Quantum Physics in words

What modern physics says about the nature of reality

John O'Neall

Table of Contents

1. The language of physics – observing and understanding....1
2. Transformations and constraints of Special Relativity....4
   2.1. Some prerequisite – systems, reference frames and all that....4
   2.2. Effects of Special Relativity....4
3. Fields: electric, magnetic and gauge....6
4. Operators in Quantum Mechanics (QM)....8
5. Relativity and symmetry in QM: spin....11
6. On to QFT....13
   6.1. Second quantization – many particles....13
   6.2. Interactions – local transformations and gauge invariance....14
   6.3. How about mass? The Higgs mechanism....17
   6.4. Path integral method and Feynman diagrams....18
7. Summing up....19
8. Equations of quantum electrodynamics (QED)....19

Being one more attempt to convey in words how modern physics explains the Universe. The language of physics is mathematics, so rather than invent analogies, I will try to explain it from the math, but using words, not mathematical symbols or equations.

Here, to start out, is the result:

> "Every particle and every wave in the Universe is simply an excitation of a quantum field that is defined over all space and time."[1]

If you are already cool with quantum mechanics, special relativity and gauge fields, you can skip to section 6.

1. The language of physics – observing and understanding

Although mathematics is indeed the language of physics, you must know how to use it and how to interpret it. This document is concerned with the second of those two points.

First, two basic assumptions:

- Without going into ontology, we will assume that there is a reality "out there" to observe, that what we perceive with our senses (and instruments) really exists in some meaningful and useful way.
- Understanding it will only be possible if it is governed by universal natural laws[2] and so we also shall assume their existence. We speak of a ***law*** when we have determined that a given previous event

1. Lancaster and Blundell, 1.
2. This idea is as old as Epicurus or Lucretius.

(called the cause) in certain specific and well-defined conditions always gives rise to the same subsequent event (the result) across countless observations. The word “always” here means “every time we have observed it”.[3] So we consider that this will always be true in the future too. The Sun will indeed appear to rise tomorrow morning, although probably not in 5 billion years. If you disagree, stop reading now.

Just as mathematics starts with some unprovable definitions or axioms, so without thinking about it do we employ other metaphysical, because unprovable, preconceptions (like the existence of a dimensionless point). As math builds a structure on these axioms, so do we build one on these assumptions plus observed, ***empirical*** facts. The result is a representation, or model, of nature which we call a ***physical theory***.

With or without special equipment, observation proceeds when our sense organs are activated by any of many signals -- light or sound waves, chemical substances, molecules of all sorts or the touch of larger objects. We do not really perceive what is “out there” because only a very limited subset of this raw date is able to trigger our sense organs. These then send electro-chemical impulses into our nervous system, where our brains massage and organize them into meaningful and useful forms or patterns. In this way, observation leads to perception. Those entities we apprehend are called **observables**, a concept of great significance in QM. An electron’s spin, although hidden inside another space, can be an (indirect) observable, but a gauge field, in spite of its importance, cannot. (Both these subjects will be described later, so stay tuned.) The boundary between what we can directly observe and what we infer from abstract mathematical models is often blurred and difficult to distinguish.

Current cognitive theories think that this data analysis is done by continually executing cognitive processes which compare the incoming data to structures already in the brain. The brain then uses the comparison in order to improve the perceived event image (à la Bayes). According to this model, understanding depends partly on information already in the brain and this may be inexact or even wrong. (Think of optical illusions.) So our perception, our handling of sensory data, depends on the state of the brain previous to the incoming data – to what the brain expects to observe. This is obviously a source of error. Kant could not have agreed more.

A million or so years ago, “useful” data was that which helped our remote ancestors to survive. As we evolved through natural selection, our cognitive processes matured and developed along with our understanding of the world around and within us. In order to express and communicate our thoughts, we started using language, spoken then written. Language helped us cooperate -- or not -- not only to defend ourselves but to come up with new ideas for survival -- or simply to make life more pleasant.

Our languages are tool kits which allow expression of some ideas better than others. Our ancestors needed words like tree or tiger, they did not need gene or quark. Today we do. As we have discovered new phenomena which our former vocabulary was not able to describe, we have invented new words and concepts, often making use of older concepts by extending their meaning or putting them together in new ways. In physics, we have been obliged to consider quantum objects as behaving like waves and particles at once. Would it not be better to have a new term for something which is neither – or both? If that word were, say, wavicle, would that aid in our description of nature, once we assimilated the word the way we have assimilated the words wave and particle?

On the other hand, there is another point of view based on fields alone as the stuff of nature and those are the subject of this document. Then the problem of wave versus particle is no longer relevant.

In any case, language is a cognitive function which is therefore limited by the structure and functioning of our brains.

Just as languages are imperfect but evolving means of expression, our mathematics evolves too.[4] We build on one mathematical theory (say, algebra or Euclidean geometry) until it is no longer adequate and we are forced to find a new one (calculus or Riemannian geometry of curved space). So our use of math evolves along with our concepts and paradigms. Just compare Newton’s equations to Einstein’s or those of quantum mechanics. (Don’t worry about understanding them, just admire.)

Newton (force): $$F = ma \tag{1}$$

3. This second assumption comes in fact from Hume, among others, so we’re in good company.
4. Math is perfect. Just ask a mathematician.

Newton (gravitational force):	$F = G\frac{mM}{R^2}$	(2)
Einstein (gravity):	$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -\kappa T_{\mu\nu}$	(3)
Schrödinger (QM):	$i\hbar\frac{\partial\Psi(x)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(x) + V(x)\Psi(x)$	(4)
Dirac (QFT):	$(i\gamma^\mu\delta_\mu - m)\vert\psi\rangle = 0$	(5)

Dirac's equation may look simpler than Schrödinger's, but trust me, it's not. The first just uses multiplication (m times a)[5]; the second, division; the third, tensors (that R thing with subscripts); the fourth, derivatives; and the fifth the whole lot, even if that's not obvious from the compact way it is written. This evolution suggests that one day our equations may get still better (and maybe more complicated) and approach even closer to describing reality precisely, i.e., become better approximations to a description of nature. An essential feature of the later equations (Einstein, Schrödinger, Dirac) is that when they are applied to the situations Newton studied, they approximate to Newton's equations. That is a Good Thing.

If the math of physics has become quite hairy, as shown in the above equations, understanding what it means in terms of observable, intuitive physical phenomena also has become more and more difficult. (Remember the wavicle?) Some of us may have just reached the point where we can almost imagine everything as made up of tiny particles separated by relatively vast stretches of empty space. But now, physics tells us that everything is made up of fields. This is the idea behind ***quantum field theory*** (***QFT***). Fields are where the buck stops, at least for the moment.[6] Like the bottom turtle, they are not made up of anything else (well, as far as we currently know).[7] The particles which we see as the constituents of all the stuff around us are vibrations in quantum fields, fermion fields for matter, boson fields for forces. If you have trouble imaging a proton field interacting with an electron field through a vector boson field, you are not alone. It's much easier to imagine them as particles, so that's what we do. Is that because our brains are built to comprehend particles better than fields? Or just habit? Or the problem of representing our ideas with language?[8] Do we need a new vocabulary? (Probably.) Or do those preexisting "hard-wired" structures in our brains, proposed by some philosophers as already mentioned, hinder our understanding of new phenomena? Should we just give up and leave the scientists with their equations?

In any case, our understanding of nature is constantly evolving. We labored for a long time with Newton's relatively simple equations, then added Maxwell's more complicated ones, then Schrödinger's and Einstein's. Since then, we have added math evolved by Dirac, Weinberg, Salam, Higgs, Feynman and many others. The language is math, but – just like English or French – it's evolving in terms of what it expresses.

We must avoid going Platonic here. There is neither cave nor any reason to consider that we are looking at shadows of the Ideal, whatever that might mean. In spite of our initial assumptions, we don't really know what's out there. Which doesn't keep us from trying to cope with it by proposing models. Especially since our methods seem to work so well. The results are quite astounding in their accuracy and precision. But we must not forget that our perception of the Universe is limited both by our sense organs and by the structure of our brains.

In summary:

- We assume the existence of some reality of which our senses give us a meaningful although limited perception.
- We also assume the existence of universal natural laws which we study by empirical methods, confident that this procedure can lead to a better understanding of nature.
- We understand our empirical data by analysis using mathematical methods which evolve with our understanding.

---

5. I'm cheating some here, since the a for acceleration is the second derivative of position.
6. We are ignoring far-out ideas for which there is no evidence, like strings, branes and loops.
7. Some scientists think that space itself is made up of elements. This is the theory of loop quantum gravity.
8. Is language necessary for thought?

2. Transformations and constraints of Special Relativity

First, three concepts are fundamental in all physics – the notions of system, observables and reference frames.

2.1. Some prerequisite – systems, reference frames and all that

Physicists can't take on the whole Universe at once, so they choose for study a limited, isolated part of it they refer to as a ***system***. The system could be a hydrogen atom or a black hole or many other things. The essential requirement is to be able to consider it as independent of everything else in the Universe. It's an approximation, of course, but usually quite adequate. This method of taking on only a limited, isolated collection of objects is called ***reductionism***. Although it is decried in some quarters, it works pretty damn well.

It is also essential that the system include properties which we can measure, called ***observables***, otherwise why bother? How we handle the observables -- how we measure them, what values they might take on and how their manipulation affects the system -- are to some extent dependent on the physical theory. Classical mechanics, QM and QFT employ them differently, maybe as simple numeric parameters, maybe as operators (more later).

Thirdly, we can not simply "measure" things. Measure relative to what? A snail and a flying bird are going to see things differently. For a given experiment, we must define what coordinates of position and time we are using. One hundred kilometers east does not designate the same thing in New York or Paris. So we must always specify how we assign position, times, lengths and so on and this is the job of the ***reference frame***. It could be stationary relative to the surface of the Earth or to the center of the Milky Way galaxy, for instance. Example: In General Relativity (GR), we often refer to the rest frame of an object, the reference frame in which that object is stationary. The procedure is much simpler for a moving train than for a photon.

One of the most fruitful concepts physicists have discovered – and that is an understatement -- is the study and classification of ***symmetries***. We ask what we can change about a system without changing the system itself. Typical are such transformations as rotating it in space, shifting it in time or translating (moving in a straight line) it from one place to another. If the system itself remains unchanged under such ***transformations*** (physics speak), then it is said to be ***symmetric***, or ***invariant***, under the transformation. The essential thing is that its observable properties and its physical behavior not change.

Some examples. You can rotate a sphere, say a baseball through any angle in any direction and, if you ignore the seams, it still looks the same: This is spherical symmetry. You can rotate a candle through any angle around its lengthwise axis and it looks the same: This is obviously cylindrical symmetry. A cube is more special; you can rotate it through an angle of of 90° around an axis through the mid-points of two opposite sides and the result is indistinguishable from the original. A pancake must be either up or down, not in between.[9] Although we are talking about the system's looking the same, the important point is that this requires that the equations which define it be the same before and after, regardless of the values of the parameters which you plug into them. And that helps us with the math: Knowing that the system possesses a symmetry forces us to eliminate non-symmetric forms of the equations, thus (hopefully) simplifying the problem.

The examples just mentioned illustrate symmetry in our everyday world, but they can lead to misunderstandings in physics. In considering a symmetry such as a rotation in quantum physics, we do not mean to take a macroscopic object and rotate the whole thing about one point, such as its center of mass. On the contrary, we mean rotation of every point. As an analogy, consider a lawn: A quantum-physics rotation rotates each blade of grass in place; it does not upend the whole lawn.

2.2. Effects of Special Relativity

The transformations just mentioned are static, each happens once then is finished. But you can make dynamic (moving) transformations too. You can look at the system from a moving vehicle like a train. As long as the observation vehicle moves with a constant speed and direction relative to the system it's observing, or vice versa, the equations describing the system must look the same. Such a constant-speed, unaccelerated observation platform, or ***reference frame***, is called an ***inertial system***. The mathematical changes due to going from, say, a stationary reference frame to an inertial, moving one are called ***Lorentz transformations***, part of the theory of ***Special Relativity*** (***SR***) published by Einstein in 1905 (just to give an idea of the time frame

9. Ok, bad example. One side of a pancake has all those little holes in it.

involved), although the transformation had been discovered the year before by Lorentz. In addition to such ***boosts*** due to relative motion at constant velocity, Lorentz transformations also include ordinary rotations.

SR has a real surprise for us and it is crucial: It says that whether you are sitting still (relative to something) or moving inertially, then if you measure the speed of light, you will always get exactly the same result. To the nearest meter per second, this is 299,792,458 meters per second (mps), but we will make do with the approximate value of 300,000 km/sec. Think: If I am moving past you in my car at 100 Km/sec and a high-speed TGV[10] Is moving past you at 200 Km/sec in the same direction, I will measure the TGV's speed relative to me as 100 Km/sec. This is intuitive. But If I am moving at half the speed of light past you, 150,000 Km/sec, and a photon (light particle) is moving past me, both you and I will measure the photon's speed as 300,000 Km/sec. I will not think it is moving at 150,000 Km/sec. This is not intuitive but it is true, and it has been confirmed in countless experiments.

- ==> According to SR, the speed of light (In a vacuum) is constant – always the same, as long as you measure it from an inertial (non-accelerating) reference frame.

The invariance of the speed of light is what requires Lorentz transformations. Physicists then say that in order to be valid, their equations all must be invariant under Lorentz transformations. Starting from this "law", it is actually quite easy to derive the equations of the Lorentz transformation by using only algebra and simple geometry. Without doing the math, it is difficult to seize the importance and usefulness of this requirement.

The invariance of the speed of light has important consequences. We measure speed in, say, kilometers per second, or km/sec, length divided by time. This means that the invariance of the velocity of light imposes a constraint on the relation between length and time. They must adjust themselves so as to always give this value for the speed of light and the result is a mixing of the two, which is clearly exhibited in the Lorentz transformations, those compatible with SR. The result is that we can no longer consider physical events taking place in space and time separately, but only in a four-dimensional ***spacetime***. Fortunately this constraint is not very important for velocities well below that of light, which is why we managed quite well without it for a long time, as in the first example with the TGV.

In addition to the two assumptions about reality and laws of nature, we now have acquired three no-longer-new requirements.

1. The speed of light in a vacuum is constant, always measured by an observer in an inertial frame of reference to be 300,000 km/sec.
2. The equations of physics must be invariant under Lorentz transformations – rotations and boosts. This is referred to as being ***covariant***. It means the equations which describe nature must take the same form before and after the transformations. This Is so important I will restate it: The laws of physics must be Lorentz-invariant.
3. The invariance of the speed of light imposes constraints on the relation of space and time which means we must understand them not as two concepts, but connected together as spacetime.

Mathematically, the equations of the Lorentz transformation show clearly that we can no longer take space and time to be independent. We must consider the four coordinates of time and location together as what's called a ***4-vector***, usually written in the order (t, x, y, z).[11] A Lorentz boost changes not only where the studied object is (its spatial coordinates) but also how long it is there (its time coordinate). The equations lead to fun facts like these:

- The faster you go relative to somebody else, the slower he thinks your clocks run, including your internal body clock. This leads to what appears to be a paradox, the (in)famous twin paradox: A twin who takes a joy ride in a space ship comes back younger than one who stayed at home on Earth. This was portrayed quite vividly in the movie "Interstellar". Note that the errant twin must accelerate up to speed and then stop (i.e., decelerate) and turn around (a rotation) to come back, so she is not always in an inertial state of constant relative motion. That is why she does not think the twin who stayed home is younger than she is.
- The faster you go, the thinner you get along the direction of motion. A result of this is that ...

10. Train Grande Vitesse, high-speed train. The French TGV cruises at something like 250 km/sec
11. 4-vectors are manipulated mathematically using tensors and matrices, so the necessary amount of math just shot up.

- … it is often impossible to say whether one event occurred before or after another, which means that the notion of simultaneity is no longer tenable.

All these phenomena have their origin in the constancy of the speed of light.

I can't keep myself from jumping ahead and pointing out that Einstein's theory of gravity, ***General Relativity*** (***GR***), says that clocks run more slowly in a stronger gravitational field. This means that a clock on Earth at sea level runs slower than one on the International Space Station – or on a GPS satellite. You might want to say that gravity slows down time, but that's not really true. In their own frame of rest (where they are not moving, meaning their position is constant and only time changes), all clocks run at exactly one second per second.[12]

Here's a fun example about that subject of simultaneity. Leonard Susskind has updated the classic example of a pole vaulter and a barn to a stretch limousine and a garage for a VW beetle. As seen by an observer who is stationary relative to the garage, which has doors open at both ends, the limo, if it moves at a speed close to that of light, will be contracted so that it might fit all into the garage at once. In particular, the observer will see the following sequence of events:

1. Limo front enters garage front door;
2. limo tail enters garage front door, because the contracted limo can fit in the garage;
3. limo front leaves garage back door.

But the limo driver rather sees the barn as being contracted, so there is no way he can fit into it all at once. He sees the following sequence:

1. Limo front enters garage front door;
2. limo front leaves garage back door;
3. limo tail enters garage front door.

Note the reversal of the order of events 2 and 3. Time ordering and, so, simultaneity are out the door and not only of the garage!

In summary:

- The speed of light in a vacuum, measured in any inertial frame of reference, is always 299,792,458 m/sec.
- This leads to the Lorentz transformation laws between two inertial systems and the necessity of using 4-dimensional spacetime.
- The equations of physics must be invariant under Lorentz transformations (covariant, in physics speak).
- Among the interesting results of Lorentz-invariance is the loss of the notion of simultaneity and the time ordering of events.

3. Fields: electric, magnetic and gauge

Physics is not just about moving objects (***mechanics***), it's also about electricity and magnetism, together constituting ***electromagnetism*** (***EM***). This subject was explained already in the 19th century by James Clark Maxwell, whose four famous equations describe all of EM. Maxwell published these equations in 1861, 44 years before Einstein published SR. These equations possess two amazing properties.

1. Maxwell's equations can be solved as a wave traveling at a speed which is explicitly given in terms of two well-measured quantities and calculation shows it to be 300,000 Km/sec – the speed of light! And there's just one such speed. So Maxwell prefigured Einstein's SR.
2. Maxwell's equations are already Lorentz invariant. (Newton's are not.)

Maxwell's equations also confirmed one more concept which was to become primordial for modern physics – the concept of a ***field*** – and, in addition, a gauge field (much more later).

---

12. Whatever that may mean.

We know that EM talks about electricity and magnetism and so about electric and magnetic fields. These fields had been discovered by Michael Faraday in 1821. A ***field*** is a physical quantity which has a value at every point in space and time. It may be a ***scalar*** which simply has a value everywhere, like temperature, about 22°C indoors as I write and about 6°C outside. Or it could be a ***vector***, meaning it has a direction as well as a value. An example of this would be the wind, which here now has, say, a value of 25 Km/sec and a direction of westerly, meaning it blows from west to east. Electric and magnetic fields are also vectors. The electric-field vector points from a negative electric charge toward a positive one. The magnetic-field vector points from the south to the north pole of a horseshoe magnetic, or from the South Pole to the North Pole of the Earth. (The particular directions are conventions, they could be reversed and all would be well.)

Maxwell found that both the two EM fields could be defined in terms of two other fields, one a scalar and the other a vector. And lo, together they constitute a 4-vector, that thingy in 4-dimensional spacetime which transforms by the Lorentz transformations. So we can think of EM as really about one thing, a 4-vector field called the ***EM vector potential***.

One can do a transformation of the EM potential which is said to be ***local***, meaning the change is not constant everywhere but varies from one point to another. Until now, we have considered transformations which were the same at each point in space. A rotation, for instance, can be described by the same angle of rotation everywhere, so it is considered to be a ***global transformation***.[13] Local transformations can be quite different. Imagine (if you can) rotating one corner of a cube through 90° about some point and another corner through only 45° about that point. Our poor cube would not be much of a cube any more.

Here's what is so great about that. It turns out that there is a type of local transformations (not the same everywhere) of the EM vector potential which does *not* change the form of Maxwell's equations! The vector potential itself is not something we can measure. We can only use it to calculate the electric and magnetic fields, which we *can* measure. In the jargon, we say that Maxwell's equations are ***invariant*** under ***local transformations*** of the EM vector potential.

For historical reasons, the EM vector potential is called a ***gauge field*** and its allowed transformations are ***gauge transformations***. Think of trains in the old days, when different countries had different distances, called gauges, between the rails. The passengers didn't notice the gauge – once they had changed to appropriate cars. This is likely the origin of the term gauge in physics. Like it or not, the word "gauge" remains to describe one of the most important concepts in modern physics.

To take home: We can have identical physical situations for different values of the gauge fields. This is because we can not measure gauge fields, only the other, physical fields which are calculated from them. The fact that different gauge fields give the same measurable field can be seen as a ***redundancy*** in our description of nature – different things which lead to the same result.

Just to whet your appetite, here is where we are going. *All* the four known forces of physics are due to invariance under local transformations of gauge fields. The ***four forces*** are the EM force, gravity, the strong force which holds nuclei together and the so-called weak force which makes some particles decay into fragments. Stay tuned...

Notice that so far, we have not used the dreaded word "quantum". It's time now to take the plunge.

But first, a summary: [?? WHERE TO (PLANE) EXPLAIN WAVES??]

- Maxwell's equations for EM fields not only can be expressed in a Lorentz-invariant way, so compatible with SR, but solutions show the existence of plane waves with a speed of 300,000 km/sec -- light waves.
- Maxwell's equations can be expressed in terms of a 4-d vector potential.
- Certain local transformations of the vector potential do not change the physical fields of EM. These transformations are called ***gauge transformations*** and the vector potential a ***gauge field***. The vector potential is not measurable and so a transformation of it represents a redundancy in the system of EM.

In brief: changes → local changes → gauge fields.

13. Remember, rotation means that all objects are rotated about a common point, the center of rotation.

4. Operators in Quantum Mechanics (QM)

Now things get hairier, so you may want to rest now with a hot cup of coffee … or a cold beer.

Up until now, we have discussed what is called ***classical*** physics. In physics, the word “classical” has nothing to do with style or expressiveness, as in music or poetry; there is no such thing as “romantic” physics.[14] But as soon as we bring in QM, we are not talking about classical physics any more. QM is really a mathematical theory with two conceptual parts:

1. definitions of terms (objects and principles) and prescriptions for doing calculations concerning them;
2. rules relating the terms to observables, quantities we can measure.

Although the operations involved are not always intuitive, we should be able to sketch out their meaning.

But what is it all about, what exactly is a quantum? You probably have heard that on a microscopic scale, things we measure no longer are required to have a continuous range of values. For instance, instead of taking on any value between 1 and 10, say, 7.39, they may only have integral values: 1, 2,... 10. On a microscopic scale, quantum phenomena clearly show evidence of this granular nature of some things.

As an example consider a hydrogen atom, which consists of a nucleus composed of one proton and one electron moving around it. The electron does not move around the nucleus in well-defined circular orbits, but it does move in different allowed states, called ***energy levels***. Each such level has a certain, specific energy and these levels are separated, like rungs on a ladder. This is because they depend on parameters which are integers, not continuous numbers. So if an electron "drops" from one level to a lower one, it must lose exactly the difference in energy between the two levels. This tiny packet of energy is called a ***quantum***. If we fire light with photon energy equal to this quantum, we can re-excite the electron in the atom. The word "excited" is used by physicists simply to describe a state which is not in its lowest allowed energy level, i.e., not on the lowest rung. Other, more complex atoms work in similar, but more complex ways.

In brief: A ***quantum*** is the smallest possible chunk of energy a system can gain or lose when its energy comes only in fixed steps rather than continuously. It's like steps on a staircase; you can't take half a step.

When you heat a piece of iron, like the coil on an electric stove, it glows red because the electrons In the atoms are radiating electromagnetic energy with the wavelength we perceive as the ***color*** red. These phenomena are things that physicists can observe and measure.

Microscopic and sub-microscopic phenomena like this are explained by using ***quantum mechanics*** (***QM***). ***QM*** is a set of rules for calculating things and is quite general in scope. By itself, it doesn’t explain anything. In order to study a particular phenomenon, we must furnish more information -- the properties of the system in question, the forces at play and any initial conditions. These might be the initial positions and momenta and the masses of two elementary particles. There are two ways of getting to the quantum version of phenomena:

- If classical equations for the domain exist, we can “quantize” them.
- Otherwise, we are forced to invent the quantum version. In order to do this, we use clues and tricks, probably the most important one of which is the requirement of Lorentz symmetry. We will use this method in QFT.

We now have reached a crucial point: The very basis of QM is based on the notion of *commutation* rules for *operators*. What does this mean?

In classical physics, we measure things like mass, position and velocity. In fact, physicists prefer ***momentum***, which is the product of mass and velocity in low-energy, non-quantum, classical physics. How do we go from classical equations of continuous energies to quantum ones? Mathematically, what we do is modify our equations by replacing those physical measurables, position and momentum, by ***operators***. These are not like telephone operators, if you remember them, but more like surgeons, who operate on someone, thereby changing them. The process of defining operators in place of simple variables is called ***quantization***, because it leads us to QM.

14. There is no such thing as “woke” physics either, although a few individual physicists may be more or less woke.

For instance, the operators for position and momentum have the curious but essential property that if you apply them in one order, say position before momentum, you won't get the same result if you measure them in the opposite order, momentum first. Physicists say these operators do not ***commute***. The difference between the two orders of operation is a very tiny number whose magnitude is equal to ***Planck's constant***,

$\hbar$ = 6.62607015×10^34 Joule·Hz.

This non-commutation of operators, the fact that the result of two operations depends on the order in which you do them, is one of the *fundamental* characteristics of QM. Similar results hold for other pairs of quantities, like energy and time, which also depend on their order of measurement.[15] Such pairs of variables are called complementary or ***conjugate pairs***. Many physicists consider these ***commutation relations*** to be not a result of QM but its very *basis.* They lead to the famous ***Heisenberg uncertainty principle***[16], which says that in the case of a pair of conjugate variables, the precision with which we can measure one of them is limited by the precision of measurement of the other. You can have one or the other, not both.

One physical interpretation of the simultaneous measurement of position and momentum, is that in order to measure momentum, we look at a moving particle whose position we therefore cannot precisely measure. But the uncertainty is not a result of insufficient measuring equipment, it is built into QM itself. We will never be able to measure the two quantities better.

This is as good a moment as any to insist on the importance of two fundamental numbers used in QM – I and $\hbar$.

- The first is called, both by physicists and mathematicians, i and is equal to the square root of -1, $i = \sqrt{-1}$. Yes, minus one. You may protest that such a number cannot exist, it is not real. We agree, it is not real, so it is called ***imaginary***. In math, it allows the construction of so-called ***complex*** quantities, composed of real and imaginary parts, and there are rules governing their manipulation. The use of complex quantities allows us to employ math ideas without which we just could not do QM.
- We've already met the other number, ***Planck's constant***, h, usually divided by 2π and written $\hbar$.

The difference of the non-commutating variables is just the product of i and $\hbar$. Both the momentum operator[17] and the Schrödinger equation cited above contain these two quantities. In spite of this, the values of all measured quantities, including momentum and energy, turn out to be real, not imaginary.

Some everyday examples, really analogies, of non-commuting operations might be filling and emptying a cup of coffee. Filling and then emptying does not give the same result as emptying and then filling; only the latter ordering leaves you a delicious brew to drink. Simple math gives examples like addition and multiplication. Suppose I act on the three numbers 2, 4 and 6 by adding the first two and multiplying by the third: That gives (2+4)*6 = 6*6 = 36. But if I multiply first, I get (2*4) + 6 = 8 + 6 = 14, not at all the same result.[18] As already stated, ordering is fundamental In QM,.

This is all very well and good, but if position and momentum are operators, what do they operate on? The answer was published by Erwin Schrödinger in 1926, thereby assuring him a Nobel Prize as well as having his picture on post-war Austrian banknotes until he was bumped aside by the generic (and fake) architectural designs of euro notes. But I digress... Schrödinger derived an equation in which the operators operate on a thing called the ***wave function***. In the same year, Max Born showed that the absolute square of the wave function could be interpreted as the probability that the system studied was in a particular state. And so QM became probabilistic. The wave function was therefore referred to as a ***probability amplitude***, called an amplitude because it must be squared in order to give a probability.

That's why we no longer draw electrons in well-defined orbits around the nucleus in an atom. We don't know the precise place where the electron will be at any moment, only the probability that it is there. So instead we draw what looks like clouds (smears, if you are being less positive) where the density of the cloud is proportional to the probability the electron is located there.

If such operators operate on particular wave functions called ***eigenfunctions***, the result is simply the same wave function multiplied by the value (the ***eigenvalue***) of the quantity represented by the operator. These eigenvalues

---

15. Actually, time usually is not considered an operator, which is important. See my paper on symmetry and groups.
16. Which is therefore not an independent result.
17. In the x representation. but you don't need to know that.
18. A still different result comes from 2*(4 + 6) = 2*10 = 20.

are the set of sometimes-but-not-at-all-always discrete (separate, discontinuous) values the quantity represented by the operator is allowed to take on. It's really simple:

(operator for X) x (eigenfunction of X) = (eigenvalue of X) x (eigenfunction of X).

More mathy, an energy eigenfunction might behave like this:

$$\hat{H}\psi_E = E\psi_E,$$

where $\hat{H}$ is the energy operator (designated an operator by its hat}, ψ_E is the energy eigenfunction with energy E, the eigenvalue of the energy. This is a very simple form of Schrödinger's equation, a hairier form of which we saw in equation (4). The solution for the eigenfunction is then a probability amplitude, so we must take its square in order to get the probability that the system has energy E.

The message to take home is that assuming physical quantities like position and momentum to be operators ***quantizes*** them and renders them non-commuting. Then the Schrödinger equation may only have solutions for discrete (discontinuous) values of parameters like energy, as in the case of orbital electrons around the nuclei of atoms. The solution to the equation then can be used to calculate the probability that the system is in the state given by the eigenvalue. This procedure shows clearly the much-discussed probabilistic nature of QM.

You may want to read that last paragraph one more time.

About now, you may be wondering where this wave function is. Answer: In an abstract mathematical space called a ***Hilbert space***.[19] Much of physics and abstract mathematical spaces of physics are inventions of the mind — imaginary in the sense of imagined, not $i = \sqrt{-1}$. They help us organize and make sense of the world. Multi-dimensional spaces let us distinguish patterns and structure in what would otherwise be a disorganized tangle of data. Some of these spaces are closely tied to reality and so easier to understand, such as configuration space, which tracks the positions of several objects at once and therefore has dimensions equal to three times the number of objects. Others are more removed and purely abstract, such as momentum-energy space, which is sometimes used to describe systems in terms of their momentum and energy rather than position and time. In quantum physics, the most important of these spaces are a special type of mathematical space invented by the German mathematician David Hilbert (a friend of Einstein) and now named after him.

What you need to remember. Physics proposes the existence of many spaces, not only for "ordinary" spacetime but also for other quantities like spin or isospin. These are referred to as "internal" spaces because we cannot go into them from our 4-d spacetime. We'll get to that subject shortly.

In summary:

- QM "quantizes" a system by turning variables into operators, conjugate pairs of which may not commute. Non-commutation leads to the Heisenberg uncertainty principle, which expresses limits on the simultaneous measurement of the pair. It is not due to faulty experimental equipment, but is built into QM.
- The allowed values of physical variables are given by their eigenvalues, which are in some cases discrete or separate, not continuous.
- The solutions to Schrödinger's equation are wave functions which represent specific states of the system. The square of the wave function[20] gives the relative probability that the system is in the state of the wave function.

- *So QM, because of the quantization of variables and the resulting non-commutivity of the operators, introduces possible discrete values of parameters; and it trades in definite values for probabilities, certainty for uncertainty.*

Remember we said earlier that QM is really a mathematical theory, a set of objects and of prescriptions for doing calculations on them. It's a way of calculating. We have not yet considered a real physical object.

19. Much of QM concerns abstract objects which only exist as mathematical concepts or formulas. But images (like analogies) of these abstract objects can be set up, or imitated, in real spacetime in such a way that the real objects behave like the abstract-math ones. Such analogical objects are called ***representations***. If that's too much, don't worry about it.
20. Really the square of the modulus, or absolute value.

5. Relativity and symmetry in QM: spin

What we have discussed so far is standard, non-relativistic QM, meaning that it is not covariant under Lorentz transformations. Now we have to apply what we know about symmetry under transformations . The first step is to require the equations we use to be compatible with SR, i.e., Lorentz-invariant. The equations should look the same if we observe the system after rotating it or while we are moving inertially, with constant velocity. Now we are working in relativistic QM, or RQM. We then come up with not one, but three possible equations for describing free particles (Don't worry, you don't have to remember these.):

- one for particles with spin zero, the Klein-Gordon equation;
- one for particles with spin ½, the Dirac equation;
- and one for particles with spin 1, the Proca equation.[21]

Oops, did I say spin? Uh... an explanation is obviously in order.

But first, why bother? We will see reasons below, but an important one comes from a theorem published by *mathématicienne* Emmy Noether in 1918. The theorem explains the link between symmetries, like those we are considering, and conservation laws. Reminder: In physics, conserved means constant, unchanging. Conservation of quantity X means that we can measure X at any point in the life of the isolated system and it will always be the same. ***Noether's theorem*** says that symmetries give rise to conservation laws and explains how (mathematically, of course). Specific examples include the conservation of momentum, due to symmetries under translation (displacement); of angular momentum, due to rotational symmetry; and of energy, due to symmetry over time.

In the absence of forces, like friction, momentum is conserved, meaning you don't speed up or slow down, but coast along forever. Such momentum is more commonly referred to by non-physicists as inertia. A similar thing is true when you are not moving along in a straight line, but rotating, like a spinning ice skater. Yes, I said rotating, like we were talking about at the beginning of section 1. Angular momentum is conserved, meaning that in the absence of forces, you just keep on turning. It's the conserved (constant) angular momentum of a top which keeps it upright, before it slows down and falls over. A gyroscope also functions by conservation of angular momentum.

Spin is a sort of angular momentum. We can't see anything spinning, nor does the theory suggest anything that does. But the equations for spin behave just like those for angular momentum. In fact, they are identical in their form. It's just that they are not for ordinary angular momentum, but for something else, which is called ***spin***. The spin of the electron was discovered in 1922 by Walther Gerlach, acting on an idea of Otto Stern. The experiment showed not only that electrons behave in an inhomogeneous magnetic field as if they were spinning, but that the spin must have value of $\frac{\hbar}{2}$ and the component of it we measure is oriented either up or down relative to the magnetic field. Conservation of angular momentum must take spin into account in the calculation.

Now hold on tight. As I said, we cannot see what is turning to produce spin. In order to explain this angular-momentum-which-is-not-quite-angular-momentum, we suppose that it is not turning in the same 3-dimensional space as angular momentum like that of our top. It is in its own separate space, a space called an ***internal space****,* as opposed to the exterior space of spacetime. We already talked about such spaces and spinor space, as it is called, is an excellent example..

Yep, there are other spaces than the 4-d one we think we live in – mathematically, at least. Spin occurs in such a space, which can have different dimensions than the 3-d one we all know and love, or the 4-d spacetime of Einstein. Because it lives in another space, we cannot measure spin directly. But it can have effects which are observable in our day-to-day spacetime, so we can infer it from experiments like that of Stern and Gerlach. In a reaction, it is the total angular momentum, i.e., the sum of the spacetime angular momentum and the spin which is conserved. This is another reason for considering spin like angular momentum.

So the three equations mentioned above occur for spin spaces of different dimensions. If the spin space has no dimensions (imagine that!), the particle described is a ***scalar*** with spin 0. If the spin space is 2-d, the particle is a ***spinor*** of spin ½ and is said to live in ***spinor space***. If it is 4-d (really twice 2-d), the particle is a ***gauge boson***

21. When we say a spin value of, say, 1, we mean one unit of spin measured in quantum terms as a multiple of $\hbar$, the Planck constant divided by 2.

of spin 1.

If you prefer, you can think of these other, internal spaces as just convenient interpretations of the math, which behaves and calculates like a space but in terms of other quantities than length, time, momentum and so on. But I find the Universe of all those spaces to be far richer, indeed, quite fascinating. If the equations look like the equations of space and are transformed (more or less) like equations of space, chances are they represent a duck … I mean, a space. Your choice.

And yet, the spinor space is not completely independent of “ordinary” spacetime. Indeed, one can show mathematically (but I won’t) that a rotation in spacetime through a certain angle entails a rotation spinor space too, but of *half* that angle.[22] The factor of ½ originates in the spin value of ½ $\hbar$. In particular, what we see as a difference of up an down in ordinary space (180°) corresponds to a change through only 90° in spinor space. Hmm...

There is yet another aspect to this situation. Particles whose spin is an odd multiple of ½ (½, 3/2,...) are called ***fermions***. Particles whose spin is an integral number (0, 1, 2, etc.) are called ***bosons***. The essential point here is the distinction between fermions of half-integral spin and bosons of integral spin. The existence of these different values of spin comes out of the theory of ***Lie groups*** (Lie is pronounced like Lee), which describes transformations like rotations or Lorentz transformations, the latter being the source of spin.[23] Classical angular momentum is due to symmetry when rotations are done in normal spacetime; spin comes from rotations in spin space, which is rotated by only half the angle.

You may now skip the next paragraph, but if you do, you’ll miss something.

If you are still with me, the group which describes Lorentz transformations is in fact a product of two groups, each having either zero or half-integral spin. Since pairs made up from combinations of 0 and ½ can give 0 (½ - ½), ½ (½ + 0) or 1 (½ + ½), these are the allowed values of spin for the particles represented by the three equations already mentioned. No experiment has discovered an elementary fermion with spin other than ½, although many composite particles exist with spin 3/2, 5/2 and so on.

While we’re on the subject of fermions (remember, half-integral spin) and bosons (integral spin), you will be interested to know that:

- Fermions are the particles which constitute matter. This is because, according to the ***Fermi exclusion principle***, no two fermions are allowed to occupy the same state. If they could, they would, and there would exist absolutely no structure of any kind, such as that of baobabs, aardvarks or us.[24]
- Bosons do not constitute matter but are the particles which carry forces. (We will see why shortly.)

In summary:

- We must make QM relativistic by requiring its equations to be Lorentz-invariant. This procedure leads to three relativistic equations for free particles of spin zero, ½ or 1. Their solutions represent ***fields***.
- ***Spin*** is so called because it behaves mathematically exactly like the equations of normal angular momentum. But spin is in an internal space proper to the particle, not in the external space of spacetime. Nevertheless, conservation of angular momentum requires the inclusion of spin. In an interaction, spin may not be conserved, but the total angular momentum, spacetime angular momentum plus spin, is. And that’s a second good reason for considering spin to be a form of angular momentum.
- Particles of half-integral spin -- ½, 3/2 and so on -- are called ***fermions*** and make up the ***matter*** in the universe. Particles of integral spin – 0, 1, 2 and so on – are called ***bosons*** and carry ***forces***.

So much for the basics. Now for stuff yet farther out!

22. This is too complicated to derive here. See my document on symmetry and QFT.
23. The Lie algebra, an instantiation or example of the Lie group, for the Lorentz transformation is really a pair of rotational (or unitary) groups (SU(2), since you asked), each in 2-d, for a total of 4-d. *Intéressant, non*?
24. Again, this is too complicated to go into here. Suffice it to say that the commutation relation for fermions is not the difference in applying two operators in opposite orders, but the sum.

6. On to QFT

Welcome back. First, a quick review:

- According to SR, the speed of light in a vacuum will always be measured by an observer in an inertial (non-accelerating) reference frame to be the same, 300,000 Km/sec. Velocity is a spatial distance divided by a time, so keeping it constant puts constraints on both and that leads to mixing of space and time in equations and in real life. It thereby prohibits our considering space and time as separate. We must work in 4-d spacetime.
- SR requires the equations of physics to be Lorentz invariant, unchanged under Lorentz transformations, which may be rotations or boosts, changes of velocity which pass us from one inertial system to another. Maxwell's equations of EM are already Lorentz invariant.[25]
- The way to QM is through quantization of variables, replacing them by operators on a wave function. Operating on eigenvectors, they return eigenvalues of the variable they represent. The magnitude of the square of the wave function evaluated at a particular state is the probability that the system is in that state.
- Elementary particles have spin of values 0, ½, 1. Spin behaves mathematically and physically like angular momentum, but in its own internal space.
- Particles with half-integral spin are fermions. Elementary fermions, the stuff of ***matter***, only have spin of ½. Particles with integral spin are bosons and carry ***forces***.

Stated thus simply, that's not so hard. The next step can follow one of three different paths, all of which are valid and useful, some more so in certain situations.

6.1. Second quantization – many particles

Using all this and symmetry, we can deduce the equations of motion for three types of *free* particles: scalar (spin 0), spinors (spin ½, fermions) and vector bosons (spin 1). By "free" particles, I mean with no other particles or fields around to complicate things, it's as if each was all alone in the Universe. Having written the equations, we can solve them. The solutions are fields which are expressed as sums of terms, each one multiplied by a "weight" parameter. The "terms" themselves represent plane waves of differing energy and momentum. They behave a lot like harmonic oscillators, such as pendulums. By "behave like", I mean their equations are similar. This is nice, because physicists have been studying these beasts for a long time and understand them really well. But studying one lone particle, with no others about to drop in and chat, is not very interesting for the particle or for us. So we need to go further.

It was Pascal Jordan in 1926, followed closely by Dirac the next year, who came up with the simple-once-you-think-of-it idea of second quantization. The trick is to repeat what we already did in order to get to QM, where we quantized the position and momentum by converting them to operators. For QFT, we do the same thing again, but this time, we quantize the free-particle *fields*, the solutions to the three equations of motion. Since each field is a sum of terms, we must also quantize the coefficients of the terms, meaning we treat them as operators too. When we do this, we discover that the quantized coefficients do not all commute: The result of applying one of them after another depends on their order. Because of this property, they now can be added together or multiplied (no more complicated than that) in ways which make them operators for the creation, annihilation or counting of quanta (chunks) of energy. And hey, we can take these energy quanta as … particles! So this trick, logically called ***second quantization***, allows us to express the results of field theory in terms of particles and gives us a mathematical technique for creating or destroying any number of them. This is a great improvement over equations for a single particle. And an essential one.

As Matthew Schwartz says: "At the risk of oversimplifying things things a little, that [second quantization] is all there is to quantum field theory. The rest is just quantum mechanics."[26]

Our original free particle needs no longer to feel rejected and abandoned. It now has company. Note, though,

25. To be complete, we should require invariance under transformations of the Poincaré group, which comprises Lorentz transformations plus translations (displacements in space).
26. Schwartz, Matthew D, *Quantum field theory and the standard model*, 20.

that all its new friends are just like it, the same kind of particle. Not just social interaction, but some variety would be nice.

Schematic summary:

- particle = sum of nodes → one node = harmonic oscillator → 2nd quantization ==> quanta!

6.2. Interactions – local transformations and gauge invariance

It turns out there is another trick we can use, based on symmetries. We already know that the equations which describe the motion of the particles are Lorentz invariant (compatible with SR) because we wrote them that way. But that was under global Lorentz transformations, changing things the same at all locations (physics speak for "positions"). Taking the gauge invariance of EM as a clue and remembering that rotations are a subset of Lorentz transformations, we can look to see what happens if we make our equations *locally* invariant under appropriate rotations. The rotations in question are called ***unitary*** transformations and they have the property that they conserve a type of wave-function product called an ***inner product***. Inner products are used to calculate the probability amplitudes for the different states of the system, so this is a Good Thing. We do not want the observable properties, even if they are only probabilities, to change. Invariance, remember?

An essential component of our mathematical toolkit is something called a ***derivative***. A derivative calculates the rate of change of a quantity. That means how much one quantity is altered as another, variable quantity, such as position or time, changes. The first quantity is then said to be a function of the second. The speed of your car, for instance, is the difference in location at one moment and the next divided by the time elapsed between the two moments. The distance from Lyon to Paris is about 460 Km, so if you drive it in four hours you are driving really fast at an average speed about 460/4 = 115 Km/hr. That is a time derivative, because you have a difference in time in the denominator of the calculation. If you are walking in the mountains, the slope you are struggling to conquer is the difference in the altitude divided by the horizontal difference between the two points. The gradient is a space derivative, since the denominator is a distance in space. In QM, we have both types, because we are using 4-vectors which measure space and time and we want the rate of change with respect to both.

The problem is that when we take a derivative of our locally transformed function, we can't calculate its change just by taking the difference of the values of the function at the two points. We must also take into account the fact that the function is transformed differently from one point to the other. Rather than just taking the difference between the function values at the two points, we must take into account the difference in the change between the two points. Slightly more mathy, we must do something to the function at the second point in order to measure it as if it were evaluated in the frame of the first point. Only then can we take the difference. This calculation can be done and the result adds a new, extra term to the derivative. Since this term allows us to connect the state of the function at one point to its state at another, it is called a ***connection***. When it is added to the derivative, the result is called the ***covariant derivative***. The equation in terms of the covariant derivative is covariant too, meaning it does not change under the local transformation. With an ordinary derivative, the equation would no longer be covariant.

Here's an example. It's really from General Relativity, but the principle and, indeed the vocabulary, is the same. When we move about on the Earth, it appears to us as if we are moving on a flat surface. Geometrically, we can view it as a two-dimensional surface. But it's not really flat when viewed from space. Let's go for a very long walk, about 30,000 km. As in figure 6.1, we start at a point A on the equator. We hold an arrow so it points straight in front of us, eastward in this case. We then walk (and swim) a quarter of the way around the Earth while being careful to keep the arrow pointing in the same direction, parallel to its original direction *in space*. This is called ***parallel transport***. Now at point B we turn left 90° toward the North Pole, but we turn our bodies, not the arrow, which we keep parallel to its original direction in space, so it is still pointing generally eastward. We plod on up to the North Pole, keeping the arrow parallel to itself. At the Pole. we turn left (south) and mosey back on down to the Equator. You can see from the diagram that the arrow is now pointing behind us, to the north. We are back to where we started but the arrow has changed directions in spite of our resolutely keeping it parallel to itself. Initially pointing east, it now points north. The reason is because we kept the arrow oriented parallel to itself in space, but the 2-d surface on which we were walking was changing constantly because of its curvature – a local change.

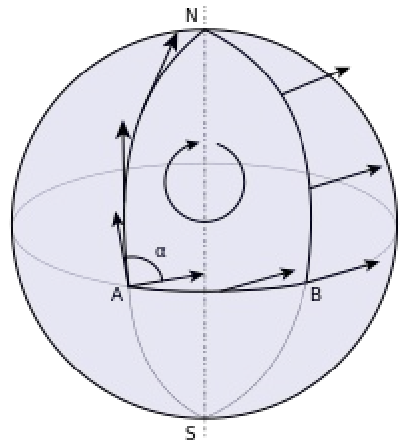


Figure 6.1. Parallel transport of a vector on the surface of the Earth.

Try it with a globe and a matchstick. Turn the globe while keeping the matchstick stationary in the room. You'll see. At the end of the circuit, the matchstick has returned to its original position, but no longer points in the same direction. If you do the same thing on a flat surface like a sheet of paper, but with a rectangle instead of a triangle, your toothpick comes back to the same direction. It's the fact that the Earth's surface is curved that makes the difference.

This particular path configuration is the most striking. The effect is far less obvious with shorter paths or smaller triangles.

An analogy comes from a simple model of finance, calculating the difference in the prices of a commodity in two different countries. You can't just subtract the price in one from that in the other, you must take into account the difference in currencies, the exchange rate. If I want the difference in price in something between France and the USA in euros, I must first change the USA price in American dollars to euros by using the exchange rate, and then take the difference. The exchange rate is therefore a kind of connection.[27]

Let's consider the relatively simple case of a single electron. Its equation does not stay the same under a local unitary transformation, but picks up an extra term. In order to deal with this, we change our derivative to a covariant derivative with an added connection which is a product of constant terms and a new 4-vector field. Consistency of the equations also dictates how the field must transform and – lo, behold – it is just like the vector potential of classical EM. In fact, it *is* the vector potential, the gauge field.

This result is a striking success for several reasons.

- By its transformation rule, the connection field can be recognized as the EM vector potential, the gauge field, used to calculate the electric and magnetic fields.
- Using the covariant derivative, the equation is now Lorentz-invariant under the transformation.
- If we write out the covariant derivative, the extra term is a product of the electron (fermion) and vector fields, so it represents the interaction between the two. This seems logical, since we started with the equation for an electron and identified the connection (gauge) field as the vector potential, the EM field whose particle is the photon. [Uh...] We can now calculate the interaction between an electron and a photon.

The particle form of light, the photon is a massless vector boson (spin 1), and from the Proca equation we also know how that transforms under the same complex (local) rotation. Since we want the interaction of the electron with a photon, we need to include the photon too. Whew! So now we have the electron plus the photon plus the product (interaction) term.

(electron) + (photon) + (interaction).

And when we look at the result of the local transformation of this equation, we see that it is conserved unchanged. And it's the same ***gauge invariance*** found in Maxwell's equations of EM.

27. Schwichtenberg, Physics from finance. Maldacena, The symmetry and simplicity of the laws of physics and the Higgs boson. On-line.at arxiv.org/abs/1410.6753.

For your delight, here is the actual equation for the covariant derivative of a system composed of an electron in an electromagnetic field:

$$D_\mu \psi(x) = \partial_\mu \psi(x) + iqA_\mu \psi(x). \qquad (6)$$

This perhaps daunting looking formula is simply decomposed into its elements.

The covariant derivative D_μ of the electron field $\psi(x)$ on the left is the ordinary derivative ∂_μ of the electron field $\psi(x)$ on the right plus something else. The second term on the right is the interaction term and it is indeed a product of I, the complex factor $i = \sqrt{-1}$; q, the charge of the electron; the EM vector potential (the gauge field) A_μ; and the electron field function $\psi(x)$.

Requiring local transformations had led us to the interaction of a photon and an electromagnetic field!

Once more, with feeling. We start by doing a local unitary transformation on the equation of state (Dirac equation) for an electron. Rather than adding the extra interaction term due to locality of the transformation to the equation, we can include it by redefining the derivative in the Dirac equation. This is logical, since the derivative depends on how the system changes from one point in spacetime to the next. This ***covariant derivative*** now includes an extra term which takes into account the change in the transformation between nearby points. Since the additional term serves to connect the two points, it is called a ***connection***. This is generally considered to be a more satisfying reasoning for inclusion of the extra term. It Is the one assumed when giving the usual explanation of gauge invariance.

The field governing the interaction is due to transformation by the gauge field.[28]

How do we interpret these results? Well, we have already said that electrons are fermions and so are matter particles, and that photons are bosons and si are force-carrying particles. The photon was necessary in order to guarantee gauge invariance of the free-electron equation and so the photon is called a ***gauge particle***. We can extend this kind of treatment to a group of two or three fermions, and we have the following cases:

- In the single-fermion case, the single force-carrying gauge particle is the massless photon of EM.
- In the two-fermion case, there are three force-carrying gauge particles, which are interpreted to be the $W^\pm$ and the Z^0 particles responsible for the weak interactions of particle physics. Alas, experiments show that these three particles are very far from being massless or of equal mass. But don't worry, we have a plan, to be divulged in the next section.
- In the case of three fermions, interpreted to be the three ***quarks*** in a proton or neutron, there are eight force-carrying gauge particles, which are the massless ***gluons*** which carry the strong interaction.

The difference in the number of gauge particles comes from the dimension of the unitary group describing the initial n-fermion case, U(1), SU(2) or SU(3), the number of gauge particles being $n^2 - 1$ for the special (SU) groups. So we know that the standard model of elementary-particle physics obeys the three symmetries, the product of which is expressed mathematically by $U(1) \otimes SU(2) \otimes SU(3)$.

Hey, isn't that great! All particle interactions are tied together by the requirements of symmetries (or redundancies) under local transformations, which give us the interaction terms between different particles. And the very existence of the gauge bosons comes out of the same calculations. How to really comprehend this? Here are some attempts.

- Symmetry of a free-particle relativistic Lagrangian (a mathematical entity used to derive the three equations of motion referred to above) under local transformations gives rise to an interaction term which includes a gauge field which represents a force-carrying field or particle. Local transformations are somehow related to forces. But how?
- The requirement of local unitary symmetry leads to
  - the very existence of the connection, the vector gauge field A_μ,
  - its transformation law and
  - the definition of the covariant derivative.

28. Or have I got things backwards here?

    - It also indicates how to construct possible Lagrangians that will be invariant under the local symmetry.
- This can be viewed in two equivalent ways:
    - Local symmetry transformations require connection terms which are expressed in terms of the force-carrying boson fields, or
    - force-carrying boson fields provide the connections that relate local symmetry transformations at different points in space and time. They are somewhat analogous to the exchange rate between countries.

Look out. We are not talking about any old local transformations nor about any particles. The original equations are for free fermions and the transformation groups are U(1), SU(2) and SU(3).

6.3. How about mass? The Higgs mechanism

We're still stuck with the problem that gauge symmetry leaves the W and Z bosons with zero mass..

First, several words about potential energy. (Skip this if you already know about it.) These days, physicists don't talk so much about forces, more often about energy. The two notions are equivalent in their results. Indeed, most of the forces of nature can be derived from potential energies. Imagine a potential energy whose value on a graph looks like a U centered around x=0 and with the y axis being the value of the potential energy. Thermodynamics tells us that systems like to be in a state of minimum energy, so at the bottom of the U. That's in terms of energy, but we can also see it in terms of forces. The force is the negative of the rate of change (yes, the derivative) of the potential. So in the middle of the U, where the potential-energy curve is approximately flat, the derivative is zero and there is no force pushing the system away from that point. Go to the right some, though, to where the curve is turning up in the U. Now the curve is steeper so it's negative derivative is pushing the system back toward zero. Farther to the right, the curve is steeper, so the force pushing the system back toward zero is stronger yet. So now we can talk about potential energy curves, not forces, knowing that our system wants to be at the minimum value of the potential-energy curve.

Meanwhile, back with the W and Z bosons, let's accept the math and assume that these particles *did* have zero mass – just after the Big Bang (in the first 10^{-12} seconds or so of the Universe) – but then something happened to spoil this. In the late 1960s, Weinberg, Salam and Glashow came up with the electroweak theory of EM and weak interactions. Their assumption is that during those first bits of a second, not only the W and Z particles but also the electron had zero mass, and that the EM and weak interactions were therefore symmetric and were really the same. In 1964 a trick was proposed, originally thought of by Higgs, Englert and at least four other physicists.[29] A form of potential energy field was proposed which looks not like a U, but like a sombrero or the bottom of a wine bottle[30] and has its minimum energy value at a non-zero value of the field. This was done simply by adding in an extra term to the potential. You can show this on a graph where the vertical z-axis is the energy of the field and the x and y axes (or radius and angle, if you prefer) are the value of the field. Usually, zero is zero and gravitational potential energy in classical mechanics, for instance, has its minimum energy at zero of the field (taken to be the surface of the Earth). But the Higgs field can have its minimum-energy value out a way from zero, therefore at a non-zero value of the field, like the circular lowest part of the wine bottle (or hat). Saw off all but the bottom couple centimeters of a wine bottle[31] and pose a marble on the high spot in the middle, corresponding to zero of the field. It is obviously not stable, but it is symmetric: Whether the marble looks left or right, it sees the same thing – a gradually increasing slope leading down into a valley. The marble will roll down into the circular trough a centimeter or two from the center. This is the minimum-energy value and it is not at the center of the field represented by the bottom of the bottle (or hat). It is also a non-symmetric state. Looking in one directions or its opposite shows a long groove the marble can roll around in. Looking perpendicular to the groove, the marble sees an upwardly sloping wall it cannot climb up. This is what we suppose happened as the Universe was around 10^{-12} seconds "old". At that time, the Universe (the marble) slipped into a particular but arbitrary place in the trough and lost its symmetry. Physicists call this process ***spontaneous symmetry breaking***.

---

29. I don't know why only Higgs and Englert were selected for the Nobel Prize. But then nobody accuses the prize committee of fairness – or courage.
30. Usually called, I know not why, a "Mexican hat" potential. I prefer wine bottles.
31. No, I haven't done this. Give it a try.

If we put this non-zero but minimal field value into the equations and fiddle around a lot (moving the coordinate system from the center out to the point in the trough, using a covariant derivative to introduce a gauge transformation plus a bit more), we wind up with an equation in which W and Z particles and electrons all have mass, but not the photon, which is satisfying. We also have a scalar (spin zero) particle with mass and this is the famous Higgs particle. Since the Higgs filed only confers mass on a particle, not giving it any (vector) acceleration, it must be a scalar.

The have W and Z particles and the electron gain mass through their interaction with the Higgs field, as is shown by the calculation. To repeat: The gain of mass of a particle results from its interaction with the Higgs field. The equations show no resulting interaction between the Higgs field and the photon, which therefore remains massless, as it is known to be.

You see why it's been described as a dense cocktail party, with Joe Doakes (the photon) ignored by everybody, but, say, Angelina Joly (the W particle) swamped by the attention and thereby slowed down in her passage toward the drinks table. It's as if she's taken on mass (sorry, Angelina).

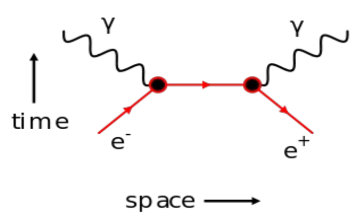
Nutshell: What we have done is to assume an energy field whose minimum is not at zero, which leads the Universe to fall into an arbitrary, non-symmetric value for the field, thus breaking the preceding symmetry. The result of all this is the existence of a Higgs field and particle, and masses for the particles which interact with the Higgs field, such as the W and Z particles, but not the photon. Going in the other direction in time, it also explains how the Universe was more symmetric before the symmetry breaking, so that EM and weak interactions could be considered components of a single theory of massless particles, electroweak theory. All that is evidence for the validity of the Higgs theory.

Calculation thus posits the hypothesis of a combined electroweak force, at least just after the Big Bang. It predicts the existence of the Higgs field and particle. The discovery of a candidate for the Higgs particle was announced at CERN on 4 July 2012 and its identity as the Higgs was confirmed within a year.

6.4. Path integral method and Feynman diagrams

Now we have many particles, thanks to second quantization, and interactions due to gauge bosons, thanks to gauge invariance. We still need a way to calculate what happens when two particles meet.

The probability of a particle's going from here to there depends on the particle, on here, on there and on what happens in between, i.e., on the path taken from here to there. Feynman's *coup de génie*, in 1948, was to realize that one should sum up the probability amplitudes (probability = square of amplitude) for all possible paths from here to there, taking into account what happens along the way, and that many of them give a negligible contribution to the total. But there was more. He showed how the contribution from each path could be represented by a diagram. The diagrams are now named after him. Here is an example, for and electron-positron annihilation into two photons.



Feynman diagram for electron-positron (e^- - e^+) annihilation to two photons γ, by bitwise via Wikipedia.[32]

Calculating the "amplitude" for this (i. e., a wave function before squaring it to get a probability) is not trivial, but Feynman broke the problem down into fairly manageable parts. The amplitude is the product of terms: one for each incoming or outgoing particle, one for each vertex and one for what happens in between vertices. It's the in-between part that is complicated, but that can be calculated by starting from the solutions to the free-particle equations for the particles and doing a lot of math. Once the formula is calculated for a particular interaction, the formula can be reused over and over again in similar diagrams. It's almost copy-paste.

As an example, the amplitude for the interaction of the figure includes"

32. https://commons.wikimedia.org/wiki/File:Feynman_EP_Annihilation.svg

- Terms for each of the incoming (advancing in time) electron e^- and positron e^+;
- a term for each of the two vertecies;
- a term for the red horizontal line, which is an exchanged vector gauge particle, straight out of gauge invariance, charged in this case in order to negate the plus and minus charges of the electron and positron;
- a term for each departing gamma ray γ.

The problem is that the results tend to be infinite, but even this can be resolved by a horrendously complicated subject called ***renormalization***, into which we will very definitely not go!

7. Summing up

In a minimum of words, we have found the following points.

- First, all the summary points of paragraph 6.
- Second quantization permits the creation and annihilation of multiple identical particles.
- Symmetry requires the use of connections based on gauge fields and these are the forces of particle interactions. It is also the basis of conservation laws.
- The Higgs mechanism shows how particles acquire mass by interacting with the Higgs particle in spontaneous symmetry breaking.
- QFT: It's all fields.

Except for SR, these are all quantum mechanics or applications thereof.

Of course, the task of putting all this together and doing calculations to predict or explain experimental data is Something Else.

8. Equations of quantum electrodynamics (QED)

You can skip this, but I can't resist these equations showing how QED comes about from symmetry considerations. And if even you don't understand all the math, looking at them piece by piece will aid in understanding what this gauge-field business is all about. If not, you can pat yourself on the back and go get a cup of coffee.

First, let's admire an equation, the beautiful (yes!) Dirac Lagrangian for a free (non-interacting) spin-½ particle like an electron.

$$\mathcal{L} = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi. \quad (7)$$

Here, Ψ is the electron wave function and $\bar{\Psi}$ is its Hermitian adjoint.[33] What that means, for our purposes, is that the total (integrated over space) $\bar{\Psi}\Psi$ is equal to one, the probability that the particle is somewhere or anywhere.

What do we see in this equation? The simplest thing is the term with m, and since m is a constant, this gives us

$$\bar{\Psi}m\Psi = m\bar{\Psi}\Psi = m,$$

the mass of the electron. Actually, we give it the mass, which we have determined by experiment elsewhere.

The other term is more interesting and has two parts. One is that Greek thingie, the γ^μ. Dirac included this because he needed the equation to be consistent with the relativistic energy-momentum relation (the dispersion relation, in physics speak)

$$E^2 = m^2 + p^2.$$

33. The Hermitian adjoint matrix is inverted complex conjugate of the original matrix. Don't worry about it.

Note that if the electron is sitting still, its momentum p is zero and this equation reduces to Einstein's famous $E = mc^2$. (You may not recognize it because in the other equations, we have set the speed of light, c, and the Planck constant, $\bar{h}$, to 1, a trick to simplify the equations.)

In SR, a Greek subscript or superscript like μ indicates a value of 0 to 3, corresponding to time and the three perpendicular (orthogonal, in math speak) spatial directions. So there are four quantities γ^μ, which I present for your contemplation and enjoyment.[34]

$$\gamma^0 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$
$$\gamma^2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}, \quad \gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

These are therefore 4x4 matrices which multiply the 4-d partial derivative ∂_μ. That's because of the μ super and subscripts Why all this? Because electrons have a spin which – remember – exists in its own space, spinor space, and these are the operator part handling the spin.

Now for the third part. The derivative term ∂_μ expresses the rate of change in the system in time (μ = 0) and space (μ = 1,2,3). If the wave function Ψ is transformed by a constant operator, as is the case with a global transformation, everything looks the same afterwards in their relative positions. If, however, the transformation is not constant, then the change here will be different from what it is over there and this fact must be taken into account in calculating the change in Ψ. This line of reasoning requires a change in ∂_μ which will lead us to the covariant derivative, D_μ, presented in section 6.2.

Here's how the math goes. We consider the local U(1) unitary transformation

$$U = e^{i\alpha(x)},$$

which is a phase rotation through an angle which varies from point to point in space. Under this transformation, the Dirac Lagrangian changes as follows:

$$\mathcal{L}_{Dirac} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi \to \mathcal{L}_{Dirac} - \partial_\mu\alpha\bar{\Psi}\gamma^\mu\Psi, \tag{8}$$

because of the change due only to the transformation. What we then do is, we write a function which translates the field there into the field here. We can use this to write a derivative in which the two terms are coherent and we can do the subtraction. But that is only done at the cost of an extra term in the derivative. Adding this term makes the Dirac equation invariant under a local U(1) transformation. The derivative with its new term is now the covariant derivative,

$$D_\mu\psi(x) = \partial_\mu\psi(x) + iqA_\mu\psi(x). \tag{9}$$

Consistency[35] then requires that the A_μ transform as

$$A_\mu(x) \to A_\mu(x) - \frac{1}{q}\partial_\mu\alpha(x), \tag{10}$$

which is the transformation of the Proca (spin = 1) equation. The result is an added term to the Dirac equation,

$$-q\bar{\psi}\gamma^\mu A_\mu\psi, \tag{11}$$

which is a product of the two types fields, electron and photon, and is just the interaction term we were looking

34. These are in the co-called chiral basis. Don't worry if that means nothing to you.
35. I'm not saying consistency with what because the what is the translation function, which I haven't written. For details, see my paper on "Symmetry, groups and quantum field theory".

for.

Now we must consider the photon with which we want this electron to interact. Since a photon has spin 1 and zero mass, we use for it the m=0 Proca equation,

$$\mathcal{L}_{Proca} = -\frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu) \quad (12)$$

which stays the same (is invariant) if the local $U(1)$ transformation if A_μ transforms like (9), which it does! For any $\alpha(x)$ this is a representation of the same $U(1)$ group for spin equal to 1. It also is the gauge transformation of the standard EM vector potential from classical EM and Maxwell's equations.

So here is the Lagrangian for QED, quantum electrodynamics:

$$\mathcal{L}_{Dirac+int+Proca} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + qA_\mu \bar{\psi}\gamma^\mu \psi - \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu).$$

Including the gauge field to form the ***covariant derivative***

$$D_\mu = \partial_\mu + iqA_\mu \quad (13)$$

"simplifies" this to looking pretty much like a sum of the Dirac and Proca equations (7) and (12).

$$\mathcal{L}_{Dirac+int+Proca} = \bar{\Psi}(i\gamma_\mu D^\mu - m)\Psi - \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu).$$

So in this sense, equation (13) for the covariant derivative tells us that it is the EM vector field A_μ itself which connects one point to the next. At the same time, the field is the photon and gives us the interaction term for the electron and photon. All by insisting on covariance under a local U(1) transformation. This is a big deal indeed. Remember from second quantization that in fact the photon is an excitation of the field A_μ.

We can do similar calculations for two or three fermions under local SU(2) or SU(3) transformations and will find the results already mentioned in the bulleted list of section 6.2. In each case, the covariant derivative has added to it connections which are the fields of which the force-carrying particles are excitations. This gives us the interaction term we need. Poof!

Thoughts and remarks:

- The non-local gauge transformation requires inclusion of the interaction which adds a connection term which is none other than the force-carrying field of the interaction.[36]
- Turning this around, we can say that *a force (interaction) is due to a non-local gauge field which connects one point to the next*. (It must be understood that we mean infinitesimally separated points connected by a translation, rotation, boost or unitary transformation.) This would seem to say that the non-local field is the force, in the sense that the force-carrying particles are excitations of this field. It is related to the transformation through its change under the transformation, equation (10).
- WHAT THE HELL DOES THIS MEAN? Is it the chicken or the egg, the force or the non-local transformation?

If you are not saturated now, consider one more thing. It is the custom to factor out a coupling constant q like we have done in (11). Then the Noether current

$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi_i)}\delta\Psi = -q\bar{\Psi}\gamma^\mu\Psi,$$

is the ***electric four-current***. The zeroth component of this is the electric charge density, so the total charge is the integral of this quantity:

$$Q = \int d^3x J^0 = -q\int d^3x \bar{\Psi}\gamma^0\Psi = -q,$$

36. Ok, it's multiplied by some constants, but they don't change.

because of normalization. So by Noether’s theorem, global U(1) symmetry means electric charge is conserved. Similar results are found for momentum, energy, spin, isospin and more.