns out that there are 4 independent parameters in the CKM matrix.

Just like we have no understanding of why the Cabbibo angle takes its particular value, nor do we have any good understanding of the CKM matrix. As you can see, it's not far from a diagonal matrix, with the Cabbibo terms being the only ones that aren't tiny. We don't know why.

It's worth pausing to take in a bigger perspective here. In the first part of this chapter, we described how the matter content of the Standard Model interacts with the different forces. There we found a beautiful consistent picture – a perfect jigsaw – in which the interactions were largely forced upon us by the consistency requirements of the theory. For a theoretical physicist, it is really the dream scenario. This, however, contrasts starkly with the story of flavour. Even focussing solely on the quarks, we find that there are 6 Yukawa couplings that determine their mass, plus a further 4 entries of the CKM matrix that determine their mixing. And none of these parameters are fixed, or understood at a deeper level.

Somewhat ironically, much of this complexity can be traced to the simplicity of the Higgs. The strong and electroweak forces are described by Yang-Mills type theories, and these come with mathematical subtleties that are ultimately responsible for the quantum consistency conditions that constrain their interactions. But there are no such subtle constraints for the Higgs boson. It is a simple, spin 0 particle, that can do as it pleases and the result is the plethora of extra parameters that we've seen.

Turning this on its head, there is a possibility that the flavour sector of the Standard Model may well offer a unique opportunity. The structure of quark masses, together with the CKM matrix, surely contains clues for what lies beyond the Standard Model. Why the hierarchy of masses? Why these values of the CKM matrix? Moreover, there is a peculiar pattern: with $V_{us} = \sin\theta$, the other remaining values of the CKM matrix are not far from $V_{cb} \approx \sin^2\theta$ and $V_{ub} \approx \sin^3\theta$. Is this coincidence, or telling us something deep and important? We don't know. But hopefully, one day, we will find out.

4.3.4 CP Violation and Time Reversal

There is one last surprise waiting for us in the CKM matrix. We started this chapter by describing how the symmetry of parity is not respected in the microscopic world: the laws of physics look different when reflected in the mirror. As we saw, this was a pivotal moment in the history of particle physics, giving an important clue about the chiral structure of the Standard Model.

The violation of parity goes hand in hand with the violation of another symmetry, know as *charge conjugation*. We could ask: are the laws of physics the same for particles and anti-particles? And the answer is no. Anti-particles do not feel the same forces as particles. Moreover, this follows immediately from parity violation: the weak interaction involves left-handed neutrinos, or right-handed anti-neutrinos. So if you just change a particle to an anti-particle – say a left-handed neutrino to a left-handed anti-neutrino – then it doesn't experience the weak force anymore.

However, it remains a logical, and indeed compelling, possibility that the laws of physics are invariant under the simultaneous action of parity and charge conjugation. This symmetry is called CP, the C for "charge conjugation" and the P for "parity". A universe that exhibited CP symmetry would have the property that the laws of physics look the same if we swap all particles with anti-particles *and* look at them reflected in the mirror.

So are the laws of physics actually invariant under CP? You may not be surprised to hear that the answer is no.

CP violation was discovered in 1964. We describe the experimental evidence in Interlude D.2, focussing here on the theory. While violation of parity has an enormous effect on the theoretical underpinnings of the Standard Model, the violation of CP appears to be almost an afterthought: it turns out that the laws of physics violate CP if the CKM matrix contains a complex number! It does, and so CP is violated.

There is a way to quantify how much CP is violated in the Standard Model. From the CKM matrix, one can define a single number called the *Jarlskog invariant*, J. The maximum theoretical value it can take is $J \leq 1/2\sqrt{3} \approx 0.3$. Its experimentally measured value is $J \approx 3 \times 10^{-5}$. We see that the CP violation, at least as manifested among quarks, is really very small.

The experimental consequences of CP violation among quarks are, to put it mildly, rather subtle. It shows up in the decay of certain neutral mesons involving higher generations (like the kaon $K^0 = d\bar{s}$). We'll explain this in more detail in Section D.2.

Why should we care about CP violation? Well, there are two reasons. The first is that our universe is, rather fortunately, full of matter but with very little anti-matter. It's thought that this imbalance occurred naturally in the early universe, but for this to happen there have to be processes where matter and anti-matter behave differently. This, it turns out, requires CP violation. So although small, it may well have had extraordinarily large consequences. We'll discuss this further in Section 5.3.3.

There's also a twist to this tale. It turns out that a world with just two generations of fermions has no room for CP violation in the CKM matrix! In that case, the only quark mixing comes from the Cabbibo angle (4.7) and there's no complex number in sight. So, while it may seem like the bottom and top quark aren't much good for anything today, without them we may well find ourselves in a universe with as much matter as anti-matter.

Time Reversal

The second reason that CP symmetry is important is because it's closely related to yet another symmetry, that of *time reversal*.

A theory is said to obey the symmetry of time reversal if there's no way to distinguish between the laws of physics running forwards in time and running backwards. Of course, in our macroscopic everyday world, the arrow of time is obvious. Underlying this is the concept of *entropy* which characterises the tendency for things to become disordered and muddled for the simple reason that there's more ways to be messy than to be neat. Ultimately, the entropic arrow of time can be traced to initial conditions. For a reason that we don't fully understand, the universe started off neat and tidy, so that it could ultimately descend into disorder, creating interesting structures like galaxies, planets and life along the way.

This entropic arrow of time is a powerful idea, largely because it doesn't depend on what's happening on the microscopic scale. In particular, it cares little for whether the laws of physics at the fundamental level are time reversal invariant. Nonetheless, it's interesting to ask whether time reversal is a symmetry of the world. If we were able to take a movie of interactions of fundamental particles, then play it backwards, would we notice the difference?

It turns out that the answer to this question is very closely tied to CP violation. The three symmetries of parity, charge conjugation and time reversal are an intwined triumvirate. There is a theorem that states that all laws of physics must be invariant under the combination of all three symmetries, what is known as CPT, with T standing for "time reversal". This means that if you make a movie of some dance performed by fundamental particles, then it's guaranteed to look the same if you watch it played backwards, in a mirror, with all particles exchanged for anti-particles.

The CPT theorem tells us that if CP is violated, then so too is T, for only then can the combination CPT be saved. The upshot is that the experimental discovery of CP violation is, admittedly indirectly, telling us that the laws of physics must look different running forwards and backwards in time. In other words, even at the microscopic level, there is an arrow of time in the universe.

Now, that sounds like a very big deal. Nonetheless, it's striking how little impact this fact has, not just on our daily lives, but on our deeper understanding of physics. As we mentioned above, it is likely that CP violation or, equivalently, time reversal violation is responsible for the preponderance of matter over anti-matter in the universe and that's not something to sneeze at. But that happened a long time ago! Has the violation of time reversal really not had any significant consequences since?! Moreover, it's not even straightforward to give a clear description of the microscopic process that runs differently forwards and backwards in time. The experimental observation of CP violation is described in Section D.2, and involve some rather subtle aspects of meson decay. We'll do slightly better in Section 4.4 when we discuss possible CP violating effects in neutrinos where an interpretation in terms of time reversal is easier to come by.

However, on the theoretical level it's interesting to compare the effects of parity violation with the effects of time reversal violation. At first glance, they seem very similar: one is a flip of spatial coordinates, $\mathbf{x} \to -\mathbf{x}$, the other a flip of time $t \to -t$. Yet the discovery of parity violation had profound consequences, leading immediately to the need for chiral matter, the associated delicate consistency conditions between left- and right-handed fermions that, ultimately, gives us much of the structure of the Standard Model. This stands in sharp contrast to the theoretical consequences of time reversal violation which imply only that a single parameter, buried within the CKM matrix, is a complex rather than a real. It makes you wonder if there's something that we're missing!

4.3.5 Conservation Laws

Some things never change. In physics, we call these conservation laws. They comprise some of the most useful and powerful laws of nature. Among the conservation laws that we learn early in our physics education are the conservation of energy, momentum, angular momentum, and electric charge.

The familiar conservation laws listed above are all exact. No known process violates these laws. Moreover, they are sewn in the mathematical fibre of quantum field theory at such a deep level that it seems likely that they are here to stay.

In addition to these exact conservation laws, we have a number of "almost conservation" laws, together with a couple of "probably almost conservation" laws. These are the subject of this section.

First, the "almost conservation" laws. These are things that are conserved if you ignore one kind of force or another, but ultimately do not hold when you consider the whole of the Standard Model. For example, the electromagnetic and strong interactions don't change the type of quark. This means that the number of up quarks, down quarks and strange quarks are each individually conserved by any electromagnetic or strong process. Of course, as soon as the weak interaction kicks in, the down quark can decay into an up quark as we have seen repeatedly in this section. But because the weak force is, as the name suggests, weak, these decays can take a long time. This means that there will be situations where we can ignore weak decays and view the number of up and down quarks – or, equivalently, the number of protons and neutrons – as an effectively conserved quantity. This almost conservation law is sometimes called *isospin* and was mentioned briefly in Section 3.

As we saw in the last section, a strange quark can decay only through the weak force and, even then, only through the process of quark mixing. This causes yet further suppression in decays of mesons containing strange quarks, and this is seen in the relatively long lifetimes of the kaons where now "relatively long" means around 10^{-8} seconds. Again, we can think of this as an "almost conservation" of a quantity called *strangeness*. Combining isospin with strangeness gives the eightfold way that we described in Section 3.

There is a similar story in the lepton sector. If we ignore the weak force, the number of electrons, muons, taus and their associated neutrinos are all unchanging. When we include the weak force, the muon and tau can both decay. Moreover, as we will see in the next section, the neutrinos also undergo a mixing process, analogous to the one we've seen for quarks, so the individual neutrino species are not, ultimately, conserved quantities either.

Nonetheless, when the dust settles there do seem to be two exact conservation laws that follow from the Standard Model as described so far.

- **Conservation of quark number**

  If you start with one kind of quark, it can decay into a different kind, usually ending up as the up quark since this is the lightest. But, if we count quarks as $+1$ and anti-quarks as -1, then the total number of quarks can't change.

  Confinement means that we don't see individual quarks, but rather the protons, neutrons and mesons that they bind into. The mesons contain a quark and anti-quark, and so there's no such thing as meson conservation. But the conservation of quark number means that the number of baryons (again, counting anti-baryons as -1) can't change. For this reason, "conservation of quark number" is usually referred to as *baryon conservation.*

  We see a good example of this in beta decay. The neutron decays, but leaves behind a proton. The total number of baryons before and after is the same.

- **Conservation of lepton number**

  Just as the total number of quarks can't change, neither can the total number of leptons. Again, we see this in beta decay. The neutron decays into a proton, electron and anti-neutrino. It creates an electron, which changes the lepton number by $+1$, but this is accompanied by an anti-neutrino which, because of the "anti-" contributes -1 to lepton number. The overall lepton number remains unchanged.

So the laws of physics seem to give two conserved quantities, baryon conservation and lepton conservation. And it's a true statement that we've never observed any process in which either these conservation laws are violated. Nonetheless, there are good reasons to think that they are not true conservation laws of nature. For this reason, I'll call them "probably almost conservation laws".

There are two reasons to think that baryon number and lepton number are not exactly conserved. While theoretically sound, neither of these is going to be high on an experimenters list of things to do. However, as I explain below, there are at least two experiments that might, with some level of optimism, tell us more in the future.

Reason 1: Electroweak Instantons

The first reason that baryon and lepton number are not likely to be precisely conserved quantities is, in many ways, the most subtle since it hinges on some of the deeper ideas relating geometry and quantum field theory. When we first met Feynman diagrams in Section 2.2, we explained that they were a way to give approximate answers to questions in quantum field theory. But there are some effects in quantum field theory that Feynman diagrams miss completely! These effects are, in some sense, smaller than any given Feynman diagram. (They are also closely related to the ideas of quantum anomalies that we discussed in Section 4.1.3.) And, in the Standard Model, there is such an effect that can turn a baryon into a lepton.

This process is due to an object known as an *electroweak instanton*. It turns out that it can't turn a proton into a positron: the proton is absolutely stable in the Standard Model. However, it can turn a collection of three baryons into three leptons. (The factor of three is actually related to the existence of three generations!) This means, for example, that the ^{3}He nucleus is unstable to decay into three leptons, say a couple of positrons and an anti-neutrino.

Now, we haven't ever observed such a decay. And for good reason. If you compute how long it would take for a helium nucleus to decay by this process, you get a silly number: something like 10^{173} years. Our universe has lasted around 10^{10} years. Clearly, it's unrealistic to think that we would ever observe this process. Nonetheless, strictly speaking in the Standard Model the baryon and lepton numbers are not individually conserved. Instead, there is only a single conserved quantity

$$\text{Conserved quantity} = \text{(number of baryons) - (number of leptons)}$$

This quantity is usually called simply $B - L$.

Reason 2: Black Holes

Black holes aren't black. Hawking taught us long ago that they slowly emit radiation due to quantum effects. While there is much that we don't understand about quantum gravity, the existence of Hawking radiation stands out as one of the few robust and trustworthy calculations that we can do. The prediction of this radiation follows from the known laws of physics, and doesn't rely on any speculative ideas about what lies beyond.

If we wait long enough (and, again, we're talking ridiculously long times here), any black hole will eventually evaporate and disappear. So we can ask: what became of the stuff that we threw in?

First, the black hole can't destroy electric charge. If you throw, say, an electron into a black hole then the black hole itself now carries the electric charge. Moreover, this is visible outside of the event horizon because the black hole emits an electric field. That electric field can't just disappear. So, as the black hole evaporates, it must eventually spit out a charged particle – maybe an electron, maybe an anti-proton – which carries the electric charge. The process of black hole evaporation must respect conservation of electric charge. Similarly, black hole evaporation respects the conservation laws of energy, momentum and angular momentum.

In contrast, there is nothing to prevent black holes from destroying baryons and leptons. When a black hole forms from the collapse of a star, it will typically contain around 10^{57} protons, and roughly the same number of electrons. But there's nothing like an electric field outside the black hole that tells you how many baryons and leptons are sitting inside. Furthermore, as the black hole evaporates there's no reason that it should spit out these particles intact. In fact, the vast majority of the mass of a black hole will be emitted in gravitational and electromagnetic radiation rather than baryons or leptons. In this way, we expect black hole evaporation to respect neither baryon number nor lepton number conservation.

Before we go on, I should stress an important point. There is an interesting and long standing problem about whether the *information* of what's thrown into a black hole is lost. We think that the answer is no. But, importantly, this isn't in contradiction with the violation of conservation laws!

To see why this is, consider the analogy of burning a book. In principle, the information written on the pages isn't lost: it's encoded in some impossibly complicated way in the correlations of the light and smoke that are emitted, and in the cinders that remain. Although the information is retained, the individual letters in the book are clearly lost.

In this analogy, the baryons in a black hole are like the letters. If you're clever enough and persistent enough, you may be able to detect that baryons were once present in the subtle correlations of the photons emitted by the black hole. But that doesn't change the fact that the baryons themselves have, almost certainly, gone for good.

Possible Experiments

No experimenter is lining up to test either of the theoretical arguments above. Nonetheless, the reasoning relies only on known, established laws of physics and tells us that, unlike electric charge, there is no fundamental reason for the conservation of lepton or baryon number. One might then wonder whether the violation of these conservation laws can be seen in some less extreme circumstance, one that could be tested here on Earth. There are two classes of ongoing experiments designed to test this.

Proton Decay

As mentioned above, within the Standard Model, the proton is absolutely stable. Nonetheless, it may well be that there is some physics beyond the Standard Model that causes protons to decay. Obviously, if we ever observed such a thing it would give us an important clue for what's happening on the next level.

So far, all we have are lower bounds. From our failure to detect proton decay in experiments we can infer that the lifetime of the proton is greater than around 10^{34} years. This is already quite an impressive statement, since its significantly longer than the age of the universe which is about 10^{10} years! But every litre of water contains about 10^{25} hydrogen atoms, so if you take 100 million litres of water, stare at it for a year, and fail to see a proton decay then you start to get close to the bound. There are a number of experiments around the world doing exactly this. The best current bounds come from the super-Kamiokande water Cerenkov detector in Japan. (We'll learn more about this detector in Section D.4 when we discuss experiments on neutrino oscillations.)

Neutrinoless Double Beta Decay

There is a rare, but well understood phenomena in which two neutrons in a nucleus simultaneously decay to two protons. This is called *double beta decay.*

For example, ^{48}Ca is a rare isotope of calcium, making up less than 0.2% of the naturally occurring element. It consists of 20 proton and 28 neutrons. A standard beta-decay process would take ^{48}Ca to ^{48}Sc, but this isotope of scandium has a smaller binding energy than calcium and so the beta decay process doesn't happen. Instead, ^{48}Ca decays through the much rarer double beta decay process

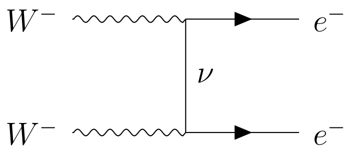
$$^{48}\text{Ca} \longrightarrow {}^{48}\text{Ti} + 2e^- + 2\bar{\nu}_e$$

The half-life is long: about 10^{20} years.

The double beta decay was first observed in the 1980s. As you can see, it violates neither baryon nor lepton number. However, it does raise the possibility for an even rarer occurrence: double beta decay without neutrinos. This would be a decay process of the form

$$^{48}\text{Ca} \longrightarrow {}^{48}\text{Ti} + 2e^-$$

With no neutrinos in the final state, lepton number is violated in this process. From the perspective of Feynman diagrams, the two neutrons would decay to protons by emitting two W^- bosons. These subsequently behave as follows:



Note that there's no arrow on the intermediate neutrino. Moreover, it's not possible to draw an arrow that lines up with both the external electron legs. Such a diagram is only valid if there's no way to distinguish between a neutrino and an anti-neutrino; indeed, this is what the observation of neutrinoless double beta decay would mean. We'll see how this might come about in the next section.

Despite many ongoing experiments around the world, neutrinoless double beta decay has yet to be observed. If it were seen, it would be a very big deal. Indeed, as we explain in the next section, it would give a key insight into the nature of neutrinos.

4.4 Neutrinos

No one would accuse a neutrino of being gregarious. They interact less than a first year undergraduate mathematics student forced to sit next to their theoretical physics professor at a matriculation dinner (to give a weirdly specific yet shudderingly memorable analogy).

For example, in the time it takes you to read this sentence, around 100 trillion neutrinos will have passed through your body. Most of them came from the Sun, but a significant minority have a cosmic origin, and have been streaming through the universe, uninterrupted since the first few seconds after the Big Bang. Moreover, in contrast to photons, the number of neutrinos hitting you doesn't change appreciably as day turns into night. The neutrinos from the Sun will happily pass right through the Earth and out the other side. This is vividly demonstrated in the picture of the Sun at night shown in Figure 39.

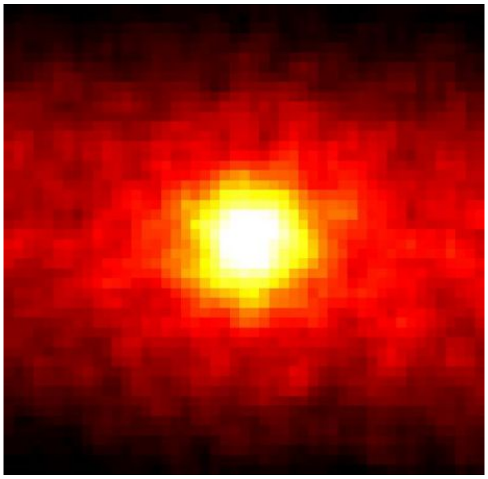


Figure 39. The Sun at night. This is a picture, taken by Super-Kamiokande, shows the neutrino flux coming from the Sun. The utterly remarkable fact is that the picture was taken at night, with the neutrinos passing through the Earth before hitting the detector.

There are two reasons why neutrinos are so intangible. The first is that they are the only particle to interact solely through the weak force. And, as we've seen, the weak force is weak. The second reason is that their mass is much much smaller than any other fermion which means that on the rare occasion that they do interact, they don't deliver much of a punch. The purpose of this section is to describe some properties of neutrinos in more detail.

4.4.1 Neutrino Masses

There is much that we don't know about neutrino masses. But we do know that the masses are not zero.

At the moment, we have no direct measurement of the mass of each neutrino. But we do have some precious information. First, we know that one neutrino must have a mass greater than

$$m_\nu \gtrsim 0.05 \text{ eV}$$

We'll explain *how* we know that neutrinos have a mass this big in Section 4.4.2.

Second, constraints from cosmology give us an upper bound on the sum of all neutrino masses. This comes from the imprint that neutrinos in the early universe leave on the cosmic microwave background radiation. (We'll say more about the intersection of

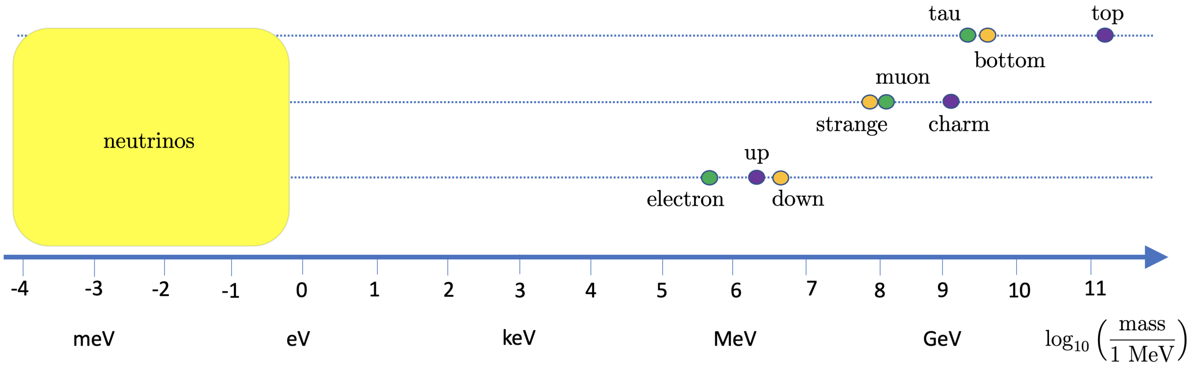


Figure 40. Fermion masses, arranged by generation. The charged leptons are green, the $-1/3$ quarks are orange, and the charge $+2/3$ quarks are purple. The neutrinos are way off to the left.

cosmology and particle physics in Section 5.3.). This bound is

$$\sum_\nu m_\nu \lesssim 0.25 \text{ eV}$$

There are still lots of possibilities consistent with these bounds. It may even be, for example, that one neutrino is massless while others have mass ~ 0.1 eV or so. Still, our ignorance notwithstanding, a rough summary of the masses of all fermions is shown in Figure 40. Even with our limited knowledge, it's clear that neutrinos aren't like the other particles. There is six orders of magnitude separating the mass of the top quark from the mass of the electron. Then there is a gap of another six order of magnitude before we get to the neutrinos. The first question we should ask is: why?

We don't have a definitive answer to this question. But we do have a plausible answer. As I will now explain, there are a number of different possibilities for the way that neutrinos get a mass, but some of these offer a clear explanation for why neutrino masses are so much smaller than those of the charged fermions. To understand this, we will need to delve once again into some intricacies of quantum field theory and the Dirac equation.

Let's first recall some facts that we learned previously. A massless fermion can come in one of two kinds: left-handed or right-handed. Furthermore, if a particle has one orientation then it's anti-particle necessarily has the opposite. Applied to neutrinos, this means that the particles that interact through the weak force, which are necessarily the particles we detect in beta decay and other processes, are:

- A left-handed neutrino
- A right-handed anti-neutrino

When we refer to a "left-handed neutrino" below, this is shorthand for "left-handed neutrino and the right-handed anti-neutrino". They come together as a pair. You can't have one without the other.

In contrast, we've never observed the parity counterparts of these objects. This means that we've seen neither a right-handed neutrino nor a left-handed anti-neutrino in experiments. When we talk about a "right-handed neutrino" below, this will be shorthand for "right-handed neutrino and the left-handed anti-neutrino". Again, these come as an entwined pair.

The final important fact that we need is that massive particles arise by gluing together a left-handed fermion with a right-handed fermion. So the fact that neutrinos have a mass suggests that we should have both left- and right-handed neutrinos in the game, even though we can only directly observe the left-handed ones. However, nothing with neutrinos is as straightforward as it seems. One of the reasons they're special is because, as we saw in Section 4.1 (see, in particular, Table 2) the putative right-handed neutrino doesn't feel any force at all. It's this latter fact that is special to neutrinos and, as we now explain, opens up a new opportunity.

A Mass From the Higgs

First, it is quite feasible that neutrinos get a mass from the same mechanism as all other fermions, namely through interactions with the Higgs boson.

Recall that the interactions of all fermions with the Higgs field are characterised by a dimensionless number called a *Yukawa coupling.* The mass of the fermion is then given by (4.5)

$$\text{mass} = \frac{\lambda}{\sqrt{2}} \times 246 \text{ GeV} \tag{4.9}$$

where 246 GeV is the energy scale at which the Higgs settles down. The values of the Yukawa couplings range from $\lambda \approx 1$ for the top quark to $\lambda \approx 3 \times 10^{-6}$ for the electron.

This same story now repeats for the neutrinos. All you have to do is tune the value of the Yukawa coupling to $\lambda \approx 10^{-12}$, or whatever is needed to explain the three masses. In the context of neutrinos, this standard mechanism for generating mass is called a *Dirac mass.*

I stress that there's nothing logically wrong with this approach. The Yukawa couplings for charged fermions already range from 1 to 10^{-6} and so you may not feel unhappy by stretching the range down to 10^{-12}. But it does feel like the vastly different masses of neutrinos are crying out for a different explanation. Happily, one exists.

A Mass Without the Higgs and the Seesaw Mechanism

For all other fermions, the need for a Higgs field to induce a mass can be traced to the fact that the left-handed and right-handed particles experience different forces. The same is true for neutrinos, but with the key additional fact that the right-handed neutrino experiences no force at all.

But the lack of any force brings a level of freedom that other fermions do not enjoy. This is because the force that a particle experiences is used to distinguish between the particle and its anti-particle. For example, the difference between an electron and a positron lies in its electric charge. A particle that experiences no force — like the right-handed neutrino — may well be its own anti-particle.

We now tie this observation together with some facts about fermions. First, a fermion gets mass only by coupling together a left-handed and right-handed piece. Second, if we have a right-handed massless particle, then its anti-particle is left-handed. Third, if the particle experiences no force, then it can be its own anti-particle. When the dust settles, all of this means that the right-handed neutrino is unique among particles because it can get a mass *without* interacting with the Higgs field. It does so by interacting with itself! This kind of mass is called a *Majorana mass*, named after the Italian theorist Ettore Majorana who first realised this possibility in 1937.

We will discuss some consequences of the Majorana mass below but there is one important point that we flag up immediately. Forces are not the only way to distinguish particles from anti-particles: the conservation laws that we mentioned in Section 4.3.5 provide another. Recall that if we have a theory with conservation of lepton number, we count electrons and neutrinos as $+1$ and positrons and anti-neutrinos as -1. The converse of this statement is that if a neutrino is the same as an anti-neutrino, then there can be no conservation of lepton number!

This gives the smoking gun for a Majorana mass: in any theory where neutrinos have a Majorana mass, lepton number is not conserved. As explained in Section 4.3.5, we would then expect to see neutrinoless double beta decay. This is one reason why those ongoing experiments are so important: they will greatly help complete our understanding of the neutrino sector of the Standard Model.

So we learn that neutrinos can get masses in two ways:

- A Dirac mass m, from left-handed and right-handed neutrinos interacting with the Higgs field.
- A Majorana mass M, from the right-handed neutrino coupling to itself.

Of course, when we do experiments on a neutrino, we measure just one mass. Which one do we see? The answer is rather nice. It is the combination[13],

$$\text{mass} = \left| -\frac{M}{2} \pm \frac{1}{2}\sqrt{M^2 + 4m^2} \right| \tag{4.10}$$

The ± sign here is telling us that we should see two particles, each their own anti-particle, with different masses.

Now comes the key idea. The Dirac mass m comes from the Higgs mechanism, so is given by (4.9) for some unknown Yukawa coupling λ. But the Majorana mass M is unrelated to the Higgs mechanism and could be very large, $M \gg m$, perhaps coming from some unknown physics at a high energy scale that we have yet to understand. If this is true, the two masses in (4.10) are approximately

$$\text{mass} \approx M \quad \text{and} \quad \frac{m^2}{M} \tag{4.11}$$

The particle with mass M is essentially the right-handed neutrino and is very heavy. We have yet to detect this in any experiment. Meanwhile, the other particle, which can be identified with the left-handed neutrino that we observe experimentally, has mass m^2/M. The key point is that the Dirac mass need not be particularly small; it could take the same kind of value as the other quarks and leptons. But if the Majorana mass is bigger still, with $M \gg m$, then we would see a tiny neutrino mass m^2/M. If this is the way Nature works, then the tiny value of the observed neutrino mass comes about not because the Dirac mass m is very small, but because the Majorana mass M is very large. This in known as the *seesaw mechanism*.

As an example, here are some sample numbers. Suppose that the Dirac mass of the neutrino is as high as 100 GeV, comparable to the mass of the top quark. If the Majorana mass is around $M \approx 10^{15}$ GeV which, as we will discuss in Section 5.1.1, is the scale of grand unified theories, then we naturally get a neutrino mass of the right order of magnitude $\sim 10^{-2}$ eV.

---

[13] This formula comes from solving a quadratic equation. Specifically, it comes from finding the eigenvalues of the matrix $\begin{pmatrix} 0 & m \\ m & M \end{pmatrix}$ where the off-diagonal terms are the Dirac mass and the term in the lower-right is the Majorana mass.

A Mass Without a Right-Handed Neutrino at all!

The seesaw mechanism provides a very natural explanation for why neutrinos are so much lighter than everything else. However, we are notably relying on physics at a high energy scale, way beyond our current reach, and this mechanism clearly predicts the existence of a new particle of a big mass M that we have still to discover.

There's something a little unsettling about this. Because, at the end of the day, we introduced the right-handed neutrino only to find that it has a very high mass and can be ignored, leaving behind the light left-handed neutrino that we actually detect in experiments. But if the right-handed neutrino is so heavy that we can't see it, then surely there should be a way to describe the physics that it leaves behind without the need to invoke it!

Indeed, there is. Here is the way that it works. Although the left-handed neutrino experiences the weak force, it has the property, unique among all the fermions, that it can bind together with the Higgs field to produce something that is neutral under all forces. (You can see this if you compare the charges of fermions in Table 2 with the charges of the Higgs field in Table 4.2.) This means that although we can't introduce a Majorana mass for the left-handed neutrino alone, we can introduce a Majorana mass for the left-handed-neutrino-Higgs combo.

Clearly, this mass once again involves the Higgs field. However, it's different from the Yukawa terms that give the other fermions a mass. First, it involves only the left-handed neutrino, not the right-handed. Second, like any Majorana mass, it violates conservation of lepton number and so will give rise to neutrinoless double beta decay. And finally, and perhaps most importantly, it involves *two* interactions with the Higgs field and not just one[14]. This means that if $\langle \phi \rangle = 246$ GeV is the expectation value of the Higgs field, the Majorana mass of the left-handed neutrino will be proportional to ϕ^2. But this doesn't have the right dimensions to be a mass! This means that we must introduce another scale M, so that the resulting mass of the left-handed neutrino is roughly ϕ^2/M. This is the same form as we saw in the seesaw mechanism (4.11).

From this perspective, the mysterious new scale M is associated to some novel physics at a high energy that we don't yet understand. It could be the mass of the right-handed neutrino as in the seesaw mechanism, or it could be something else entirely. Either way, the irony of the seesaw mechanism remains: detecting a very small Majorana mass for the neutrino is clearly pointing to some new physics at a very high scale!

---

[14] In more precise language of effective field theory, it is a dimension 5 operator built out of Standard Model fields, in contrast to the dimension 4 Yukawa terms.

Some Other Way?

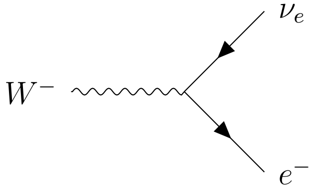
Above we've sketched the most plausible scenarios for neutrinos to get a mass. However, they're not the only possibilities. One could, for example, introduce further scalar fields that act like the Higgs boson, but carry different quantum numbers. If these too get an expectation value, it's possible to arrange for the neutrinos to get a mass. As you can see, our need to better understand the neutrino sector of the Standard Model has some urgency.

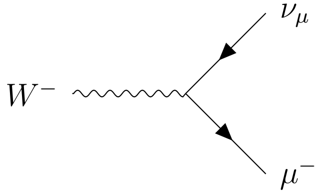
4.4.2 Neutrino Oscillations

So far we have described the different ways in which neutrinos can get a mass. But we haven't yet explained how we know that they have mass. After all, it's not like we can simply collect a bunch of neutrinos in a jar and weigh it. Instead, our information comes in a less direct manner.

We have met the key piece of physics already. In Section 4.3.3, we described how the assignment of mass to quarks is misaligned with the way the weak force acts on the quarks. This resulted in the phenomenon of *quark mixing*, described by the CKM matrix.

An entirely analogous phenomenon is at play in the lepton sector. It's simplest to explain what's going on by starting with two neutrinos, ignoring the third for now. To this end, we'll consider just ν_e and ν_μ. These are defined to be the neutrinos that couple to the electron and muon respectively, as in the following diagrams





But, just as with quarks, the ν_e and ν_μ particles that appear in these interactions are not the particles that have a well defined mass. Instead, there is a mixing and the neutrinos that have a specific mass are ν_1 and ν_2 defined by

$$\begin{aligned}\nu_1 &= \nu_e \cos\theta + \nu_\mu \sin\theta \\ \nu_2 &= \nu_\mu \cos\theta - \nu_e \sin\theta\end{aligned}$$

This is entirely analogous to the quark mixing that we saw in (4.7). It turns out that $\theta \approx 33^\circ$ for leptons, so that $\sin\theta \approx 0.55$. This is somewhat larger than $\theta_{\text{Cabbibo}} \approx 22^\circ$ that we saw in quark mixing. This, it turns out, is what storytellers call "foreshadowing".

At this stage of the argument, however, there's a slight change of perspective. In the context of quarks, when we hold a meson in our hand (metaphorically speaking) we know that it has a definite mass. The mixing then shows up because this meson interacts through the weak force with quarks of other generations

For neutrinos, this situation is reversed. If we've got a beam of neutrinos then it came from some phenomenon involving the weak force, usually associated in some way to beta decay. This means that we know our experiment emitted a neutrino like, say, ν_e with definite flavour but, at least if $\theta \neq 0$, this neutrino does not have a definite mass. What happens next is quite wonderful. The kind of particles that happily travel along without adventure are those with definite mass (known, in the language of quantum mechanics, as energy eigenstates.) But ν_e doesn't have this property. And this has a dramatic effect: as the beam travels some distance, the neutrinos oscillate from ν_e to ν_μ and then back again.

There is a fairly simple formula that describes how this happens. Suppose that the difference in the masses of ν_1 and ν_2 is

$$\Delta m^2 = m_2^2 - m_1^2$$

measured in eV. If the neutrinos have kinetic energy E (measured in GeV) and travel a distance L (measured in km) then the probability that ν_e transforms into ν_μ is given by

$$P(\nu_e \to \nu_\mu) = \sin^2(2\theta)\, \sin^2\left(1.27 \times \frac{L\,\Delta m^2}{E}\right) \tag{4.12}$$

The fact that this probability depends on sine functions is telling us that the change of flavour is an oscillation, in the sense that it goes back and forth. The formula contains two fundamental parameters: the mixing angle θ and the difference in masses Δm^2. To see oscillations, both need to be non-zero. The formula also contains two parameters that can vary from one experiment to another: the energy E of the beam and the length travelled L. In principle, by varying E and L, and seeing how one kind of neutrino morphs into another, we can determine the mixing angle θ and mass difference Δm^2.

We explain more about how these experiments are done in Section D.4. Here, instead, we focus on the results. For reasons that will become clear, I'll first describe what we know about the mixing angles and only then turn to the masses.

Neutrino Mixing Angles

With three generations, neutrino mixing is described by introducing a 3 × 3 matrix, entirely analogous to the CKM matrix that we met for quarks in (4.8). This is

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{4.13}$$

On the left-hand side we have neutrinos ν_e, ν_μ and ν_τ that interact with their counterpart electrons through the weak force; On the right-hand side we have neutrinos ν_1, ν_2 and ν_3 that have definite mass. Relating them is a 3 × 3 matrix is known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, or simply the neutrino mixing matrix.

The components of the PMNS matrix have now been measured to reasonable accuracy. The absolute values are roughly

$$\begin{pmatrix} |U_{e1}| & |U_{e2}| & |U_{e3}| \\ |U_{\mu 1}| & |U_{\mu 2}| & |U_{\mu 3}| \\ |U_{\tau 1}| & |U_{\tau 2}| & |U_{\tau 3}| \end{pmatrix} \approx \begin{pmatrix} 0.8 & 0.5 & 0.1 \\ 0.3 & 0.5 & 0.7 \\ 0.4 & 0.6 & 0.6 \end{pmatrix}$$

Some values are known fairly well; others less well. There are, for example, error bars of ± 0.1 on $U_{\tau 2}$.

The first thing to note is that the PMNS matrix is strikingly different from the CKM matrix describing the mixing of quarks[15]. In the quark sector, the CKM matrix was close to being the unit matrix, with just small off-diagonal elements. This meant that there was close alignment between the masses and the weak force.

But we see no such thing in the neutrino sector. The mixing is pretty much as big as it can be! Once again, we see that, in quantitative detail, the neutrinos really behave nothing like the charged fermions.

We do not have an explanation for the structure of the PMNS matrix. Indeed, its form came as a surprise to theorists. Surely it is telling us something important. It's just we don't yet know what!

---

[15] Recall that $\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$ with $\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.97 & 0.22 & 0.004 \\ 0.22 & 0.97 & 0.04 \\ 0.009 & 0.04 & 0.999 \end{pmatrix}$.

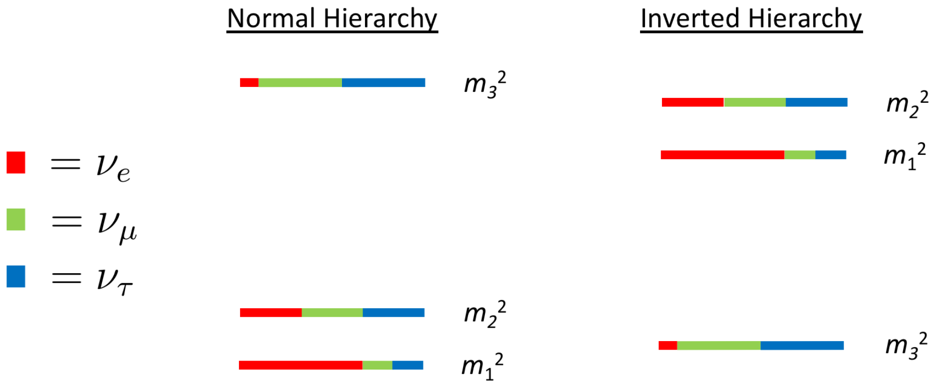


Figure 41. A colour coded description of the possible ordering of neutrino masses.

Neutrino Mass Differences

The mixing angles are a surprise. They tell us that each of the particles ν_1, ν_2 and ν_3 that have definite mass are not closely associated to the particles ν_e, ν_μ and ν_τ that have definite weak interaction.

Inverting the relation (4.13), we can make the following statements about the neutrinos ν_1, ν_2 and ν_3 that have definite masses:

- ν_1 acts like an electron neutrino two thirds of the time, and as a muon or tau neutrino the other third.
- ν_2 acts like any one of the three neutrinos one third of the time.
- ν_3 acts like a tau neutrino 45% of the time and like a muon neutrino 45% of the time. The remaining 10%, it acts like an electron neutrino.

With this in place, we can now describe what we know about the mass differences. First, ν_1 is known to be lighter than ν_2 and the squares of their mass differ by

$$m_2^2 - m_1^2 \approx 7.4 \times 10^{-5}\ \mathrm{eV}^2$$

The difference in their masses is of order $\sim 10^{-2}$ eV, an order of magnitude smaller than the biggest mass. We also know the difference between the masses of ν_3 and ν_2 but, crucially, we don't yet know which one is heavier! We have

$$m_3^2 - m_2^2 = \pm 2.5 \times 10^{-3}\ \mathrm{eV}^2$$

Of course, if we could measure the mass difference between m_1 and m_3.then we would be able to resolve this ± ambiguity. As it stands, we just don't know the order of the masses.

The two possibilities are shown in Figure 41. Given the pattern seen in all other fermions, one might expect that the electron neutrino ν_e would be the lightest. Since the ν_e has the biggest overlap with ν_1, this would mean that ν_1 is lightest. This is referred to as the *normal hierarchy*. But, as we've seen, very little about the neutrinos follows our expectation. So another possibility is that ν_3, which contains very little of the electron neutrino, is the lightest. This is called the *inverted hierarchy*.

CP Violation Among Neutrinos

Neutrino mixing provides another forum in which the Standard Model can exhibit CP violation. And, as we described in Section 4.3.4, this also entails an opportunity for the violation of time reversal invariance.

In the quark sector, the violation of CP symmetry shows up as a complex number in the CKM matrix. The same is true in the neutrino sector. It turns out that if the neutrinos have only a Dirac mass, then there is a single complex number while if the neutrinos have a Majorana mass then there is an opportunity to introduce two more.

We do not currently have a good handle on these complex numbers. (Indeed, we have no handle whatsoever on the CP violation coming from Majorana masses.) However, preliminary results suggest that the complex numbers are non-vanishing and there is CP violation in the neutrino sector.

The good news is that, although difficult to measure, CP violation in the neutrino sector is conceptually rather more straightforward than in the quark sector. In particular, CP simply exchanges all left-handed neutrinos with right-handed anti-neutrinos. This means that if CP is preserved, the following probabilities are equal

$$CP \;\Rightarrow\quad P(\nu_\alpha \to \nu_\beta) = P(\bar{\nu}_\alpha \to \bar{\nu}_\beta)$$

where you can replace ν_α and ν_β with your favourite choice of flavour from ν_e, ν_μ and ν_τ. To detect CP violation, we (simply!) have to do an experiment that shows the probability for a neutrino to change flavour is not the same as the probability for its anti-particle to do the same thing.

Similarly, the violation of time reversal is also easier to state among neutrinos. In a world that is time-reversal symmetric, the probability of a neutrino morphing into a different one would coincide with the probability that this process is reversed,

$$T \;\Rightarrow\quad P(\nu_\alpha \to \nu_\beta) = P(\nu_\beta \to \nu_\alpha)$$

This is an experiment that could, in principle, be performed to discover a violation of time reversal invariance in our world.

Finally, as we discussed in Section 4.3.4, there is a mathematical theorem that says the laws of physics must be invariant under the combination CPT. In the language of neutrinos, this theorem tells us that

$$CPT \;\Rightarrow\quad P(\nu_\alpha \to \nu_\beta) = P(\bar{\nu}_\beta \to \bar{\nu}_\alpha) \tag{4.14}$$

A mathematical theorem is all well and good but it's nice to confront it with experiment. While the expectation is that CP and time reversal will both be found wanting in the neutrino sector, if it was found that the probabilities (4.14) don't coincide, that would surely rock our understanding of physics.

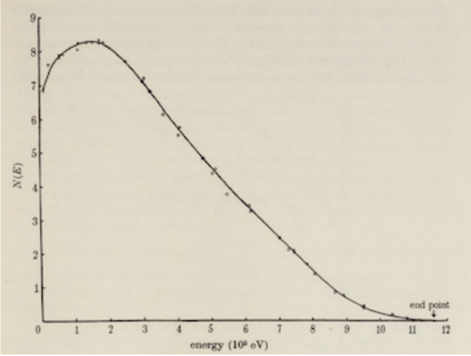
D Interlude: Big Science for Weak Things

For many decades, there was just one known manifestation of the weak force: beta decay. Much of our early understanding of the weak force came from careful study of this phenomenon.

We described some of the initial flurry of discovery in Interlude A. Radiation was found, largely by accident, in the darkness of Becquerel's desk drawer in 1896. By 1898, Rutherford had determined that the radiation from uranium consists of two different types – α rays and β rays – characterised by their penetrating power. And by 1900 it was understood that β rays were composed of the newly discovered particle called the electron. The next step was to further study the properties of these rays.

That, it turns out, was not easy. It had long been understood that α particles are emitted from the nucleus with a fixed velocity, and the general expectation was that the same would be true of β rays. However, the experiments were significantly harder. The leading experts in this game were Hahn, Meitner, and von Baeyer. Their first experiment confirmed that electrons were emitted with a uniform velocity; their second suggested a spread of velocities; their third a collection of distinct, but fixed, velocities. The situation was, to put it mildly, confusing.

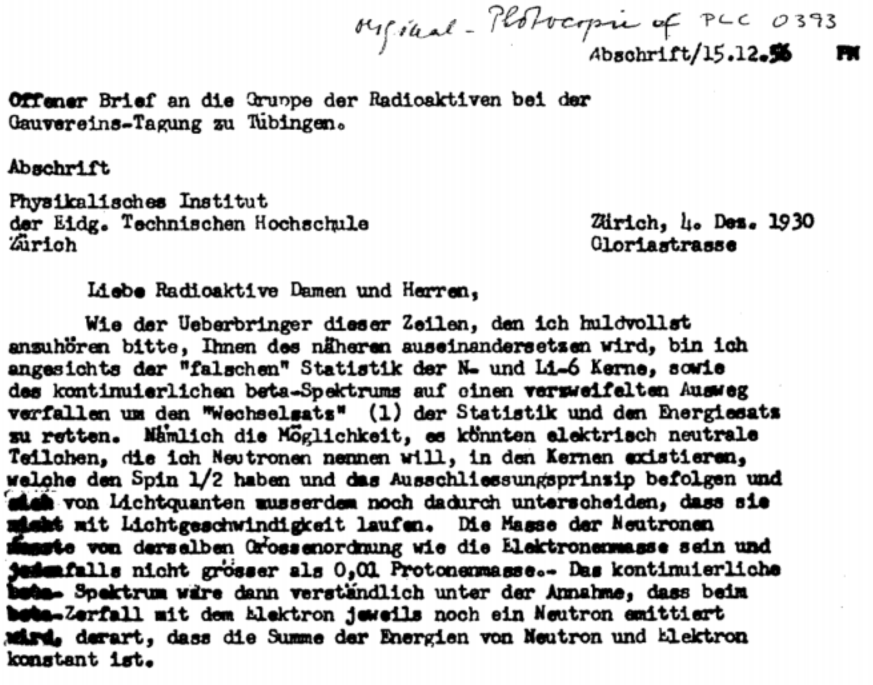
The fog finally lifted in 1914. James Chadwick had completed his PhD under Rutherford and gone to work with Geiger in Berlin. (We met an older version of Chadwick in Section A.4 where we recounted his discovery of the neutron.) While the earlier experiments had used somewhat temperamental photographic plates to detect electrons, Chadwick used the counter recently invented by his postdoc mentor. His experiments made it clear that the electrons were emitted with a range of different velocities. A typical plot of energies is shown in the figure on the right[16].



1914 was not a great year to be an Englishman working in Berlin. Shortly after finishing his experiment, Chadwick was arrested and interned in the stables of a racecourse where he would spend the rest of the war, the monotony broken only by regular

---

[16] This was taken from a paper by G. J. Neary entitled "The β-Ray Spectrum of Radium E". (Radium E is now known as ^{210}Bi.)



original - Photocopie of PLC 0393

Abschrift/15.12.56 FM

Offener Brief an die Gruppe der Radioaktiven bei der
Gauvereins-Tagung zu Tübingen.

Abschrift

Physikalisches Institut
der Eidg. Technischen Hochschule
Zürich

Zürich, 4. Dez. 1930
Gloriastrasse

Liebe Radioaktive Damen und Herren,

Wie der Ueberbringer dieser Zeilen, den ich huldvollst anzuhören bitte, Ihnen des näheren auseinandersetzen wird, bin ich angesichts der "falschen" Statistik der N- und Li-6 Kerne, sowie des kontinuierlichen beta-Spektrums auf einen verzweifelten Ausweg verfallen um den "Wechselsatz" (1) der Statistik und den Energiesatz zu retten. Nämlich die Möglichkeit, es könnten elektrisch neutrale Teilchen, die ich Neutronen nennen will, in den Kernen existieren, welche den Spin 1/2 haben und das Ausschliessungsprinzip befolgen und sich von Lichtquanten ausserdem noch dadurch unterscheiden, dass sie nicht mit Lichtgeschwindigkeit laufen. Die Masse der Neutronen müsste von derselben Grossenordnung wie die Elektronenmasse sein und jedenfalls nicht grösser als 0,01 Protonenmasse.- Das kontinuierliche beta- Spektrum wäre dann verständlich unter der Annahme, dass beim beta-Zerfall mit dem Elektron jeweils noch ein Neutron emittiert wird, derart, dass die Summe der Energien von Neutron und Elektron konstant ist.

Figure 42. Liebe Radioacktive Damen und Herren.

deliveries of *Nature* and the occasional piece of scientific apparatus to help him pass the time.

Chadwick's work was not the end of the story, and a great deal of to-ing and fro-ing took place, largely between Lisa Meitner's group in Berlin and the pair of Chadwick and Charles Ellis, now back in Cambridge. But by the late 1920s, the situation was decisively settled: the spectrum of beta rays was continuous.

What to make of this? It was in sharp distinction to both alpha rays and gamma rays, where the radiation was emitted with a definite energy. Moreover, the emission of alpha and gamma rays was well understood as the transition between different states, just like the clean spectra of atoms. What could a continuous spectrum possibly mean?

D.1 The Neutrino

Two proposals were put forward in 1930. Bohr, ever the revolutionary, was keen to ditch energy conservation. Pauli, however, had a different idea: a new, hitherto undetected, particle. If two particles were emitted in beta decay, then the energy could be shared among them in different ways. There would then be no reason for the electron that we observe to carry a unique energy.

Pauli's first made this proposal public, just days after a messy divorce, in a famous "Dear Radioactive Ladies and Gentleman" letter sent to a conference in Tubingen. This is shown in Figure 42. The first paragraph translates as

> "I have come upon a desperate way out regarding the 'wrong' statistics of the N- and Li 6-nuclei, as well as to the continuous β-spectrum, in order to save the 'alternation law' of statistics and the energy law. To wit, the possibility that there could exist in the nucleus electrically neutral particles, which I shall call neutrons, which have spin 1/2 and satisfy the exclusion principle and which are further distinct from light-quanta in that they do not move with light velocity. The mass of the neutrons should be of the same order of magnitude as the electron mass and in any case not larger than 0.01 times the proton mass. The continuous β-spectrum would then become understandable from the assumption that in β-decay a neutron is emitted along with the electron, in such a way that the sum of the energies of the neutron and the electron is constant."

It is clear from the letter that Pauli is trying to solve two problems at once. We now know that these two problems in fact require two new particles. The first problem was described in Section A.4 and arose from the prevailing view that the nucleus is comprised of A protons with $A-Z$ electrons making up the charge difference. Pauli doesn't dissent from this viewpoint, but proposes that there are further neutral particles in the nucleus which ensures that the nucleon spin agrees with the sum of its constituents. As we saw, this problem was solved in 1932 by the discovery of the neutron.

The second problem addressed by Pauli is the one of interest here. The continuous spectrum of β-decay is resolved if a light, neutral particle is also emitted. Pauli calls this the neutron, but obviously that name was later taken. The Italian name *neutrino* was coined by Fermi in 1933.

Detection

Pauli was unduly nervous about his proposed neutrino. Indeed, his letter continues

> "I don't feel secure enough to publish anything about this idea, so I first turn confidently to you, dear Radioactives, with a question as to the situation concerning experimental proof..."

The neutrino hypothesis was accepted long before the particle was seen directly in experiment. It was an integral part of Fermi's theory of beta decay, which agreed too well with the observed phenomenology. Moreover, the same theory made it clear how difficult a direct detection would be. As we've mentioned previously, a typical neutrino emitted in beta decay will interact only once as it travels through a light-year of lead.

The resolution to this problem is not to focus on a single neutrino. Instead, you need to go to a place where there are many. This was the goal of a team lead by Clyde Cowan and Fred Reines, working at the Los Alamos lab. Originally their idea was to detect the neutrinos emitted in a nuclear explosion, but they soon realised that a nuclear reactor offered a better (and presumably safer) alternative. Because of the ghostly nature of neutrinos, they named their original experiment "project poltergeist".

They set up their experiment in the Savannah reactor site, a nuclear reservation in South Carolina, USA. The nuclear reactor emits a flux of 10^{13} neutrinos per cm^2 per second, through beta decay

$$n \longrightarrow p + e^- + \bar{\nu}_e$$

The experimental challenge is to take the resulting neutrinos and observe the inverse process

$$\bar{\nu}_e + p \longrightarrow n + e^+$$

The bad news is that such processes are extremely rare. The good news is that both products n and e^+ have a distinctive signature. The experimental set-up consisted of two tanks, each holding 200 litres of water. These provide the targets for the neutrinos. The resulting positron quickly annihilates with an electron in the water, emitting two gamma rays. Slightly later, the neutron is captured. To aid this process, Cowan and Rienes dissolved cadmiun salts in the water. These are known to efficiently capture neutrons, emitting a third gamma ray in the process with a well understood energy spectrum. The signature of neutrinos is then two coincidental pulses of gamma rays, the first a pair; the second, delayed by a few milliseconds, a single photon, all of which were registered by liquid scintillators which surrounded the two tanks.

The experiment ran for almost 1400 hours and, when the reactor was on, detected roughly 3 coincidental pulses an hour.

Announcing a major discovery must be both exciting and nerve-wracking. Announcing the discovery to Wolfgang Pauli, not known to suffer fools gladly, doubly so. But on June 14th 1956, Cowan and Rienes sent Pauli a telegram that read:

> "We are happy to inform you that we definitely detected neutrinos from fission fragments by observing inverse beta decay of protons. Observed cross section agrees well with the expected six times ten to minus forty four square centimeters"

For some reason, Pauli's reply was never sent. It exists only in the form of a draft discovered later in his papers and reads: "Thanks for the message. Everything comes to him who knows how to wait."

The Muon Neutrino

It had long been known that when charged pions decay to a muon, they also emit a second spin 1/2 particle with all the properties (or lack thereof) of a neutrino.

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \quad \text{and} \quad \pi^- \longrightarrow \mu^- + \bar{\nu}_\mu$$

We now know, of course, that this is a different kind of neutrino from the one that appears in beta decay. The question is: how to tell ν_μ and ν_e apart?

One way is to run the inverse beta decay experiment again. What do you see if this neutrino collides with a proton? The two obvious possibilities are:

$$\begin{aligned} \nu_\mu + p &\overset{?}{\longrightarrow} n + \mu^+ \\ \text{or} \quad \nu_\mu + p &\overset{?}{\longrightarrow} n + e^+ \end{aligned}$$

If the muon neutrino ν_μ is, in reality, the same object as the electron neutrino ν_e, then both of these processes should occur, presumably with comparable frequency. In contrast, if ν_μ is truly distinct from ν_e, then you would expect to see only the first process and not the second.

At this stage, accelerators once again come to the fore. A synchrotron can accelerate protons to, say, 30 GeV which, upon a collision with a fixed target, create a beam of pions. 25 meters down the road, these pions decay into a beam of muons and neutrinos. At this point, you need to put some shielding in place to remove the muons, leaving behind just the neutrinos. In the original experiment, performed at Brookhaven in 1962 by Lederman, Schwartz, Steinberger, and others, this shielding was achieved by 5,000 tonnes of steel from a decommissioned battleship (often falsely claimed to be the USS Missouri, which seems unlikely given that this ship, although mothballed in the late '50s, was subsequently reactivated and even served in the gulf war).

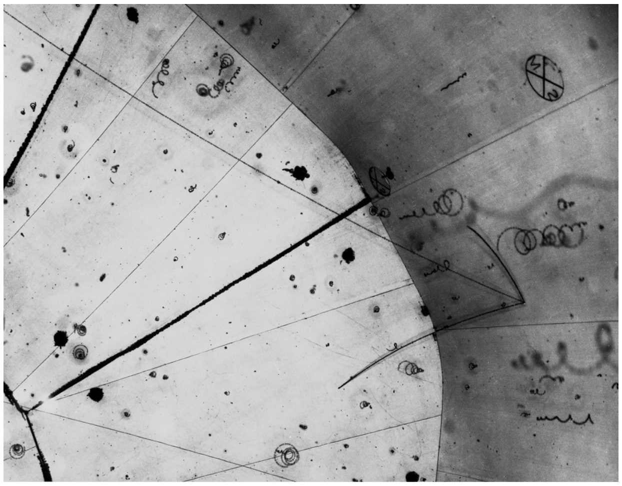


Figure 43. Muonic inverse beta decay, seen in a bubble chamber. The neutrino entered from the right where it hit a proton at the vertex where three lines meet. The proton is the short line moving up and to the left; the pion the slightly longer line moving down and to the left. The longest line of all, exiting top-left, is the muon. This image was taken at Argonne National Laboratory.

Waiting on the other side of the battleship, was a novel kind of detector known as *spark chamber*. This consists of an array of plates, sitting at a high voltage, with gas between them. When a charged particle passes through it ionises the gas which, when it hits the plates, leaves a trail of visible sparks. One can easily distinguish the track a muon from the much lighter electron.

One advantage of using accelerators is that the neutrino cross-section increases with energy. As Pais puts it, while you need a lightyear of lead to stop an MeV neutrino, a few million miles will do the job for a GeV neutrino. And, indeed, the 30 GeV machine was just sufficient to see the desired result: after several months of experiments, starting with roughly 10^{17} protons hitting the target, 29 anti-muons were seen and no positrons. This means that the subscript matters: ν_μ is indeed different from ν_e.

In subsequent years, muonic inverse beta decay could be seen in bubble chambers. The first such example is shown in Figure 43.

D.2 Not P and Not CP Either

We started Chapter 4 with a description of parity violation, since this sets the scene for much of the subsequent structure of the Standard Model. As we reviewed there, parity violation was first seen in the realm of atomic physics through the beta decay of cobalt atoms. But, prior to this, there was a hint of parity violation in the world of particle physics.

Heavy mesons provide the setting to see violation of both parity and of CP. Here "heavy" means mesons that include strange, charm or bottom quarks. The lightest of these is, of course, the strange quark and this is where the phenomena were first observed.

The story starts with what we now call the charged kaon, $K^+ = \bar{s}u$, which weighs in at 494 MeV. Back in the 1950s, before quarks were understood, this particle was seen in experiments. But it had an unusual property: sometimes it decayed to two pions, and sometimes it decayed to three pions. In fact, this difference was so striking that, at the time, it was assumed that the experiments must be seeing two different particles. They gave these particles names θ and τ. Both names have since been retired and, in the case of τ, upcycled into the name of a lepton, but back in the 1950s these were two of the most interesting particles around. They decayed as

$$\begin{aligned} \theta^+ &\longrightarrow \pi^+ + \pi^0 \\ \tau^+ &\longrightarrow \pi^+ + \pi^+ + \pi^- \end{aligned}$$

But then there was a surprising coincidence that needed an explanation: as far as the experiments could tell, θ^+ and τ^+ had exactly the same mass and lifetime! Why on earth would that be? This was known as the *theta-tau puzzle*.

Of course, we know now the resolution to the theta-tau puzzle: it's that both particles have the same mass because they're actually the same particle – the kaon K^+. But physicists in the 1950s were reluctant to draw this obvious conclusion because it violated one of their cherished principles of physics: that the world should be invariant under parity.

It's not so easy to explain why this is the case without going into the mathematics. But it turns out that the decay to two pions looks identical when reflected in the mirror, while the decay to three pions does not. This isn't because of any obvious reason due to the directions in which the pions fly out. Instead, it shows up only in the subtle fact that the wavefunction of three pions differs by a minus sign upon reflection.

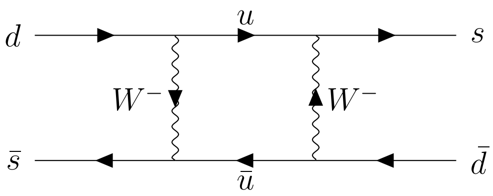
Given this experimental observation, parity could be conserved only if θ^+ and τ^+ were genuinely different particles, with θ^+ staying unchanged when reflected in a mirror, while τ^+ picked up a minus sign in its wavefunction. Conversely, if you want to identify the θ^+ and τ^+ and say that they're the same particle, then you can do so only if you sacrifice the symmetry of parity.

Needless to say, the argument above is somewhat subtle and was far from enough to convince physicists that parity was indeed broken in the weak interaction. The first physicists to take this possibility seriously were T.D. Lee and C.N. Yang (yes, Yang-Mills Yang) who, in 1956, did a systematic analysis of parity symmetry in various experiments. They came to the conclusion that, while there was overwhelming evidence for parity in the strong and electromagnetic forces, the jury was still out when it came to the weak force. They also proposed the experiment involving the decay of cobalt atoms that was subsequently performed by their colleague, C.S. Wu.

Wu's experiment showed that parity isn't just violated in the weak force a little: it is violated as much as it possibly could be. This spurred particle physicists to find other pieces of evidence. Confirmation came quickly from showing that muons arising from $\pi^+ \longrightarrow \mu^+ + \nu_\mu$ decay are polarised in a manner that can only be explained by parity violation. Moreover, the signature was so stark that even though the experiment was initiated after hearing of Wu's results, it was completed before she had finished. In the end, the two papers establishing parity violation in very different ways were published back to back.

Neutral Kaons and CP

Kaons also played a starring role in the discovery of CP violation. This time it was the neutral kaon $K^0 = d\bar{s}$ and its anti-particle $\bar{K}^0 = \bar{d}s$ that were of interest. The phenomenon of quark mixing that we met in Section 4.3.3 means that heavy, neutral mesons of this type have an interesting property: over time, the particle can change into its anti-particle and then back again! This happens through a Feynman diagram of the form



As we learned when discussing quark mixing, those intermediate u and $\bar{u}$ quarks could also be replaced by c and $\bar{c}$, or t and $\bar{t}$.

This means that as a K^0 wends along its merry way, it's constantly switching between a K^0 and $\bar{K}^0$. The actual particles that we observe are some mix of the two. But what mix? Here's where things get interesting.

Under CP symmetry, a K^0 switches with a $\bar{K}^0$ (because particles swap with anti-particles). So if the kaon was to arrange itself in accord with CP, the right mix would simply be

$$\begin{aligned} K_1^0 &= K^0 + \bar{K}^0 \\ K_2^0 &= K^0 - \bar{K}^0 \end{aligned}$$

Then K_1^0 reflects into itself under CP, while K_2^0 reflects into minus itself. (These words make more sense when written as quantum mechanical equations! The right statement is that the two combinations above are CP eigensates.)

Moreover, just as parity conservation (if it existed!) would dictate the possible decays of particles, so too would CP conservation. CP symmetry means that we should see the two different species of kaons decay in different ways

$$\begin{aligned} K_1^0 &\longrightarrow \pi^0 + \pi^0 \\ K_2^0 &\longrightarrow \pi^0 + \pi^0 + \pi^0 \end{aligned}$$

So do we?

The K^0 and $\bar{K}^0$ do indeed mix into two different combinations and, experimentally, these are distinguished by their lifetime. There is a long-lived neutral kaon that is called K_L and a short-lived neutral kaon that is called K_S. At first glance it seems as if all is good, since these two different combinations decay as

$$\begin{aligned} K_S^0 &\longrightarrow \pi^0 + \pi^0 \text{ in around } 10^{-10} \text{ s} \\ K_L^0 &\longrightarrow \pi^0 + \pi^0 + \pi^0 \text{ in around } 5 \times 10^{-8} \text{ s} \end{aligned}$$

This makes it look like we can identify $K_S^0 = K_1^0$ and $K_L^0 = K_2^0$, in which case CP would be preserved.

However, a careful examination shows that this isn't quite the case. In 1964, Christenson, Cronin, Fitch and Turlay created a beam of neutral kaons, and let it travel for 18 m. Travelling at close to the speed of light, it takes around 6×10^{-8} s to travel 18 m, meaning that the short-lived K_S^0 kaons had long since departed and the beam contained only K_L^0 kaons. The goal was to see if they did, indeed, all decay to three pions as expected by CP.

They did not. Of the roughly 23,000 decays, they found 45 decaying into two pions (both $\pi^0 + \pi^0$ and, more commonly, $\pi^+ + \pi^-$). This is only 0.2% of the sample, but was enough to show that CP is not a symmetry of our world. The kaon states that have a definite mass are not quite the K_1^0 and K_2^0 states compatible with CP, but instead take the form

$$\begin{aligned} K_S^0 &\approx K_1^0 - 0.002\, K_2^0 \\ K_L^0 &\approx K_2^0 + 0.002\, K_1^0 \end{aligned}$$

That tiny extra 0.002 piece is the sign of CP violation.

The signature of CP violation in quarks is remarkably subtle. The effect is stronger in mesons that contain bottom quarks, but the essence of the idea remains the same. Recall from Section 4.3.3 that among the indirect implications of this result is the statement that the laws of physics are not invariant under time reversal. It seems surprising that the observable consequences of something so profound can only currently be seen in something so subtle and seemingly insignificant as the decay of certain mesons.

D.3 The Bosons of the Weak Force

Before the discovery of the W-boson, there was beta decay. Before the discovery of the Z-boson, there were *neutral currents*. This is the name given to the process in which a neutrino sneaks in, gives a charged lepton a gentle kick, and then quietly leaves again. For example,

$$\nu_e + e^- \longrightarrow \nu_e + e^-$$

Since we don't actually see the neutrino, the signal is rather subtle: the electron simply flies off without warning.

The first detection of neutral currents came in 1973 at CERN. A proton synchrotron (known simply as the PS) was used to create a beam of muon neutrinos which then entered an enormous bubble chamber, given the poetic name Gargamelle. You can see the discovery picture in Figure 44.

By this time, the theory of the weak force, with its W- and Z-bosons, was in place and it was possible to get a ballpark figure for their mass. But to find them required a machine that could reach the required energy. Something just shy of a 100 GeV or so should do it.

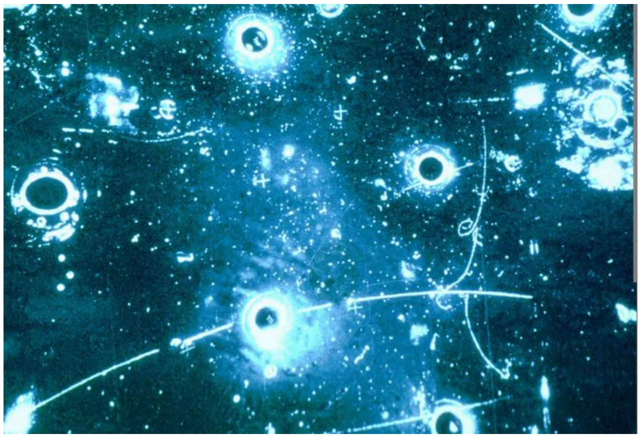
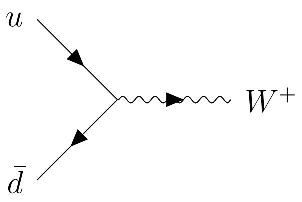


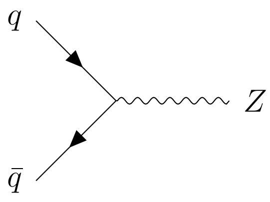
Figure 44. The first detection of neutral currents. The neutrino came in somewhere at the top and sent an electron barreling downwards, creating a shower of e^+e^- pairs, with their tell-tale spiral tracks. A black and white version of this appears in the discovery paper "Search for Elastic Muon-Neutrino Electron Scattering". The same journal contains a second paper with evidence for neutral currents from hadrons.

The W and Z Bosons

By 1976, CERN had turned on its latest machine, uninspiringly named the Super Proton Synchrotron or SPS. The proton beam reached 300 GeV before colliding into a fixed target. Sadly, as we mentioned before, much of the energy in fixed target machines ends up in the kinetic energy of the final product so only $\sqrt{2m_pE}$ of the energy, or about 25 GeV in this case, is available to do something useful. And that's not enough.

In the same year that the SPS turned on, Rubbia, Cline and McIntyre wrote a paper proposing to add a counter-rotating beam of anti-protons to CERN's newest accelerator. The centre-of-mass energy is then $300 + 300 = 600$ GeV. It turns out that around half the momentum in a proton is carried by the gluons, with the remainder split evenly between the three valence quarks. That means that each valence quark carries roughly 1/6 of the energy, so 600 GeV should be sufficient to create both W and Z-bosons through Feynman diagrams of the form





where q and $\bar{q}$ can be any quark-anti-quark pair.

Not everyone in CERN was overjoyed at the prospect of their new toy undergoing such an extensive overhaul so soon after turning on but Carlo Rubbia was, by all accounts, a persuasive if not particularly likeable man. The argument that he might move to Fermilab and do the experiment there was perhaps the decisive one and the SPS was upgraded to the SppS, or Super Proton-Anti-Proton Synchrotron. A key part of the design was an idea to focus the beam known as *stochastic cooling*, introduced by Simon van der Meer.

The first collisions occurred in December 1981 and were recorded in two detectors known as UA1 and UA2, the "UA" standing for "underground area", referencing the fact that both beam and detectors sit 100m below the surface. The goal was to see the decay products.

First, the W-boson. This decays about 70% of the time to quark-anti-quark pairs, such as $W^+ \rightarrow u\bar{d}$, and these appear in the detector as jets. But jets are ten a penny in $p\bar{p}$ collisions. For this reason, you need to focus on the 30% of the time that the W-boson decays as $W^+ \rightarrow \bar{l}\nu$ or $W^- \rightarrow l\bar{\nu}$ with the lepton l either an electron or muon.

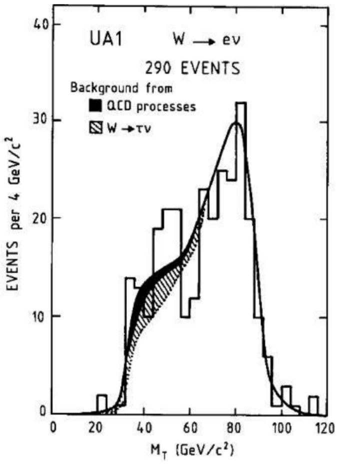
If the W-boson was sitting at rest when it decayed, we would get back-to-back leptons with momentum $\mathbf{p}_l = -\mathbf{p}_\nu$. Of course, you miss the neutrino in the detector, so you should see the electron or muon fly off with its kinetic energy[17] given by $|\mathbf{p}_l| = M_W/2$, half the mass energy of the W-boson.

In reality, the W-boson usually isn't at rest when it decays, but travelling at some unknown velocity aligned with the direction of the beam. Furthermore if the lepton also travels in the same direction of the beam, then you're simply going to miss it: there's too much going on down there and no detector. The trick, therefore, is to look at the *transverse momentum* $\mathbf{p}_T$ of the lepton, meaning the momentum orthogonal to the beam. This obeys $|\mathbf{p}_T| \leq M_W/2$, with equality occurring only when you're lucky, and the W-boson was at rest *and* the lepton emerges perpendicular to the beam. However, one can show that the expected distribution of $|\mathbf{p}_T|$ from many W-boson decays takes a specific shape, peaking at $M_W/2$ and then quickly dropping off.

---

[17] For slowly moving particles, where you can ignore relativity, the kinetic energy is quadratic in the momentum: $E = \frac{1}{2}mv^2 = \frac{1}{2}p^2/m$ where the momentum $p = mv$. But this formula is no longer right when particles move fast. The correct relativistic formula combines the rest mass energy and kinetic energy as $E = \sqrt{m^2c^4 + p^2c^2}$. When particles move extremely fast, with $pc \gg mc^2$, we can neglect the rest mass energy and the kinetic energy is approximately linear in the momentum: $E \approx pc$. This is the formula we've used in the text, with units $c = 1$.

Using this method, the first clear signs of the W-boson were seen in January 1983. The discovery was announced soon afterwards. The figure on the right shows that data collected by UA1 between 1982 and 1985. The data is plotted using a variable related to $\mathbf{p}_T$ called *transverse mass*, a slightly odd name when you first hear it given that mass has no direction.



The search for the Z-boson is cleaner because this time there's no neutrino to miss and you see back-to-back electron-positron pairs. However, the events are rarer, with a frequency of around 10% of the W-bosons decay. For this reason it took a few months more before the Z-boson was also discovered.

While W- and Z-bosons were first discovered at a $p\bar{p}$ collider, their detailed study came with the next upgrade at CERN. This was the Large Electron Positron Collider, or LEP, a 200 GeV machine built in a new tunnel, some 27 km in circumference. While hadron collisions are a mess, e^+e^- collisions allow for exquisitely precise measurements. It is from this experiment, with its four detectors Aleph, Delphi, Opal, and L3, that we have our most accurate understanding of the weak force. The current values of the masses are now known to be

$$M_W = 80.379 \pm 0.0021 \text{ GeV} \quad \text{and} \quad M_Z = 91.1876 \pm 0.0021 \text{ GeV}$$

with comparable precision on their decay widths.

However, there was one last part of the Standard Model that was tantalisingly just out of reach of the LEP machine. This was. . .

The Higgs Boson

By the end of the last century, the LEP data told us that the Higgs boson must be heavier than 114 GeV. But how much heavier?

To answer this, LEP was dismantled and the Large Hadron Collider, or LHC, constructed in its place, designed to collide two proton beams with a centre of mass energy of 14 TeV. The full CERN accelerator complex is shown in Figure 45. There are a number of detectors around the ring, the most important of which are ALICE, used for heavy ion collisions, LHCb, used to study B-mesons, and two multi-purpose detectors called ATLAS and CMS. These latter two detectors were the ones to find the Higgs.

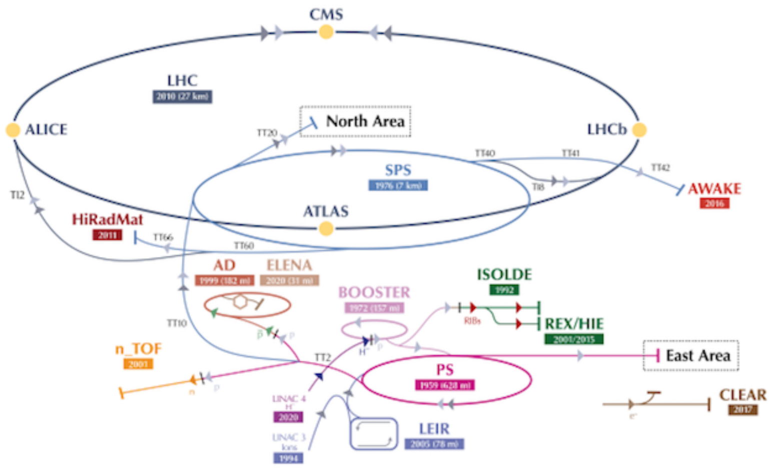
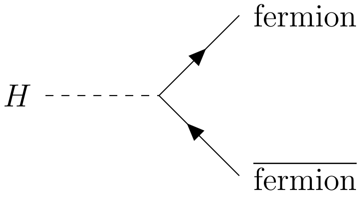
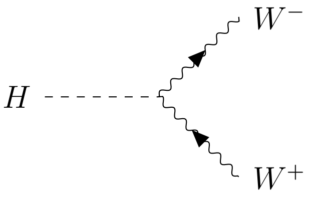
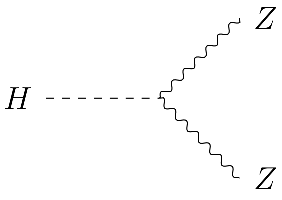


Figure 45. The CERN accelerator complex. You can see the older PS and SPS accelerators which now act as feeders for the LHC.

The Higgs, like the top quark of Interlude C.3, is detected through its decay products. It can decay either into fermions, or into W and Z bosons through diagrams like this







As we explained in Section 4.3, the strength of the interaction is proportional to the mass of the particle, because it comes from the same physics that gives fermions and W and Z bosons a mass in the first place. This means that the Higgs preferentially decays into the particles with higher masses.

The coupling to the top quark is the strongest, but because the top is heavier than the Higgs that decay route woud appear to be ruled out. That means that the decay $H \to b\bar{b}$ is the most common, but it is difficult to distinguish the resulting two jets from the background of a pp collision. The next two decay modes, $H \to W^+W^-$ and $H \to ZZ$, are significantly cleaner since both W and Z can subsequently decay to leptons through $W \to l\bar{\nu}$ and $Z \to l\bar{l}$, giving two lepton and four lepton events respectively where the leptons are either electrons or muons. (Remember, no one sees the neutrino.) The decay through Z-bosons to four leptons is particularly clear and sometimes referred to as the golden channel.

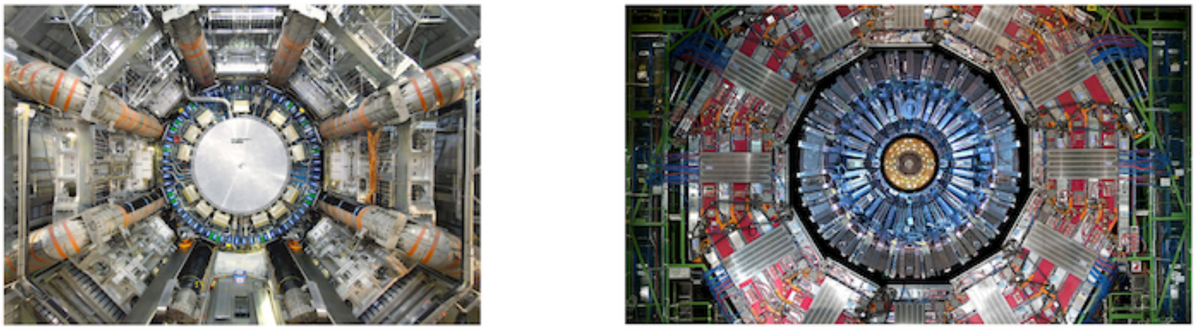
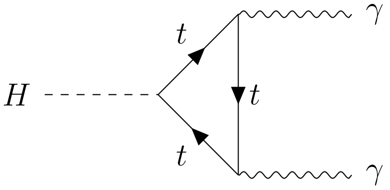


Figure 46. Iconic images of the ATLAS detector, on the left, and CMS, on the right, not to convey any information but simply to let you stare in awe at one of the great engineering feats of all time.

Given that the Higgs decays preferentially to massive particles, it is somewhat ironic that one of the best signals comes from the emission of two photons. These arise from top quarks which, although too heavy to appear directly as decay products, couple so strongly that the one-loop Feynman diagram is comparable to other tree level processes:



On July 4th, 2012, ATLAS and CMS held a joint seminar in which they announced the discovery of the Higgs boson, the final piece of the Standard Model jigsaw. There is nothing as pretty as a bubble chamber photograph to show, merely a small bump above the background for the emission of two photons shown in Figure 47, revealing that the Higgs boson has mass

$$M_H = 125.10 \pm 0.14 \text{ GeV}$$

Since then, a number of different Higgs decay channels have been seen, all impressively (but, to some, disappointingly) in perfect agreement with expectations from the Standard Model.

There is much that we still don't understand about the Higgs, not least its self-coupling and more detailed information about the shape of the Higgs potential $V(\phi)$. There is hope that a future e^+e^- collider – perhaps the proposed ILC – could change

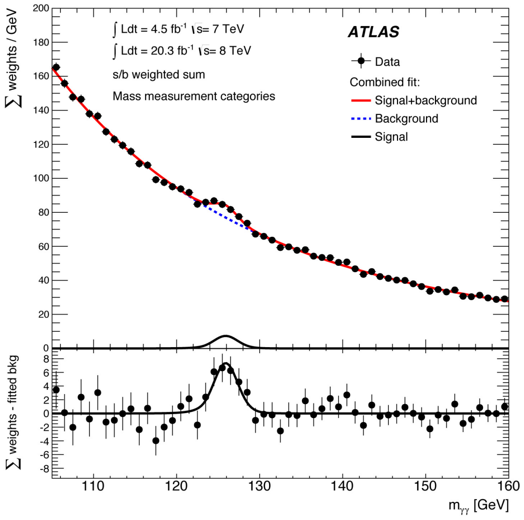


Figure 47. The tiny bump that revealed the Higgs, taken from the ATLAS experiment.

this, allowing us to understand the Higgs sector with the same precision that LEP explored the rest of the electroweak sector.

D.4 Neutrino Oscillations

We end this interlude the same way in which we began: with neutrinos. We saw how neutrinos were first discovered in Interlude D.1. Now we will learn about their mass, a long and involved story that, as we will see, is still far from complete.

Solar Neutrinos

The first hint that something was awry came when neutrinos failed to escape the Sun.

In the late 1960s, Ray Davies built the world's first solar neutrino detector, buried deep underground in the Homestake gold mine in South Dakota. The detector consisted of a large tank filled with a dry-cleaning fluid that, importantly, was rich in chlorine. The neutrinos were detected by a neutrino capture process, which is like stimulated beta decay

$$\nu_e + {}^{37}\text{Cl} \longrightarrow {}^{37}\text{Ar} + e^-$$

These argon atoms were then counted and used as a proxy for the original neutrino.

Neutrinos have the ability to induce such an interaction provided that their energy is greater than ~ 800 keV. However, they do so with extraordinarily small probability. We now know that around 10^{11} neutrinos that originate from the Sun stream through a surface area of one square centimetre here on Earth every second. Most, it turns out, aren't powerful enough to induce the reaction above, but there's still about 10^8, per square centimetre, per second, that can do the job. Yet the number of reactions observed in 600 tonnes of cleaning fluid was just a few a day. (Neutrino experiments use the unit of SNU, where 1 SNU means 10^{-36} interactions per target atom per second. The Homestake experiment detected about 2.5 SNU of solar neutrinos.)

There was a problem however. The observed solar neutrinos were a factor of 3 too small. Detailed calculations of the reactions in the Sun, largely performed by John Bachall, showed that the expected flux was around 8 SNU. Where did the other neutrinos go?

The reaction of the larger scientific community to this puzzle was mostly to shrug and move on. To particle physicists, the Sun looked like a ridiculously messy and complicated object; surely those astrophysicists had screwed up the calculations. Meanwhile, the astrophysicists were bewildered at the possibility that you could reliably detect neutrinos; surely the particle physicists had got something wrong in their experiment. After all, could we even be sure that it was detecting neutrinos from the Sun, and not some other source?

However, as time went on the problem became more urgent. Bachall's theoretical models of the Sun passed many impressive tests, leaving little doubt of their accuracy. Meanwhile further experiments confirmed and refined the Homestake results. In Japan, Kamiokande, and later Super-Kamiokande, detected neutrinos at higher energies through Cerenkov radiation emitted in scattering process

$$\nu_e + e^- \longrightarrow \nu_e + e^- \tag{D.1}$$

Importantly, this gave directional information about the incoming neutrinos and showed clearly that they were coming from the Sun. Meanwhile GALLEX in Italy, and SAGE in Russia, both repeated the experiment with gallium rather than chlorine, now relying on the neutrino capture process,

$$\nu_e + {}^{71}\text{Ga} \longrightarrow {}^{71}\text{Ge} + e^-$$

The advantage is that the threshold is somewhat lower than the chlorine reaction, needing only around 200 keV, meaning that many more of the Sun's neutrinos can partake. Indeed, the number of events seen was significantly higher, at around 75 SNU, but still below the theoretical prediction of 130 SNU. Curiously, the shortfall in these experiments is only around 40%, compared to the 70% seen in the chlorine experiments. This is telling us that the loss of neutrinos is energy dependent. Collectively, these discrepancies between experiment and theory went by the name of the *solar neutrino problem*.

As the years went on, it became increasingly clear that the resolution to the solar neutrino problem must be found in neutrino oscillations which, as we have seen in Section 4.4, requires that the neutrinos have mass. The question was: how to test this? This requires us to count not just the electron neutrinos emerging from the Sun, but also the muon and tau neutrinos. And this is significantly harder. We can't rely on the obvious neutrino capture process

$$\nu_\mu + n \longrightarrow p + \mu^-$$

because the incoming neutrinos have energies less than the 150 MeV needed to create a muon.

An unambiguous resolution to the solar neutrino problem was provided by the Sudbury Neutrino Observatory (SNO), based in the Creighton nickel mine in Ontario, Canada. One novelty was that their tank was filled with heavy water, $D_2 0$, where the hydrogen is replaced by deuterium D. It doesn't take much to split the deuterium nucleus apart; just 2 MeV of energy is enough. Moreover, neutrinos can knock apart a deuterium nucleus in two different ways. A weak interaction involving an intermediate W-boson does the job through a neutrino capture process analogous to those that occur in chlorine or gallium,

$$\nu_e + d \longrightarrow p + p + e^-$$

Only electron neutrinos contribute to such processes. However, the neutrinos can also split the deuterium through a weak interaction involving a Z-boson,

$$\nu + d \longrightarrow n + p + \nu$$

This time there is no charged lepton created, meaning that all three kinds of neutrinos, ν_e, ν_μ and ν_τ contribute.

In addition, SNO measured neutrino scattering events of the form $\nu + e^- \to \nu + e^-$. Electron neutrinos undergo such scattering events through both W-boson and Z-boson interactions, but muon and tau neutrinos only scatter off electrons when an intermediate

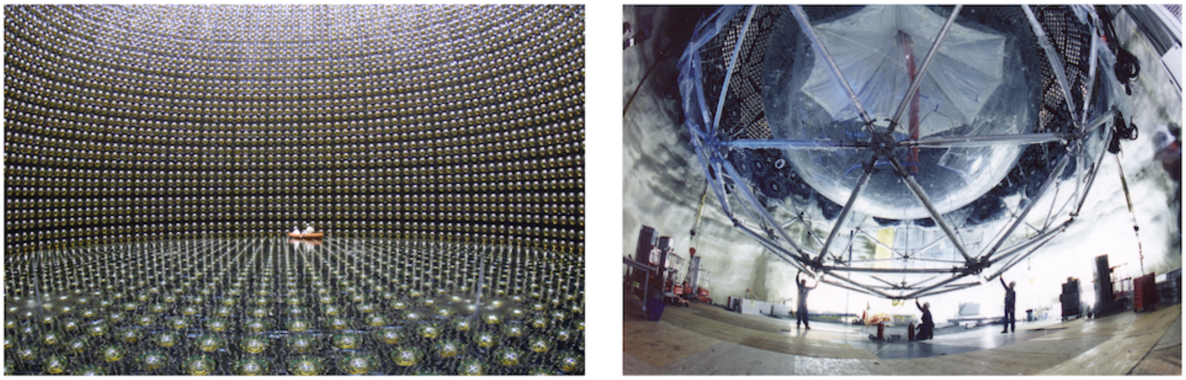


Figure 48. Neutrino detectors tend to look like the lair of a James Bond villain. On the left is a boat cleaning the Super-Kamiokande photosensors as the tank slowly fills up. On the right is the SNO tank, filled with heavy water.

Z-boson is involved. This means that the total rate of such events depends on some combination of the flux of electron, muon and tau neutrinos. (For what it's worth, it turns out to be ν_e flux + 0.15 times ν_μ and ν_τ flux.)

The upshot is that SNO was able to see everything – electron, muon and tau neutrinos. And once you see everything, nothing is missing. The end result agreed perfectly with theoretical expectations of the nuclear reactions inside the Sun. The electron neutrinos missed by previous experiments had transmuted into muon and tau neutrinos, incontrovertible evidence for neutrino oscillations.

Atmospheric Neutrinos

The story of missing neutrinos was repeated when we looked elsewhere. One key clue came from neutrinos created in the upper atmosphere. Cosmic rays, mostly in the form of protons or helium nuclei, are constantly bombarding the Earth. When they hit the atmosphere they create a constant stream of $\pi^\pm$ pions. These pions decay to muons

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \quad \text{and} \quad \pi^- \longrightarrow \mu^- + \bar{\nu}_\mu$$

and the muons then quickly decay to electrons,

$$\mu^+ \longrightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad \text{and} \quad \mu^- \longrightarrow e^- + \bar{\nu}_e + \nu_\mu$$

Indeed, as we saw in Interlude B, this is how both the pion and muon were first discovered. Now, however, our interest lies in the neutrino by-products. These "atmospheric

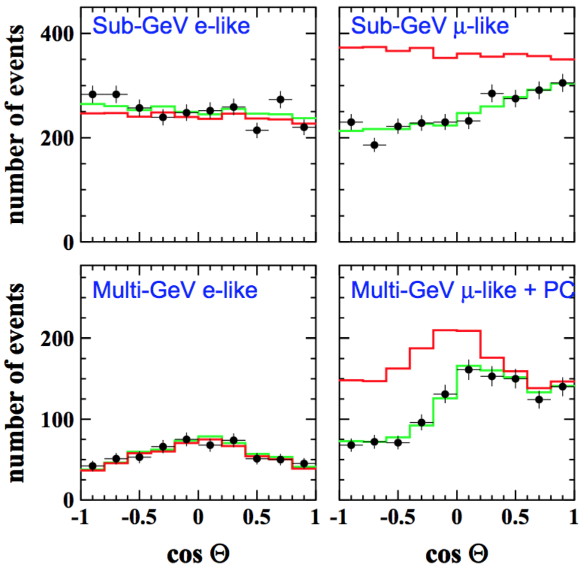


Figure 49. The observed flux of electron neutrinos (on the left) and muon neutrinos (on the right). The top boxes show low-energy neutrinos; the lower boxes high-energy neutrinos. The red line is the theoretical expectation without neutrino oscillations, and the black boxes the data.

neutrinos" have significantly higher energies than solar neutrinos; often around a GeV or higher. Given the decay processes described above, each collision should result in two muon neutrinos (strictly one ν_μ, one $\bar{\nu}_\mu$) for every electron neutrino. The question is: can we find them?

The answer, given by Super-Kamiokande, is interesting and shown in Figure 49. These show plots of the neutrino flux (on the vertical axis) against the angle at which the neutrinos come into the detector (on the horizontal axis). An angle $\cos\theta = 1$, on the far right, means that the neutrinos come directly down. An angle $\cos\theta = -1$, on the far left, means that neutrinos come up, through the Earth.

The data on the left two boxes is for electron neutrinos, both for low-energy events (shown in the top box) and high-energy events (in the bottom box). The red line is the theoretical expectation; the black dots the observed flux. We see that the agreement between experiment and theory works well.

The story is more interesting for muon neutrinos, shown in the two boxes on the right. The number of neutrinos coming straight down agrees perfectly with what we expect, but there's a clear deficit for those that come up through the Earth. Why?

For any other particle, you might think that the Earth is simply getting in the way. But neutrinos pass right through the Earth without any difficulty. (Remember the picture of the Sun at night in Figure 39.) Besides: theorists aren't stupid and had taken the presence of the Earth into account when computing the red line! Instead, the key point is that the muon neutrinos have travelled further, and so had more opportunity to convert into other neutrinos, in this case tau.

Importantly, the atmospheric neutrinos clearly show us that neutrino oscillations depend on the length L that neutrinos travel. We have

$$\begin{array}{rcl} \text{Straight down: } L = 15 \text{ km} & \Rightarrow & \text{No oscillations} \\ \text{Straight up: } L = 13000 \text{ km} & \Rightarrow & \nu_e \text{ unaffected, but } \nu_\mu \to \nu_\tau \end{array}$$

Meanwhile, for solar neutrinos we have $L = 150$ million km. As we'll now see, this collection of data goes a long way to allowing us to determine neutrino masses.

Getting a Handle on the Masses

Nature is kind, and gives us two sources of neutrinos: solar and atmospheric. As we've seen, these clearly show that something fishy is going on, as neutrinos appear and disappear. But to be sure, we should design our own experiments here on Earth, in which a source of neutrinos is created and subsequently detected elsewhere.

There are now a number of these experiments up and running, with results consistent with the oscillations seen in solar and atmospheric neutrinos. The best results so far have come from the T2K experiment, which directs a beam of muon neutrinos created in Tokai, Japan to the Super-Kamiokande detector located 300 km away. They clearly see ν_e neutrinos in a beam that, at its origin, consisted purely of ν_μ neutrinos. Meanwhile, the OPERA experiment in the Gran Sasso lab in Italy has directly detected ν_τ neutrinos in a ν_μ neutrino beam from CERN, 730 km away.

From this collection of experiments, together with results from solar and atmospheric neutrinos, we can put together the picture of masses and mixing angles that we described in Section 4.4.2.