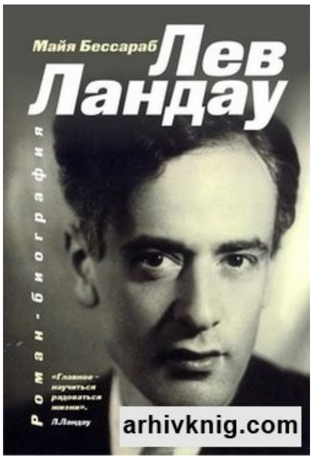
Secret Symmetry and the Higgs

What do a snowflake, a bar magnet, a superconductor and the state we are in today have in common? Not much, you might imagine. However, there is a hidden connection.

Ice is so familiar to us that we just accept the sudden and dramatic metamorphosis of water as it cools below its freezing point. A liquid that we can dive into is spontaneously transformed into a hard rock-like material that we would crack our skull on. John Locke used the following anecdote to express just how surprising this transformation really is:

"A Dutch ambassador, who entertaining the king of Siam with the particularities of Holland, which he was inquisitive after, amongst other things told him that the water in his country would sometimes, in cold weather, be so hard that men walked upon it, and that it would bear an elephant, if he were there. To which the king replied, Hitherto I have believed the strange things you have told me, because I look upon you as a sober fair man, but now I am sure you lie."[i]

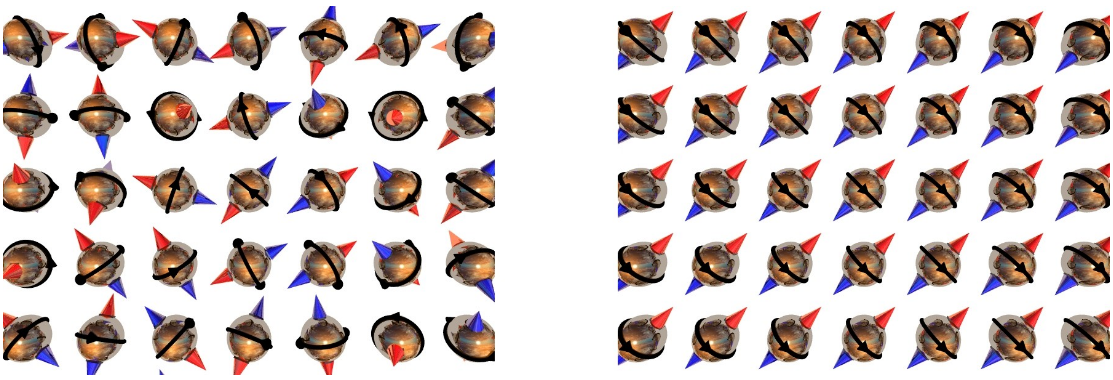
The miraculous transformation that turns liquid water into ice is known to physicists as a phase transition. Many other phase changes are known to physics where, at a well-defined temperature, the structure of a material changes completely. These include: the condensation of a gas into liquid, the transformation of liquid helium into a superfluid and the transition of a normal conductor into a superconductor, which happens in a range of materials that lose all electrical resistance at low temperatures. There are many others. We will encounter some of them below. On the face of it, these transformations all appear very different, but in the 1930s the Russian theoretical physicist Lev Landau recognised that they share common features. He realised that in each case, the transformation of the material is accompanied by a loss of symmetry, and this insight was the first step towards building a mathematical model of phase transitions. Landau's theory represents one of the most important insights in the whole of 20th-century physics. It is remarkable because, as we will see, his reasoning was very general and independent of the details of any particular physical system. For this reason, it applies to a whole range of different areas of physics. There can be few occasions where pure reason has led to so great a pay-off.



[Lev Landau]

Magnetism and Mexican Hats

Ferromagnets, such as iron, nickel and cobalt, undergo a phase transition at a temperature known as the Curie temperature after the French scientist Pierre Curie. The Curie temperature of iron is 1043K. Above this temperature a lump of iron will lose its magnetism, while below this temperature the iron will spontaneously magnetize.



[Schematic depiction of the atoms in a ferromagnet. Left: The symmetrical phase above the Curie temperature where all the magnetic fields of the atoms are randomly oriented. Right: The asymmetrical phase below the Curie temperature where the magnetic fields of all the atoms are aligned in the same direction.]

Landau's theory can be illustrated with a simplified toy model of a magnet. Imagine a collection of atoms arranged regularly in a plane, each atom has a small magnetic field that is constrained to point in one of two directions; it can either point upwards or downwards. Both directions are completely equivalent, but the energy of two neighbouring atoms is lowered slightly if their magnetic fields point in the same direction, so there is a tendency for them to align. At high temperatures thermal vibrations vigorously jostle the atoms around with the effect that their magnetic fields are continually being flipped. So, although there might be a tendency for neighbouring atoms to align their magnetic fields, any such alignments are rapidly disrupted. In this state the material will not exhibit a magnetic field macroscopically, as the randomly oriented magnetic fields of all the atoms cancel each other out. However, if the temperature is lowered, a point will eventually be reached where the thermal vibrations are no longer strong enough to disrupt the alignments and suddenly all the magnetic fields will spontaneously align in the same direction. The material will now have a macroscopic magnetic field. It has become a magnet.

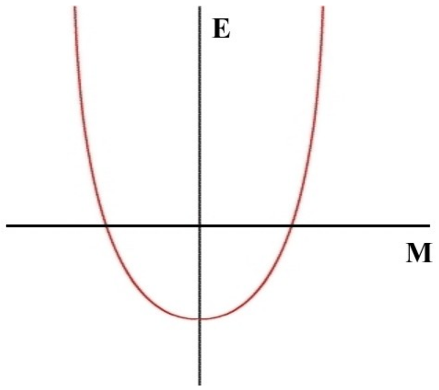
How should this be interpreted in terms of symmetry? Above the critical temperature the direction of the magnetic fields of the atoms are arranged completely at random. The magnetic field of each atom is equally likely to be pointing in either direction, so the system does not distinguish between the two directions, they are completely equivalent. In this state the system is symmetrical with respect to the up and down directions. But below the critical temperature this symmetry has disappeared as the magnetic fields of all the atoms are now pointing in the same direction. Furthermore, the loss of symmetry below the phase transition is equivalent to a gain in order, as illustrated in this toy model, and this was the key to Landau's analysis. He suggested that there must be a parameter that measures the order. Above the transition temperature this parameter must be zero – as the system is completely disordered and symmetrical. Below the transition temperature the order parameter must take some non-zero value that measures the order that has spontaneously arisen. In the case of a magnet, the order parameter is called the magnetization. Above the critical temperature the magnetization is zero, below the critical temperature it is non-zero.

The next step in Landau's analysis was to write out an expression for the energy of the system near the critical temperature in terms of the magnetization[ii] M. Close to the critical temperature M must be small, as it is zero above the critical temperature, so Landau assumed that the energy could be represented as a polynomial expansion in M. Now, our toy model is completely symmetrical with respect to the up and down directions of the magnetic fields, so the energy cannot depend directly on M, because, if it did, an upwards magnetic field would contribute +M to the energy and a downwards magnetic field would contribute –M to the energy, which would contradict the assumption that the system was completely symmetrical with respect to the two directions. In other words the expression for the energy must respect the symmetry of the system. It follows that the energy cannot depend on any odd power of M. The polynomial must therefore only contain even powers of M. Furthermore, all powers of M higher than the fourth power can be disregarded, as they are insignificant when M is small. This leaves just two terms in the expansion, terms proportional to M^2 and M^4. The result is the following expression for the energy in terms of the magnetization at temperatures close to the phase transition:

$$E = \alpha M^2 + \beta M^4$$

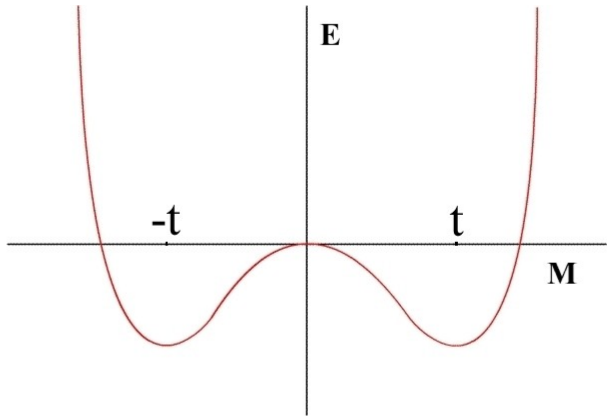
A macroscopic magnetic system, such as the one described by our toy model, will rapidly tend to thermal equilibrium, at which point the system will be in its lowest energy state. So this expression is very important as, in principle, it determines the lowest energy configuration of the system in terms of the magnetization. However, it contains two unknown terms α and β. Now Landau reasoned that β must be positive, because otherwise there would be no minimum energy and this clearly could not represent a physical system. This implied that there were just two possibilities for the shape of the energy graph.

If α were positive, then the graph would look like this:



The energy has a single minimum where M is equal to zero.

But if α were negative then the graph would look like this:

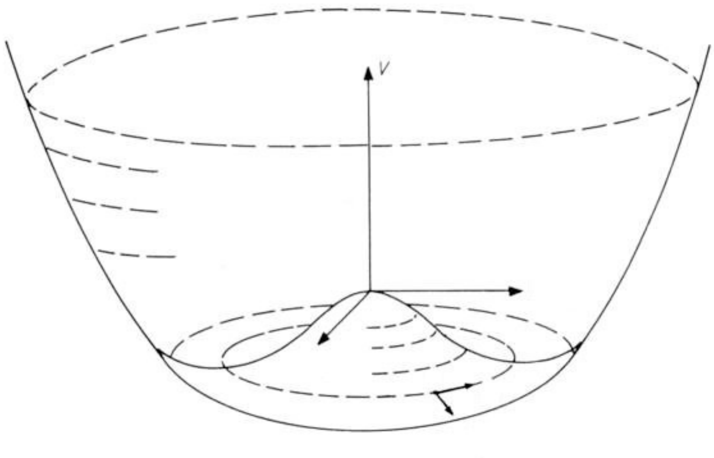


The energy graph now has two minima.[iii]

Landau reasoned that the first graph must represent the energy just above the critical temperature and the second graph must represent the energy just below the critical temperature. In other words, α is not a constant, it must vary with temperature. Above the critical temperature α is positive, below the critical temperature α is negative and at the critical temperature α is exactly equal to zero.

Landau's description of the phase transition was that above the critical temperature the system is in its lowest energy state, which is the minimum of the first energy graph. At this point the magnetization is zero, which represents a state in which the magnetic fields are oriented at random with respect to the up-down axis. However, as the temperature falls the shape of the graph changes and below the critical temperature there are two minima, and the system must fall into one of these minima. They are both equivalent, so the choice is made completely arbitrarily by random fluctuations within the system as the temperature falls below the critical temperature. In one minimum all the atomic magnetic fields will be pointing upwards, in the other minimum they will all be pointing downwards. In either of these states the system is now ordered with all the magnetic fields aligned and the up-down symmetry is lost, or at least hidden.

Landau's expression relating the energy to the order parameter is applicable to a whole range of phase transitions and the energy graphs generalize in a natural way to higher dimensional systems. For instance, if the magnetic fields were free to point in any direction in a plane rather than being constrained to lie along a single axis, then below the critical temperature the energy graph would look like the illustration below. This is known as a Mexican hat diagram. The energy minima now form a circle within the brim of the hat. At each point on this circle the energy of the magnetic system is as low as possible.



[Energy graph – Mexican hat]

Superfluid Helium

At atmospheric pressure the boiling point of helium is 4.2K, that is just 4.2 degrees above absolute zero, which is lower than the boiling point of any other substance. In the early years of the 20th century a number of laboratories were engaged in a race to liquefy helium. The Dutch physicist Heike Kamerlinghe Onnes was the eventual winner in 1908. Five years later he was awarded the Nobel Prize in Physics for his low temperature research. But no-one realised until the 1930s that if liquid helium is cooled further it undergoes a much more dramatic transformation. Below a critical temperature of 2.17K helium is transformed into a superfluid – a liquid without any resistance to its flow. A container that holds normal liquid helium perfectly well will suddenly spring numerous leaks when it is cooled below the critical temperature, as superfluid helium seeps out through ultramicroscopic pores in the container. Superfluid helium has many strange and wonderful properties. A video clip in which some of these effects are demonstrated is available at the following link:

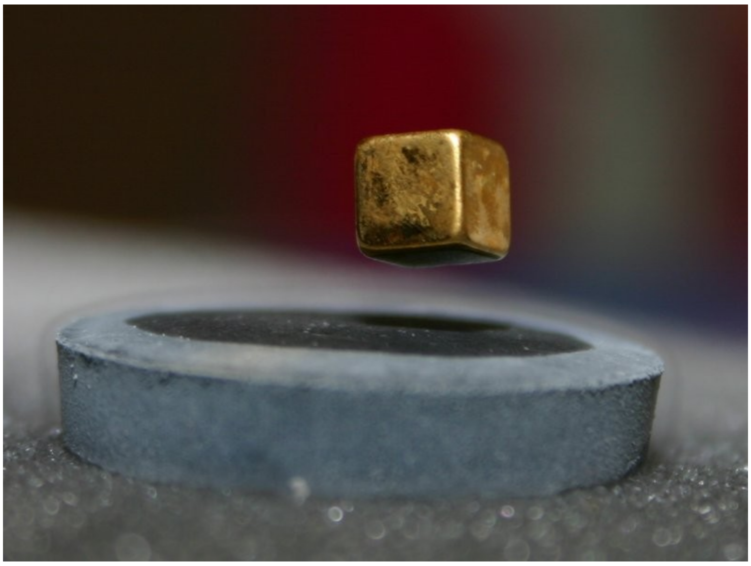
[http://www.alfredleitner.com/superfluid.html]

Landau successfully adapted his theory to the superfluid helium phase transition and in 1962 he was awarded the Nobel Prize for Physics for this research.

Superconductivity

Another remarkable low temperature phenomenon – superconductivity – was discovered in the laboratory of Kamerlingh Onnes in 1911. At a temperature of 4.2K it was found that mercury loses all electrical resistance (mercury is of course a solid metal at these very low temperatures). Superconductivity was soon shown to be common to other materials and not just a strange property of mercury. There is a characteristic temperature at which each of these materials is transformed from a normal conductor into a superconductor. For instance, niobium becomes a superconductor below 9.2K, whereas niobium-tin alloy has a critical temperature of 18K. In more recent years various substances have been discovered with much higher superconducting temperatures.

Another characteristic of superconductors is that they will expel a magnetic field. This is known as the Meissner Effect. A superconducting material when cooled below its transition temperature will levitate above a magnet, as shown in the illustration below.



[The Meissner Effect]

The barrier to the magnetic field is not abrupt at the edge of the superconductor; the magnetic field penetrates a short distance into the superconductor. Within this skin superconducting electric currents circulate and produce a magnetic field that completely cancels out the external magnetic field, leaving the innards of the superconductor completely field free.[iv]

In 1950, Landau and his colleague Vitaly Ginzburg applied Landau's theory to the superconducting transition, once again with great success. One of the outstanding triumphs of the Ginzburg-Landau theory is that it explains why there are two fundamentally different categories of superconductor. Most pure metals are Type I superconductors and they lose their superconductivity when placed in quite a weak magnetic field. But Type II superconductors remain superconducting even in very strong magnetic fields. This means that they are suitable for use as electromagnets. As there is no resistance to the electric currents within a superconductor, it is very energy efficient to produce a strong magnetic field in this way. Superconducting magnets have become a very important technology. There are now around 10,000 MRI scanners in hospitals around the world, most of which utilise superconducting magnets.

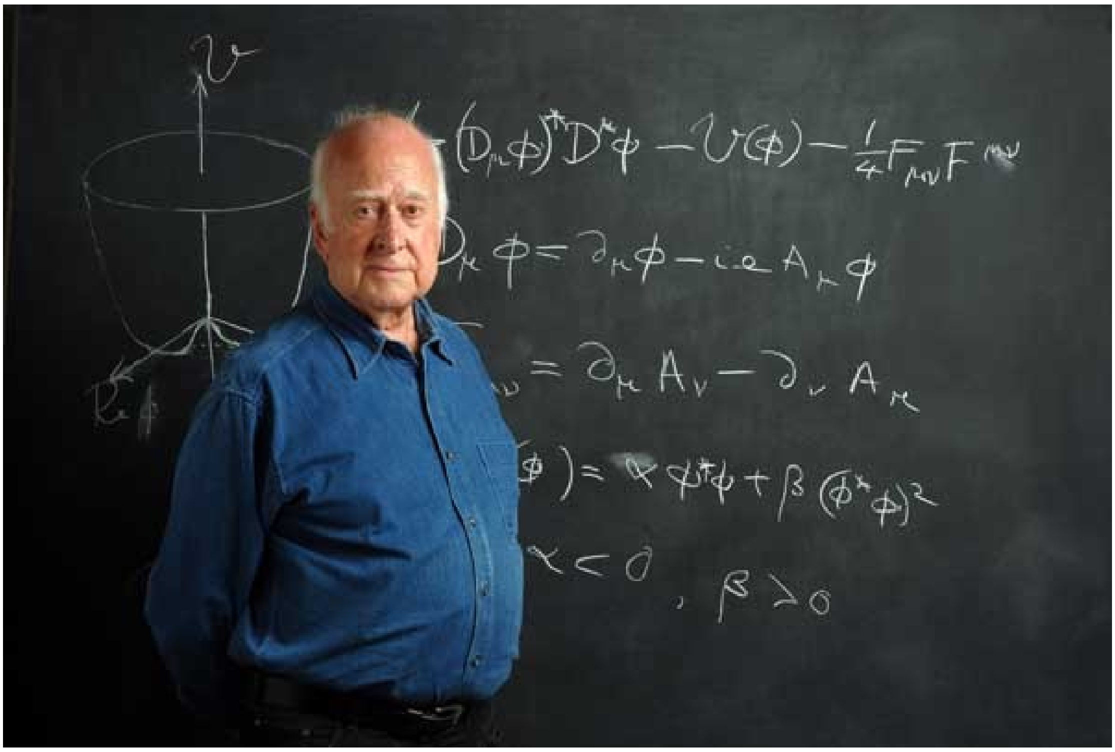
A few years after the publication of the Ginzburg-Landau theory of superconductors three Americans – John Bardeen, Leon Cooper and Robert Schrieffer (collectively known as BCS) – published an alternative theory of superconductivity that was received with much greater enthusiasm in the West. Whereas the Ginzburg-Landau theory is a top-down theory based on very general physical principles, the BCS theory is a bottom-up theory built around particle interactions. The BCS theory explains superconductivity as the result of a counterintuitive tendency of electrons to pair up at low temperatures. These Cooper pairs, as they are called, are bound together by the exchange of pulses of lattice vibrations that are passed back and forth between the two electrons. Although this binding is very weak, at very low temperatures it is sufficient to prevent the pairs from being shaken apart by the thermal vibrations of the atoms in the material. The result is that they can flow through the superconductor without any resistance.

Although the BCS theory proved to be a very powerful tool for explaining superconductor experiments, initially there was some doubt about whether it was an acceptable theory, because it implied that the symmetry of the electromagnetic force was broken within a superconductor. This was considered to be a disastrous result because the symmetry is an essential part of the theory of electromagnetism as it guarantees that electric charge will be conserved in all interactions. Without this symmetry it seemed that the theory would not make mathematical sense and would be in serious conflict with experiment. However, the relationship to symmetry breaking was clarified when the Russian physicist Lev Gor'kov proved that although the Ginzburg-Landau theory and the BCS theory look very different, the two theories are completely equivalent at temperatures close to the critical temperature. (Among other things, Gor'kov showed that Landau's order parameter measures the density of the Cooper pairs in the superconductor.) It turns out that the symmetry of the electromagnetic force really is spontaneously broken within a superconductor and this is why a magnetic field cannot penetrate deep within a superconductor. The electromagnetic force is mediated by the exchange of massless particles of light known as photons. Within a superconductor, the Cooper pairs interact with these photons and this interaction gives the photons a mass and thereby limits the range of the electromagnetic force to a distance equivalent to the penetration depth. This loss of symmetry could be explained in terms of the Ginzburg-Landau theory. Below the critical temperature, it was hidden but not destroyed so electric charge was still conserved and the theory of electromagnetism still made sense. The magic of symmetry breaking as evinced in a superconductor would inspire the most spectacular of all applications of Landau's theory in particle physics.

The Electroweak Theory

Over the last 400 years physicists have steadily revealed the connections between all the many and varied forces that control the motion of the bodies around us and give structure to the matter from which our universe is formed. By the middle of the 20th century all known phenomena could be accounted for in terms of just four forces: gravity and electromagnetism, plus two forces that play important roles in nuclear physics – the weak force and the strong force. The strong force holds quarks together to form protons and neutrons and binds these protons and neutrons together to form the nuclei of atoms, so it is fundamental to the structure of matter. The weak force is also extremely important, as it is able to transform protons into neutrons and neutrons into protons and thereby transmute the elements. The weak force is vital for the synthesis of all the atoms heavier than hydrogen within very massive stars. These atoms are dispersed into space in the supernovae explosions that end the lives of these supergiant stars. They may then be incorporated into the material that forms later generations of star systems such as our own.

Electromagnetism and the weak force have completely different characteristics. The electromagnetic force holds atoms together with negatively charged electrons bound to a positively charged nucleus. But electromagnetism also operates over very long distances. For instance, the Earth's magnetic field will attract the tip of a compass needle wherever the compass is located on Earth. By contrast, the weak force only operates over a very short distance – approximately, one thousandth of the diameter of a proton and this is related to the fact that it is an extremely feeble force. In the 1950s there were early attempts to construct a theory that united both forces within a single model. But this would require the weak force to be mediated by very massive particles, while electromagnetism was mediated by the massless photon. Unfortunately, the inclusion of these massive particles appeared to completely destroy the usefulness of the theory. But Peter Higgs recognised that what was required was a symmetry breaking mechanism that hid some of the symmetry, whilst retaining all the nice features that symmetry brought to these theories. Higgs realised that Landau's model would fit the bill nicely.



[Peter Higgs – notice the Mexican hat diagram on the left of the blackboard. (Copyright Peter Tuffy, The University of Edinburgh)]

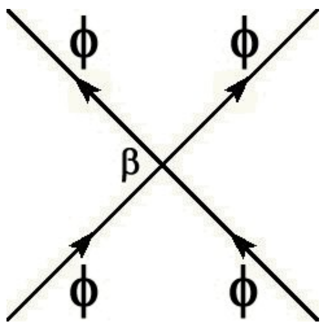
In 1964, Higgs translated Landau's ideas into the language of particle physics and constructed a model that demonstrated how, in principle, the symmetry of a force such as electromagnetism or the weak force could be spontaneously broken. This would be the basis for a unified theory of the electromagnetic and weak forces. They could now be considered as two manifestations of a single electroweak force, thus reducing the number of independent forces from four to three. This would be the greatest unification of the forces since the unification of the electric and magnetic forces by Faraday and Maxwell in the 19th century. Higgs did not attempt to construct such a theory himself. Steven Weinberg took the next step and applied Higgs' idea to an electroweak model that had earlier been studied by Sheldon Glashow. This theory is known as the Glashow-Weinberg-Salam (GWS) Model.

The Higgs Mechanism

So how does the model constructed by Peter Higgs work? In quantum field theory each species of fundamental particle is associated with a quantum field that permeates the whole of space. All these fields fluctuate at each point in space and the excitations of the various fields are what we interpret as particles. In the model constructed by Higgs there is a field ϕ that plays the role of the order parameter in Landau's theory. The potential energy of the ϕ field is:

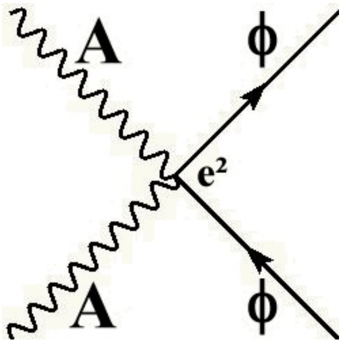
$$V = \alpha\phi^2 + \beta\phi^4$$

This is exactly the same form as the energy in Landau's theory. (You might notice that this is what is written on the blackboard behind Peter Higgs in the photograph above. The only difference is that ϕ is expressed as a complex number so ϕ^2 is written as $\phi^*\phi$.) In quantum field theory β is interpreted as the strength with which the ϕ particles interact with each other. Diagrammatically these interactions are represented as follows:



The direction of time is upwards in the diagram. At the bottom of the diagram two ϕ particles approach each other, at the centre they interact with strength β, then at the top they recede from each other again. The quadratic term $\alpha\phi^2$ is the energy cost of producing a ϕ particle. In quantum field theory, α is interpreted as the mass of the ϕ particles.

In addition to the Higgs field ϕ, the model includes a field A corresponding to a force mediating particle that would behave like a photon. The interaction between the A particles and the ϕ particles produces an additional contribution to the energy of the form $e^2A^2\phi^2$, where e is the electric charge of the ϕ particle. It is represented in the diagram below:



Time runs upwards in this diagram. It represents a process in which an A particle and a ϕ particle approach each other, interact with strength e^2 and then recede from each other again.

Higgs proposed that immediately after the Big Bang the universe went through a dramatic phase transition. At extremely high temperatures in the very early universe, the parameter α would be positive. The lowest energy configuration would then be one in which the ϕ field was zero. But, in accordance with Landau's theory, as the temperature dropped the value of α decreased until below a critical temperature of perhaps a trillion degrees it became negative. The energy graph would then look like the Mexican hat diagram. Below the critical temperature the lowest energy configurations of the ϕ field form a circle within the brim of the hat. The ϕ field now loses energy and the system falls randomly to one of the points on the minimum energy circle. ϕ now takes a non-zero value that minimizes the energy. We will give the label t to this value of ϕ. The result is that the ϕ field is no longer zero even in completely empty space, just as the magnetic field in a permanent magnet is not zero. Due to the phase transition the symmetry of the system has been spontaneously broken.[v]

The quantum field will oscillate around the minimum energy field. To make this explicit, we need to define a new field H that is zero at the energy minimum and examine the oscillations and excitations of this field. This means that ϕ must be rewritten as:

$$\phi = t + H$$

where t is the value of the constant background ϕ field that permeates empty space, and H is the Higgs field. Expanding Landau's expression for the energy in the ϕ field in terms of this new field will explicitly show the effect of the symmetry breaking. The result is a constant term, which particle physicists usually disregard as it simply shifts the energy scale, then there is a quadratic term and a quartic term just as there was for the original ϕ field, but now there is also a term proportional to H^3, so the energy is not symmetrical under H → -H. This symmetry has been lost in the new low energy configuration.

But more important is the effect that this symmetry breaking has on the photon-like A particle. The interaction between the ϕ particle and the A particle is represented by the term $e^2A^2\phi^2$. Rewriting this in terms of the Higgs field gives $e^2A^2(t + H)^2$, which expands to

$$e^2t^2A^2 + 2e^2tHA^2 + e^2H^2A^2 .$$

The second and third terms represent interactions between the Higgs particle and the A particle, but the first term is the one that we are most interested in. It represents the interaction between the A particle and the background ϕ field whose value we have labelled t. It is a quadratic term in A multiplied by a constant e^2t^2. This represents a mass term for the A particle. In the symmetric phase at temperatures above the phase transition the force carrying particle A had no mass, but due to its interactions with the background ϕ field it now has a mass equal to e^2t^2. This is the magic of the Higgs mechanism. The symmetry of the force has been spontaneously broken and the photon-like particle that mediates the force has been transformed into a massive particle. This is exactly what happens within a superconductor, but here the Higgs particle is playing the role of the Cooper pairs.

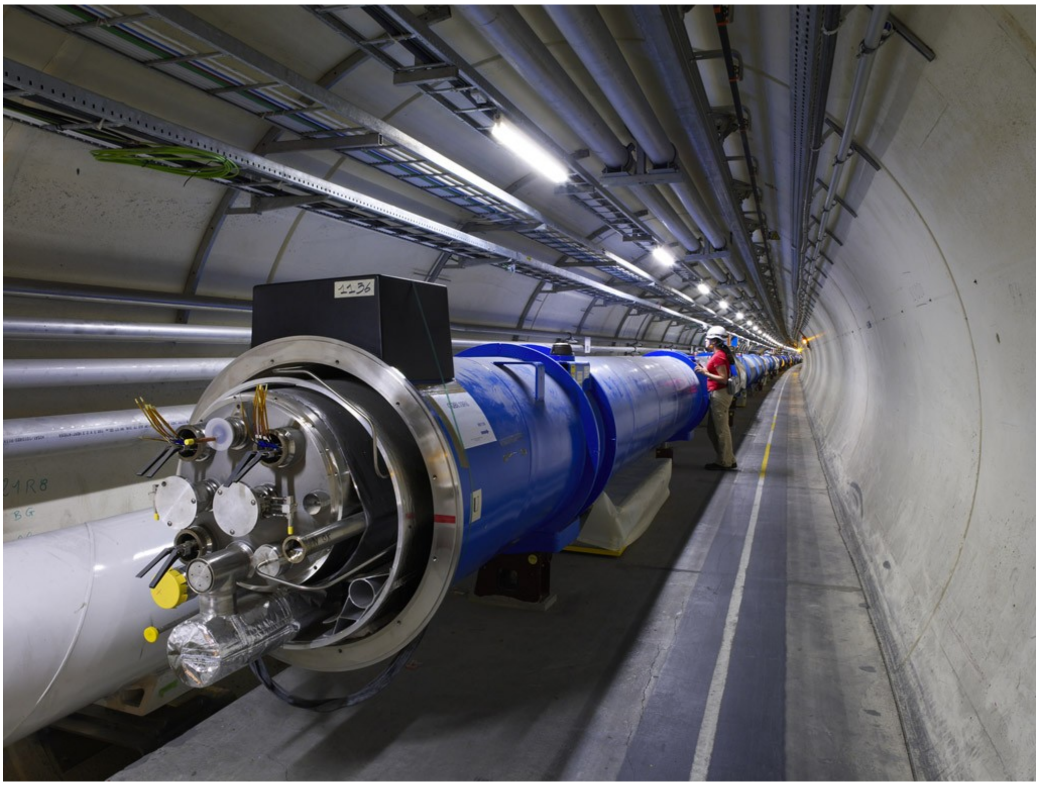
The State we are in Today

The Higgs mechanism works in the same way in the full GWS electroweak theory. In the full theory there are four force carrying particles plus the Higgs particle. The Higgs particle does not carry an electric charge, but it does carry the weak charges. For this reason it interacts with the weak force carriers, but not the photon and so the weak force is broken, but electromagnetism remains unbroken. Above the critical temperature the force carrying particles are massless, but below this temperature only one force carrier – the photon remains massless, while the other three – the W-plus (W^+), W-minus (W^-) and the Z-nought (Z^o) become massive due to their interactions with the Higgs field. The result is that the electromagnetic force mediated by the photon is a powerful long-range force, whereas the weak force mediated by the other three particles is a feeble and extremely short-range force.

What this means physically is that since the earliest moments of the universe the Higgs field has had a non-zero value everywhere, even throughout empty space. (This is similar to the non-zero magnetic field in a permanent magnet.) As the W and Z particles move through this background field they are tugged back, almost as though they were wading through treacle, and this is what gives them their large mass. The Higgs also interacts with matter particles such as the electron and the various flavours of quark and the effect on them is similar. According to the GWS model this is where the mass of each of these particles arises from. The Higgs particle plays the same role in empty space that the Cooper pairs play in a superconductor. In the words of the Nobel Prize winning physicist Frank Wilczek: “We are living within a cosmic superconductor.”

But is the theory true? Does it represent the way that the universe really works? Many of the predictions of the electroweak theory have now been verified in particle accelerators. The most spectacular of which was the discovery of the massive W-plus, W-minus and Z-nought particles at CERN in 1983. The only prediction that remains to be confirmed is the existence of the Higgs particle itself. The Large Hadron Collider, CERN’s phenomenal new machine, is now colliding ultra-high energy beams of protons in the quest for the Higgs. Superconductivity is playing a remarkable dual role in this effort. Not only is it the theoretical inspiration behind the theory of the Higgs particle, but it is also responsible for the cutting-edge technology of the machine. The ultra-high energy protons are guided around the 27km circuit of the collider by the intense magnetic fields of niobium-titanium superconducting magnets cooled in a bath of superfluid helium.

Most physicists expect that the Higgs particle will make its first appearance at the LHC in the very near future.



[The tunnel of the Large Hadron Collider during construction. The sealed ends of the two beam pipes can be seen side-by-side projecting from the end of the pipeline. (A worker in the LHC tunnel. Copyright CERN Geneva)]

[i] John Locke – An Essay Concerning Human Understanding, Of Probability, Book IV Chapter XV

[ii] The magnetization can be defined as follows:

$$M = (n_+ - n_-)/(n_+ + n_-)$$

where n_+ is the number of atoms whose magnetic fields are pointing upwards and n_- is the number of atoms whose magnetic fields are pointing downwards. $(n_+ + n_-)$ is therefore equal to the total number of atoms. Above the critical temperature, $n_+ = n_-$, so $M = 0$. Well below the critical temperature, either
$n_- = 0$ and $M = 1$, or $n_+ = 0$ and $M = -1$.

[iii] In the illustration the values of M at which the energy is a minimum have been labelled –t and t.
If you are familiar with the differential calculus, it is easy to check the form of these two graphs. The turning points of the graphs are the points at which the derivative of E with respect to M is equal to zero.

$$dE/dM = 2\alpha M + 4\beta M^3 = 0$$

If both α and β are positive, then there is just a single minimum at $M = 0$. However, if α is negative and β is positive, there are three turning points at the points $M = 0$, $M = -(|\alpha|/2\beta)^{1/2}$ and $M = (|\alpha|/2\beta)^{1/2}$. The turning point at $M = 0$ is a maximum and the other two turning points are minima.

[iv] The depth of the skin is known as the penetration depth of the superconductor.

[v] t is the value of the field ϕ at the energy minimum, so it is equal to $(|\alpha|/2\beta)^{1/2}$.