Study Guide to Deep Down Things, by Bruce Schumm

Jeff Grove, jeff1.grove@gmail.com, Fall 2012

Bruce Schumm's book Deep Down Things is a book about the Standard Model of particle physics. It is aimed at those with some background in science, but does not require an extensive knowledge of physics. It uses almost no math, yet it goes much deeper than most popular books on particle physics. In particular, it presents an explanation of gauge theory, which I have not seen outside of a textbook.

This book is more difficult than most in the popular particle physics genre. For an introduction to the theory of the Standard Model, almost as broad but more basic, I recommend Robert Oerter's The Theory of Almost Everything (Penguin, 2006). Another helpful book is Martinus Veltman's Facts and Mysteries in Elementary Particle Physics (World Scientific, 2003). This book is also high in level and covers most of the Standard Model, but omits gauge theory. Most other popular books on particle physics either emphasize only part of the story, or contain lots of historical, experimental, anecdotal, and personal information, or both. These three were selected for their coverage of the whole theory. A very condensed but not too superficial overview is at http://en.wikipedia.org/wiki/Standard_Model.

These notes were prepared in the fall of 2012 for use at the Boulder Public Library's discussion group on Cosmology and Modern Physics. Information about the current activity of this group, and a copy of this guide, may be found at http://www.sackett.net/cosmology.htm.

Readers may take this as some quantum superposition of a road map, things to be sure to notice, a few things to ignore, another view or a fuller explanation of a topic, a justification of ignorance, a rant about terminology, a flood of excess verbiage, and several other possible states. I'm sure every observer will see a different probability distribution.

I am not a physicist, but an engineer who has studied physics some. Where I am interpreting the text or the physics in ways I'm not sure of, I'll use red type. If anyone with more knowledge wants to help by reviewing these (or other) portions, corrections are welcome. Any errors, in red or not, are of course my responsibility.

Chapters 1-5

This section was improvised when I realized it would be needed. With Chapter 6, the style becomes more conventional.

Page 37 A pure wave is not localized in space and time, but a superposition of waves of similar wavelengths can be (p.38). If a particle's energy or momentum is exactly known, its wave function is a pure wave, i.e. all the same wavelength, and therefore extends throughout space. (Wavelength is inversely proportional to energy; with exactly known energy, therefore exactly known wavelength.)

Page 38. A superposition (addition) of waves of a range of wavelengths becomes localized. Consider the fluctuation in amplitude you hear when two musical notes slightly out of tune are played together. Fig 3.3/4 b). Because the two waves are nearly identical in wavelength (or frequency, or energy, or momentum), they may start out up at the same time (in phase), and addition increases the amplitude. Over time, they will drift apart, and will eventually reach the point where one is up and the other down (exactly out of phase). Then addition of the two waves cancels, giving zero amplitude if the individual amplitudes are identical. Applied to particle wave functions, this is a crude form of localization, in that there points where the probability of finding the particle is high, and other points where it is zero. If you add more waves of similar wavelength, but all in a fixed range of wavelengths, you can make the low probability regions larger. As you increase the number of wavelengths, still in the same fixed range, you can make the spacing between non-zero regions larger. Fig 3.3/4 b,c,d). Now consider the size of the region of wavelengths (momentum uncertainty) we are drawing from. If we choose our waves from a smaller range (lower momentum uncertainty) (fig 3.3 -> 3.4), the isolated high probability region gets wider (higher position uncertainty).

THE IMPORTANT POINT here is that the uncertainty principle is not mysterious. It is a simple result of the addition of sinusoidal wave functions, which are the solutions of QM wave equations. The only surprise is that nature has chosen to behave in accord with this simple mathematical foundation in producing some of its most complex and confusing phenomena.

Page 45 bottom: <u>Potential energy</u> can be a confusing concept. Kinetic energy, chemical energy, mechanical energy of pressure, spring compression, elevation, etc. are all specific kinds of energy that can be observed and tracked as it changes form. If you are following one of these forms of energy as it changes its form, you may have to account for its intermediate forms. For example, consider the <u>kinetic energy</u> of a weight bouncing up and down suspended on a spring. At the top and bottom of its motion, it is not moving. Where did the KE go? Well, at the top, the spring is less stretched, it is storing less energy, so even more energy is missing. But it is higher in space, farther from the center of the Earth, so the excess energy has gone into the elevated position. On the downstroke, this gravitational energy goes back into the KE of the weight. At the bottom, KE is again zero, and the excess height is gone too. Where is the energy now? The spring is more stretched, so it is storing more energy. The energy sloshes back and forth between these three specific forms. No potential energy to be found anywhere. Now suppose that you are only really interested in the kinetic energy of the weight. You could CHOOSE to call the energy of the spring or the energy of position in the gravitational field POTENTIAL ENERGY, to signify that they are secondary forms of energy that can be used to store and recover energy used in the primary process of interest, the motion of the weight. Potential energy is not a type of energy, it is a SITUATIONAL CATEGORY, where energy can be stored or retrieved, or produce forces, or otherwise perform some function in a system where some other form of energy is the primary interest. Potential energy, being kind of a side category, sometimes has an arbitrary zero point. In this example, the height of the weight that is designated as zero stored energy is arbitrary, at least as long as it is below the bottom point of the oscillation. As long as the weight never goes below that point, the gravitational potential energy is always positive, and all calculations on the system are unaffected. Even that restriction is unnecessary if you can allow negative potential energy in your calculation.

Page 90-1, fig. 4.12. Schumm states that the figure shows ALL of the Feynman diagrams needed to calculate... This is incorrect. It only includes the possibilities that have exactly seven vertices. Those with fewer vertices are not shown, and those with more do not contribute enough to include to achieve the accuracy stated.

Page 101. Henry Moore's sculpture "Nuclear Energy" is described at http://en.wikipedia.org/wiki/Nuclear_Energy_%28sculpture%29.

Page 135. One of several indices into the Particle Data Group web site: http://pdglive.lbl.gov/listings1.brl?quickin=Y&fsizein=1. It's huge.

Page 135-7, fig. 5.10. Eightfold way diagrams. We've seen several of these, and I have found them confusing. Yes, they are symmetrical, you can flip them around, rotate them, etc. But that isn't the kind of symmetry that they point to, or at least not a very direct manifestation of it. To make sense, they need to be viewed as part of the historical process of figuring out what's what. They aren't much on their own, but they do show the existence of some sort of pattern. Physicists know that when they find a pattern that doesn't make sense, it may be a clue, possibly very indirect, to an underlying pattern that actually means something.

These diagrams showed that the particles in them fit into some sort of regular structure. It was such a powerful clue that Gell-Mann was able to predict the existence of particles missing from the pattern before he had any idea what the real structure was. But the patterns of the particles in these diagrams are quite removed from the underlying structure that produced them.

Let's use fig. 5.10 on p.136 as an example. Remember that this diagram is only of one several diagrams involved in this process. They originated in the "particle zoo" of the early '60s, a large unruly collection of observed particles, with no apparent structure. They had an assortment of masses, charges, spins, strangeness (whatever that is), etc. How to make sense of them all? So you start to categorize them by these parameters, starting with guesses about what might be interesting and which ones to group together. If you start with electric charge, you find that they all have integer multiples of the electron's charge. This is good, because it tells you that charge is quantized. Looking at other characteristics shows that they are also quantized. But you already knew that. So you start looking at these characteristics in pairs. This is conveniently represented by plotting particles in a plane with one parameter vertical and one horizontal. This is called a scatter diagram. Since everything is quantized, all the particles fall on coordinate grid points. With all your particles on it, it's still looking pretty scattered, but not knowing what else to do, you keep playing with them. At some point you notice that if you only look at certain subsets of the particles, the patterns become more distinct. So you start looking at different subsets, and plotting them on different axes. Then, maybe the subsets start to make sense to you. Maybe if you separate them into groups in some ways, and plot them in other ways, the groups are pretty orderly, but with some points missing. So far you have explained nothing, maybe you're just doing numerology. But you have some pretty neat (partial) patterns. So you tell your experimentalist buddies to go to their accelerators and look for something with characteristics predicted by the missing point in your diagram. To everyone's amazement, they find a particle that fits in the diagram. Now, maybe numerology is becoming science, sort of. Find a few more, and you're pretty sure it means something, but you still have no idea what. Then you start playing around with fractional charges (for which there is no actual evidence), and three little imaginary doodads that combine only in certain ways. It still doesn't make much sense. Then you reinvent Lie group theory more or less from scratch, and suddenly it makes a lot of sense.

So the moral of the story is "Don't count your doodads before they're quarked unless you want to shake hands with the King of Sweden." And don't feel bad if you can't see the connection between those eightfold way diagrams and quantum chromodynamics. There's a long, long way between them, even though the diagrams are a really potent clue for future Nobel laureates.

Steve had some very interesting questions. Page numbers refer to Schumm.

> p 76-78 ...anti matter ..."going backwards in time"... in the Feynman diagrams - is this just a convenience, or if not, what is really being asserted here about "reality"?

On p.77 Schumm states that there is no definitive conclusion. I don't recall seeing any purely theoretical point about this. It flows from an interpretation of the math on p.74-75. It seems awfully persistent for having no physical meaning or literal implications. I find it confusing in Feynman diagrams, but maybe it has some use as a shorthand in drawing or reading them. My guess is that it doesn't really say anything about reality.

> p 80 ... "anything that can happen must happen" ... is it a statement about reality?

I think this one is real, although "anything" means "anything not forbidden by conservation of energy, momentum, charge, etc.", and "must happen" means "sometimes happens, with some non-zero probability". It seems to be necessary in too many observable situations to just be a mathematical device. In the case of Feynman diagrams, including all possibilities makes it possible to show how certain observed outcomes are possible, and to calculate probabilities that correspond to observations. Many observed reactions and probabilities are only possible if this is true. Like some other ideas in QM, this one might just be a kludge that gives the right answer for the wrong reason, but it is pretty firmly tied to observations. Note that "can happen" may require energy conservation or not, depending on whether the state that violates it is permanent or temporary (involving virtual particles).

> ...is there any difference when one is talking about quantum behavior?

I think the difference is that in the macroscopic world, because of the very large number of particles involved, the chance of all particles doing even a slightly improbable thing at the same time is so small that it is never seen, thus effectively impossible.

> This also raises the issue of "renormalization" -- is this just a mathematical trick to make a particular mathematical approach work out, or is there something more fundamental going on here?

This one especially smacks of a kludge, but it seems to work very well. Some people question it, but they probably use it too. It makes sense because it takes the idea of spontaneously created particles, considers all possibilities resulting from them, and uses it fiddle with the numbers to get rid of the infinite probabilities. But still, infinity minus infinity equals exactly what I need? (Mike says you can cancel the offending terms in an identifiable way. That helps.)

Here's one from Barry:

> Jeff, there is one thing about a couple of these Feynman diagrams that's bothering me. I'm referring to fig. 4.7 (b), p. 76, and fig. 4.12, p. 96. These show a photon (or photons, in the case of 4.12-1 and 2) that are stationary in space (the x axis). How can a photon be stationary in space if it is "moving" at the speed of light??

Feynman diagrams aren't meant to be taken literally as space-time diagrams, even though they are similar. In particular, particles are often drawn as stationary in *time* if they are virtual. That's because all particles either have an antiparticle, or are their own antiparticle. So by the "rotation" rule, if it slants one way, it's one, and if the other way, it's the other, and if it is virtual, either way works because it won't make it out of the area anyway, and we can't tell which it is. Drawing particles as stationary in space is less common, and I can't think of an example where it is logically useful, but it is still not meant to be taken literally.

Chapter 6

Chapter 6 is a digression from physics into <u>Lie</u> (pronounced "Lee") <u>groups</u>, particularly those used to describe the Standard Model, U(1) and SU(2), and later SU(3). A group is an abstract mathematical structure that shows relationships between members of a set. Groups may be either discrete or continuous, depending on whether the members form a continuum or not. A Lie group is a continuous group with some other properties. They are very commonly used in physics.

The main point of Ch. 6 is the mathematical representation of <u>symmetry</u>. The idea of symmetry is that you can take an object with some property, such as its shape (lots of possibilities, both tangible and abstract), and change it in some way, and it still has the same property. We are interested in two varieties of symmetry, where the <u>symmetry operation</u> is discrete (mirror reflection, the four rotations of a square), or continuous (rotation of a circle, and many others).

The mathematical language that is used to describe these ideas is the language of <u>groups</u>. Groups with discrete <u>members</u>, such as mirror reflection, will occasionally be relevant to us, but these are fairly easy to understand. Most of our effort will be in understanding groups in which the group elements form a continuum. All of these will have group members that are variations on the rotation of a circle.

Clearly there is a connection between the set of "objects", the property of them that we require to be preserved, the transformations of the objects that leave the property unchanged, the <u>members</u> or <u>elements</u> of the group (which are actually the transformations), and the <u>group operation</u> that combines group members. From the point of view of group theory, the "objects" and their preserved property are secondary, even though they are often what we really care about. The group elements and group operation are central.

We will start with the definition of a group. We aren't going to do anything formal with them, but it will come up as we talk about the properties of various groups. We will talk about rotation about various axes. Mostly, these will be abstract rotations, which preserve some abstract property. We call them rotations because they form a group which is mathematically analogous to angular rotation in some kind of space. We will use spatial rotation for examples, but don't worry about visualizing the physics in terms of rotation in ordinary space, because it isn't. We usually won't even talk much about the space the rotations are in, because it's an abstract space, and we aren't going deep enough to care about it. For us, it's just an elaborate analogy.

One point of confusion can be what the <u>members</u> or <u>elements</u> of the group are, and what the <u>operation</u> on them is. The way we will talk about the examples will lead to the impression that the group members are the "positions" of the rotated "object", such as a square or circle in a plane, and the group operation is the rotation of it. <u>This is not correct</u>. (The confusion is furthered because "object" is often used as a synonym for "element" or "member". The rotated object may be what we really care about, but it is not part of the definition of the rotation group.) The <u>members</u> of our groups will be the <u>various kinds and amounts of rotation</u> we are considering, and the <u>operation</u> will be the successive performance (called <u>composition</u>) of these rotations. This probably doesn't mean much now, but I will point it out again as we go along. Schumm first points it out on p. 149. Just remember that neither the act of transforming (rotating) the object whose property is preserved, nor the property, nor the object, nor the resulting position of it, are part of the definition. These are part of the application of the group to a specific situation. The group is the set of elements (transformations that could be applied to the object, preserving the desired property), and the rule for combining them to create another element.

On p. 141-3, we find the definition of a group. <u>A group consists of a set of elements, or members, and a binary operation that combines any two of them</u>. If a and b are members of the group, then the operation on them is written a*b or b*a. The order of the members in the operation may or may not matter, depending on the type of group. We will not actually be manipulating groups with these rules. The important thing for us is the flavor of the group structure, which underlies the physics we will discuss.

To be a group, the members and operation must follow four rules, called the group axioms.

1. For any members a and b in the group, a*b is also a member. (b*a is also a member, of course, but it might not be the same member.) This is called <u>closure</u>. It is what it is meant by an operation that leaves a certain property unchanged, i.e. still a member of the group. For rotation groups, this means that if any rotation a is followed by rotation b, the composition (successive application) of a and b is also a member of the group. This is pretty obvious for rotation in a plane, and in this case a*b = b*a, because the rotations add. For particle states, it means that any transformation of a state is always another particle state in the same group. Quarks always turn into quarks, never leptons. (In the last two sentences, notice how easy it was to gloss over the group operation of rotation in some abstract space and wind up talking about the "object" being transformed by the rotation.) Closure is the most important of the group properties for us.

2. In a succession of operations on members, it doesn't matter which operation is done first, as long as there is no change in the order of the entire sequence. i.e. (a*b)*c = a*(b*c). This is called the associative property. I don't think we will talk about it much, although it is important in calculations.

3. A group must contain a member, I, called the identity element. When the group operation combines I with any member, a, of the group (in either order), the result is a, not some other member. i.e. a*I = I*a = a for any a. (I is always the same element for any a, and I commutes with all members.) In rotation groups, I is rotation by zero degrees.

4. For any member, a, of a group, there is always another member, a', such that a*a' = a'*a = I. a' is called the inverse of a, and is different for each a (except I, which is its own inverse).

- Another property of interest is *not* a property of all groups. That is the commutative property, introduced on p. 142. It relates to whether, and in what cases a*b = b*a. For different groups, it may be true always, never, or in some cases and not others. Most of our groups are non-commutative, or non-abelian.

Again, the important thing is the flavor of this structure, not being able to perform manipulations with it.

On p. 144-6, we have an example of a discrete circle group with four elements. This is the same as the group of rotations of a square in a plane, about its center; any multiple of 90 degrees leaves the square's appearance unchanged. The last paragraph of this section is a useful reminder.

On p. 147, Schumm introduces continuous (Lie) groups, in particular, continuous rotation groups. The top of p. 149 has another useful reminder.

On p.150, we get to complex dimensions, and things start to get complicated. Schumm uses "size" to mean the magnitude of a complex number. (In conventional notation, if complex z = x+iy, the magnitude |z| = sqrt(x^2+y^2) is the length of its vector in the xy plane.) This structure allows rotation in one complex dimension to occur, where the magnitude of z is unchanged, even as x and y change. As in the unit circle, fig. 6.2, this rotation corresponds to the angle θ. These operations on a complex number turn out to be the same group as rotation of a circle about its center (360 degrees gets you back where you started), and it is commutative. The angle θ is called the phase of the number. Phase and magnitude together define the number completely, just as x and y do. We will mostly use the magnitude and phase form.

There is an error on p. 151 near the middle. 0.64+0.36i is incorrect. A better example is 0.707+0.707i, which is the 45° rotation case.

On p. 153, he introduces rotation in ordinary 3-space. As shown, this group is non-commutative; it matters in what order the rotations are done.

On p. 157, we get the idea of generators of a continuous group. I think a Lie group is just a continuous group with a finite number of generators, although there may be a few other properties as well. The generators are like the unit vectors (rotations about each of the axes) in the space of all rotations. In 2-space, there is only one generator, since there is only one kind of rotation. In 3-space, we need three, one for each axis. Any rotation in 3-space can be composed of a specific angle of rotation around each of three axes, just as in addition of multiples of basis vectors. In general, rotations with a single generator will commute. On p. 158-61, he discusses the relationships between generators, in particular which operations commute and which don't. This leads to the introduction of the Lie algebra.

On p. 161, Schumm introduces SU(2), the group of rotations in two complex dimensions. Now, we're getting pretty abstract. A two dimensional complex number is just like a two dimensional real number (a vector), except both components are complex.

> I'm going to change Schumm's notation into a more standard form. He uses x and y for his two complex components, and s_x and s_y as their magnitudes. This invokes the image of real and imaginary parts in a confusing way. I will represent our two dimensional complex number as z = (z1, z2), where z1 and z2 are each complex numbers, z1 = x1+iy1, |z1| = sqrt($x1^2$+$y1^2$), etc. The "length" of z is then |z| = sqrt($|z1|^2$+$|z2|^2$).

We could visualize rotation in <u>one</u> complex dimension as analogous to rotation in two real dimensions. But two complex dimensions - what do you do with that? My first guess was that it would be like four real dimensions, but that isn't right. Remember that rotation preserves the size and shape of things (note 6.7), and that we preserved the magnitude (size, length) |z| of a complex number z=x+iy by changing it so that x^2+y^2 remains fixed (fig 6.2b). In complex rotation, we need to preserve the magnitude of each complex component, and the magnitude of the composite, but we can allow the phase of each component to vary. We can transform z1 alone so that $x1^2$+$y1^2$ remains fixed but its phase changes, and the same for z2. That's two kinds of rotation and two generators. We can also "rotate" z1 and z2 together while keeping the "angle" between them the same, so that the composite magnitude sqrt($|z1|^2$+$|z2|^2$) doesn't change. That's another kind of rotation, and one more generator.

But there are four real numbers in two complex numbers, so what about the other combinations? Maybe we could transform z in a way that keeps $x1^2$+$y2^2$ fixed, or the other way around. We could, but then |z1| and |z2| would change, even though |z| wouldn't. So we wouldn't be preserving the "shape" of z, and that's not a rotation as we have defined it, and hence not part of <u>our</u> group (which we <u>defined</u> to preserve the magnitudes of z1 and z2 individually, as well as the magnitude of z). This distinction is the difference between two complex dimensions and four real dimensions. In four real dimensions, rotation only needs to preserve the sum of the squares of all four of the components, while in two complex dimensions, it also needs to preserve the magnitudes of each complex pair separately. This is called the "complex structure". (Are these two paragraphs reasonably clear and accurate?)

The first new paragraph on p. 164 begins a good summary of the above. Pages 166-7 talk about the difference between R(3) and SU(2), which have to do with how the spin ½ and spin 1 qualities are modeled. (Something about having to go 720 degrees to get back where you started. I think it's like a Mobius strip. I don't know if that refers to all generators or only some. I don't think I will be able to find or create an intuitive explanation.)

Page 168 has some interesting trivia, and then a statement about what we are doing. We're making a mathematical model of nature. To "explore the implications" means to see what our model predicts, and if it conforms to nature. If it predicts something we didn't know, it's not that nature conforms to the model, but only that our guess was better than we knew at the time. Nature has been gracious in that when a model conforms well, its logical consequences frequently turn out to conform to and predict features of nature we didn't know about when we created it. This is an important boost to the credibility of the model.

Chapter 7

Chapter 7 is about the application of Lie groups to particle physics, to describe the symmetries of … I'm not exactly sure what. I think Schumm covers it pretty thoroughly, so I don't have much to add. In Chapter 7, I am particularly trying to create subsection markers, and provide further clarification and explanation of specific points. It might be best to follow this guide in parallel with reading the text, referring to it whenever you need it.

Some confusion seems to arise from multiple or unclear meanings of the word forms of "spin" and "rotate", so I'll try to clarify them first. I will refer to these definitions from time to time in the text below.

Spin. Usually clear from context, after you recognize the two possibilities:

1. <u>Intrinsic spin</u>, (of a fundamental particle) p. 177. A quantum number of a fundamental particle related to its total inherent angular momentum about its hypothetical "axis of rotation". This is the largest spin value that will be measured, if the measurement axis coincides with the spin axis. Only odd or even multiples of hbar/2 can occur. It is fixed and not dependent on measurement circumstances. In general, this hypothetical axis of rotation can be in any direction. (The axis of rotation is hypothetical because the whole concept of a body rotating about an axis is only a fair analogy. The real mechanism is unknown.)

2. <u>Spin projection</u>, p. 179. The result of a real or hypothetical measurement of the spin of a particle along some chosen axis. It can be any of the allowed values between + and - the intrinsic spin. In a specific measurement, the probability distribution of the measured values will depend on the relative angle between the particle's "true" spin axis and the measurement axis. (Can also be used with rotations in an abstract space, i.e. isospin projection.)

<u>Rotate</u>. This one seems to be particularly problematic, because it is such a common word in general usage:

1. <u>Rotation of a body about an axis</u>. Although this may not be the literal origin of the intrinsic angular momentum of a particle, the similarities are strong enough, and nobody worries about it. This usage doesn't lead to much confusion with the other usages.

2. <u>Rotation of the spin axis of a particle in physical space</u>. It can point in any direction, and it can change.

3. Rotation (of whatever it is) in an abstract space refers to the various rotation groups that we will study.

Now, to the text:

In the middle of p. 174, Schumm reminds us why we care about Lie groups. "...leave...unchanged" refers to the group property of closure. The middle of p. 175 begins a brief digression on a discrete group with only two members, "reflect", and "don't reflect", which is the identity element. Noether's theorem applies, and the conserved quantity is parity. Note 7.3 explains the difference between mirror reflection and parity inversion. In ordinary 3-space, they are equivalent. Page 176-7 returns to continuous groups. This time, it's the group of rotations in space, and the fact that the laws of physics are invariant (symmetrical) with respect to them. This is connected by Noether's theorem to conservation of angular momentum.

On p.177, we get serious about angular momentum and spin. A review of the basic physics:

- h and hbar have the units of angular momentum, which are energy * time. (The units are not intuitively obvious. Remember that linear momentum is mass * velocity (M * (L/T)), and is a vector in the direction of the velocity. Angular momentum about an axis is equal to the sum (or usually, integral) of the linear momentum of each point of mass, times its distance from the axis ($M * (L/T) * L = M * (L^2/T) = M * (L/T)^2 * T$). Since kinetic energy is ½ m v^2 ($M * (L/T)^2$), so the last form of the angular momentum units is the same as energy * time.) It's quantized, as described.

- Angular momentum is perhaps arbitrarily defined as a vector pointing along the axis of rotation. Curl the fingers of your right hand around the axis in the direction of rotation, and stick the thumb out. Your thumb points in the direction of the angular momentum vector. The vector's magnitude and units are as above.

- Note that hbar = h/2pi (hence the "bar"). This is because angular momentum can also be calculated from a function of the mass distribution ("moment of inertia", I think) times the speed of rotation. Speed of rotation can be expressed as either revolutions/unit time or radians/unit time. 1 revolution (or cycle) = 2*pi radians, hence the ratio of 2pi. It's like feet and inches, except for angles. You could even use h-double-bar = h/360, if you want to measure rotation in degrees. Fortunately, they don't.

Page 178 introduces intrinsic spin, the amount of angular momentum a particle actually has. Page 179-81 describes spin projection. See "Spin" above for summaries and the distinction. Plots like fig. 7.3, 7.6, and 7.7 (the dots are the actual items plotted) are various examples of the possible spin projections in a situation. Fig. 7.8 and 7.9 are similar, except that "spin" of some sort is simultaneously projected on two different axes. These plots have dots instead of lines because the quantities involved are quantized.

Page 183 introduces the idea of an abstract space. The example he uses seems to be the same as "spin space", which he later seems to claim to be a physical space, so I'm confused about that. For now, let's just go along and try to sort it out later.

At the top of p. 183, he says the dimensions of this space are "associated with" the probability of a hypothetical future measurement giving spin projection +½ probability (horizontal axis), or -½ (vertical axis). He actually means "the square root of" the probabilities, allowing both positive and negative roots. Since the probabilities themselves must add to 1, then all possible points (see fig. 7.4) must be on the circle $x^2 + y^2 = 1$, hence they are all 1 unit from the origin, and all reachable by simple rotation about the origin (phase change). (Note that rotating 90 degrees from 3 o'clock (+½ is certain) to 12 o'clock (-½ is certain) involves 180 degrees rotation in the physical space where the electrons and their detector live. I think this is related to the "720 degrees to get home" idea.) But he still seems to be calling the horizontal/vertical space of fig. 7.4 an abstract space at the top of p. 185.

On the rest of p. 185 and top of p.186, we get a comparison of R(3) and SU(2), leading to the assertion (if I read this right), that we really do need to rotate the *physical* spin axis of our particle 720 degrees in *physical* space to get completely back to the same state we started in. I can sort of see this from fig. 7.4, but I'm still confused.

The first paragraph starting on p. 186 and on to p. 187 (except as noted below) discusses the paradoxical nature of spin – what could it be?

The second paragraph on p. 186 points out that various particles with different intrinsic spin (see "Spin(1)" above) need different wave equations to provide their wave functions. Note 7.7 is a nice summary of how this came about for spin ½ particles.

The rest of the section is a summary of our ignorance, which may really include the relationship between the physical space (where spin "lives"?) and the abstract space that describes it mathematically. But the math they use does predict the actual experimental observations, so I guess we're stuck with it, until someone finds a theory that both works and makes sense.

The new section starting on p. 187 is about isospin. It is a quantum number relating to the strong force. Note 7.9 points out that this name is a contraction of "isotopic spin", since it relates to the fact that the strong force binding energy in a nucleus depends on the total number of nucleons, not the specific number of neutrons and protons (the isotope). (We'll find out in Ch. 8 that this is not about either the strong force or the weak force. It is somewhat related to the strong force, but has the SU(2) symmetry of the weak force. I think its use here is just a convenient historical teaching device.)

Excerpts from http://en.wikipedia.org/wiki/Isospin:

Isospin does not have the units of angular momentum and is not a type of spin. It is a dimensionless quantity and the name derives from the fact that the mathematical structures used to describe it are very similar to those used to describe spin.

...near mass-degeneracy [mass-equality] of the neutron and proton points to an approximate symmetry.... The neutron does have a slightly higher mass due to isospin [symmetry] breaking; this is due to the difference in the masses of the up and down quarks and the effects of the electromagnetic interaction. However, the appearance of an approximate symmetry is still useful...[underlining mine]

In addition to isospin, we will also encounter weak isospin and weak hypercharge in later chapters. They are similar quantum numbers, related to the weak force.

The key concept in this section (p. 189-91 top) is that "rotation" in isospin space doesn't change the properties of certain sets of particles, i.e. isospin space represents a symmetry group that is closed under operation of the strong force. Just as an electron is still an electron if its spin projection changes, a nucleon is still a nucleon (to the strong force) if its isospin projection changes.

On p. 191, Schumm points out that there is nothing real or physical about isospin space. He says, "*Rotations in isospin space have nothing to do with rotations in real three-dimensional space or even the quasi-real three-dimensional space of real spin.*" This seems to confirm that he sees real (electron, etc.) spin as existing (somehow) in a *physical* space that at least acts the same mathematically as SU(2), and not in R(3) (i.e. that this use of SU(2) is really intended to be considered as a real space, not an abstract one.). A hybrid of some sort? At least I think that is what he means. Maybe there is just no better way to say it.

But again, we have to accept that nature has decreed that the behavior predicted by this math does happen, whether we think it makes sense or not (various examples on p.192-4). At the bottom of p. 194, he introduces the phrase "internal symmetry space", and says that isospin is the *first* example. This reinforces the idea that he really does mean that the electron's spin lives in a *physical* space that behaves like SU(2). (Is this the same instance of SU(2) as in the electro-weak symmetry U(1) X SU(2) ? I think not.)

On p. 195, we start a discussion about the eightfold way.

> Schumm discusses generators and the Lie algebra, and "representations" of the various symmetry groups here. Other authors also use "representation" to describe the patterns of dots in eightfold way diagrams. There is something called "representation theory" of groups, which I think also deals with generators and Lie algebras, but I can't be sure this is the same idea.

Projections of spin, isospin, and hypercharge are the axes of the plots in fig. 7.7-9. The dots are the projections of physical properties (quantum numbers) of specific particles onto these axes. The reason these plots are important is that they are where the physicists started in trying to understand where all the particles came from. By plotting experimentally observed particles by their projections of these quantum numbers, they found that the particles fell into these patterns, without knowing what the patterns meant. (These plots have dots instead of lines because the quantities involved are quantized.) For example, the bottom of p. 196 tells us that the patterns of dots in fig. 7.7 are "footprints" that turn out to be representations of groups R(3) and SU(2). Likewise, fig. 7.8-9 are footprints that happen to be representations of SU(3) (p. 197-9).

Don't worry about what the patterns mean; historically, they're just footprints. Murray Gell-Mann followed them to Lie groups in 1962 and got the Nobel Prize for it. It's a good thing that there was a big mass gap between the strange and charm quarks, because if there weren't, there would have been many more particles and the patterns would have been much more complex and harder to interpret.

With that, you may skip from the bottom of p. 199 to the bottom of p. 201. Of, if you're really digging in, you might want to look at Veltman p. 230-240 (remembering that all of this was in the days when only the three lightest quarks were available to make particles) when reading Schumm p. 200.

P. 202 compares the SU(2) symmetry of isospin space (proton/neutron or up/down quarks) with the SU(3) symmetry of <u>flavor space</u> (up/down/strange quarks, *not* the color symmetry of QCD). Then he points out that there are a variety of symmetry spaces, real and abstract, and that observing the "representations" of them, in whatever sense he means, can be used to identify the responsible underlying symmetry group. He also notes that a partially filled pattern, if it represents a real symmetry, probably predicts the missing elements. On p. 204 top, we read of an experimental confirmation. On p. 204-mid 205, there is another summary, emphasizing the degree to which these various spaces are "real". He is still calling spin space SU(2), in which electron spin has to rotate 720 degrees to get home, a real space.

Starting mid-p. 205, he has a brief aside about the limitations of direct sensory knowledge. He concludes this on p.206 with the judicious use of "might be", and refusal to speculate about what we might see if our senses worked differently. Then, on p. 206, he introduces the idea of local as opposed to global symmetries, setting the stage for Chapter 8 and gauge theories.

Chapter 8

Chapter 8 is a long and complex chapter, but it consists largely of several variations of the same theme. This theme is that the <u>global phase invariance</u> inherent in complex wave functions can and must be broadened to <u>local phase invariance</u>, which requires additional terms in the wave equation; and that those added terms can be interpreted as the potentials and <u>quanta</u> (particles) of new fields, which turn out to actually exist and behave exactly as predicted by the new terms. This surprising and pervasive fact, known as the <u>gauge principle</u>, is the core of the book. Theories that include it are called <u>gauge theories</u>. Because this idea is highly mathematical in its form, it is rarely presented outside of textbooks in particle physics. The non-mathematical presentation here is the reason this book was chosen for discussion. Another reference on the subject is at http://en.wikipedia.org/wiki/Introduction_to_gauge_theory.

My approach to the notes for this chapter is to briefly summarize blocks of text, sometimes quite large, to make the overall flow of logic apparent. Sometimes, I have left out way too much (including helpful context and partial summaries); other times, maybe not. I would suggest reading this chapter in small chunks, in conjunction with these notes. That will emphasize the cyclic structure of the chapter, and allow you to see how one cycle plays out before moving to the next.

In p. 209-13, Schumm gives a review of complex numbers, emphasizing their nature as magnitude and phase, rather than as real and imaginary parts. A wave function, found as a solution to a wave equation, assigns a complex number to each point in space and time. Wave functions are used to predict observable physical phenomena, which are represented by real numbers. Thus the wave function carries more information than needed to describe reality. For example, the likelihood of finding a particle in a particular region, its probability density, is just the square of the magnitude (or size, or length) of the complex number given by the wave function. The phase is irrelevant, and can be varied at will without affecting the physical result. This invariance is called gauge symmetry, and the freedom to change the phase is called gauge freedom. As shown in fig. 8.1, the variation of phase without changing magnitude is an inherent symmetry of complex numbers. This phase change is analogous to a form of rotation, and the fact that the Lie groups of particle physics can represent rotations of this sort is why they are used.

P. 214 gives us Schrödinger's equation, which will be our first example. You don't need to pay attention to the details of it. The important thing is that it is a form of

Kinetic Energy + Potential Energy = Total Energy,

where the potential energy arises from forces on the particle. (The other wave equations of quantum field theory, for different particles, etc., are similar energy conservation equations, including mass-energy in the relativistic cases.)

On p. 213, we see global phase invariance as a restricted form of invariance (although the name sounds more general), requiring that the phase may only be changed by the same amount everywhere. This freedom is inherent in all of the wave equations of particle physics. But we need local phase invariance, to allow the phase to be altered by a different amount at each point (local causality, p. 218. Some do not consider causality to be the source of the requirement, but it is still needed (D. Griffiths, Introduction to Elementary Particles, footnote p. 359).). (In fact, the local phases can be chosen in various ways to make specific problems easier to work with, similar to the freedom to choose a coordinate system that best fits a problem.) Because the kinetic energy term is a derivative, it responds differently from the rest of the equation to local changes, and requires a change to the wave equation to maintain this invariance.

In 1954, Yang and Mills argued that local phase changes should be allowed. From the bottom of p. 214 to the bottom of p.216 is an explanation of how the kinetic energy term responds differently to local phase changes. Yang and Mills reported (p. 217 bottom to p. 222) that the expectation of only global phase changes violated the principle of local causality (separate points in space cannot affect each other faster than the speed of light), and therefore, the theory must be modified. This requires additional terms to be included in the wave equation. (P. 219 bottom simplifies the equation by dropping the potential energy term. This is not fundamental to the argument, and the potential of the force under consideration will be reintroduced below to allow local phase invariance.)

The result of the failure of invariance of the kinetic energy term under local phase changes led Yang and Mills to add another term to the wave equation to restore local phase invariance (p.223). We will see this innovation applied in several cases. Schumm calls these new terms "cheating terms". This changed equation now describes a new physical situation (p.224 middle), where there is a new potential energy and a new force affecting the particle. The new term (when we repeat this for the relativistic case below) turns out to describe the electromagnetic interaction. The only requirement on the cheating term is that it must maintain the correctness of the wave equation under local phase changes. All of its different valid forms (for different sets of local phase changes) turn out to represent the same electromagnetic field (p.225 top). This is as we would hope, since the point is to make the observable result invariant. Beginning at p. 226 par. 4, over to p. 228, Schumm summarizes all of this nicely, also noting that the cheating term representing the electromagnetic potential is proportional to the particle's charge, as it should be.

To me, the presentation of this remarkable interrelation of ideas, the gauge principle, makes the whole book worthwhile. If you're having trouble with this, p. 208-228 are more important than anything else in the chapter. The rest of the chapter is variations of this concept, and can be considered only superficially in these notes if necessary.

In p. 228-31, Schumm applies the gauge principle to a relativistic wave equation, instead of the nonrelativistic Schrödinger case. The added cheating term leads to the introduction of the electromagnetic field, for which the quantum is the photon. With this new interaction comes a set of minimal interaction vertices, allowing it to be represented in Feynman diagram form, as in Chapter 4. The electromagnetic field theory derived this way is the same as quantum electrodynamics (QED), which was derived without it (p.68-9). He mentions on p. 230 that Noether's theorem connects the global invariance of electron phase to the conservation of electric charge.

This sketch of the gauge theory development of QED is one of several applications of the gauge principle in this chapter. Schumm is enthusiastic about the wide applicability of the gauge principle, the simplicity of the underlying concept, the pervasiveness of its influence, and the minimal information needed to set up an application of it. All that is needed to apply it to a wave equation is to identify the symmetry group and the nature of the "charge" of the interaction, and to measure the strength of the charge experimentally (p. 277). In his examples, he tends to refer to the new interaction and its Feynman diagram rather than the new equation term, particle, or field. They all are related.

On the bottom of p. 231, he tells us that the same gauge principle can be applied to the weak and strong forces as well. The QED case above had U(1) symmetry. The next example has SU(2) symmetry in isospin space (p.232). (He also used isospin space in Ch. 7. It seems to be a teaching device for introducing SU(2), and I don't think it has other relevance for us. Just because it has SU(2) symmetry does not mean it's related to the weak force. Note 8.3.)

The idea of this isospin example is that 1) protons and neutrons are equally affected by the strong force, and obey the same complex wave equation, 2) a single particle can be a superposition of both types, 3) an "angle" in the non-physical isospin space determines the relative amount of "protonness" and "neutronness" in a given particle, and 4) local phase invariance applies to all of this. 5) Therefore, we need to generalize U(1) phase invariance to this more complicated symmetry (p. 234). 6) In doing this, we find that we have three angles to account for, one for the amount of proton/neutron present, and one for the phase of each of these components (p. 234-7). 7) This symmetry is analogous to rotations in two complex dimensions, so the relevant symmetry group is SU(2) (p.237-8), originally described on p. 161-165. There is a midpoint summary on the top of p. 238.

This isospin example continues in the new section on p. 238. If gauge freedom applies, then we can apply the rotations of SU(2) to the wave function without physical effect, and they need not be the same rotations at every point in space (p. 239). The wave equation must be modified to remain correct under these changes. We do this by adding the appropriate cheating terms. SU(2) has three <u>generators</u>, i.e. three independent real angles to represent the possible rotations in two complex dimensions (p. 164, first new par.). Therefore, it turns out (p. 241 middle) that we need three extra terms. He never explains what interaction (if any) this all represents, but ignoring that, the three generators correspond to the isospin angle that determines how much proton and neutron we have, and the phases of the proton and neutron parts of the wave function. In the middle of p. 242, he blows his cover and admits that all of this isn't about anything in particular (in the scope of this book, although it has other uses), but is just a stand-in for the weak interaction to be considered next.

> I followed up the reference in note 8.3 about this isospin example. I learned that there are various gauge theories that can be constructed about various things. This one happens to be sort of about the strong force, but has the same symmetry as the weak force(SU(2)), and also has a conserved quantity, but the "force" that it calls forth is not physically useful (at least in the scope of this book). Apparently, the nuclear isospin example has been around a lot longer than the weak gauge theory, and it has been used as a proxy for the weak isospin theory before.

The next two sections, beginning on p. 243, apply this nuclear isospin analogy to its intended target, the physically unrelated but mathematically similar <u>weak isospin space</u>, and the weak interaction. In place of the proton and neutron, we have the <u>weak doublets</u> (electron-electron neutrino, up-down quarks, etc.), which all behave the same under the weak force. It still leads to SU(2), three generators, and three cheating terms. The three cheating terms, which all have the same coefficient g (weak isospin, p. 245, a quantum number conserved by the symmetry, p. 246) are now seen as the three fields that carry the weak interaction (p. 247). This new interaction leads to three new sets of minimal interaction vertices. The coefficient g is not determined by the theory, and must be measured experimentally (p.246). This time the gauge principle leads to three new fields and particles, W_1, W_2, and W_3 (or W^+, W^-, and W^0; there is some terminology mixing), that carry the force (or cause the interaction) in question. The properties of these fields and particles lead to the same weak force that we previously saw in Chapter 5 (p. 247-51), including the fact that they carry the weak charge themselves (p. 248).

P. 250-51 and fig. 8.9 illustrate the difference between weak interactions caused by the charged W's and the uncharged W^0. Interactions involving "weak charged currents", the charged W's, were known before the gauge theory of the weak interaction. The W^0/Z^0 were unknown before then, and were predicted by the new theory (p. 252 top). "Weak neutral currents", interactions involving the W^0/Z^0, were detected in 1973, providing powerful evidence supporting gauge theories, including that of the weak interaction (p. 252-3).

There is another issue not addressed yet, which is that weak isospin is not a full description of the weak interaction (p253 bottom -254). We actually want the unified electroweak interaction, with another symmetry group, U(1), representing weak hypercharge, also involved. In Ch. 9, we will see how the W_1, W_2, and W_3 are related to the more commonly discussed W^+, W^-, and Z^0. This U(1) symmetry brings in the B boson and the weak hypercharge quantum number, and the formulation of the photon and Z^0 boson as superpositions of B and W^0. This is important in the full description of the electroweak interaction, as we will see in Ch. 9.

On p. 254-5, there is another summary of the gauge principle, from the perspective of the symmetry groups and their generators, leading to cheating terms that describe the minimal interaction vertices of the interaction. On p. 256, we see that if the symmetry group is non-Abelian, the generators do not commute, still more cheating terms are required, leading to more minimal interaction vertices. These new vertices represent interactions between the new field quanta themselves, which implies that they also carry the charge of the interaction.

> Note that in QED, with an Abelian symmetry group, this doesn't happen, so photons can't carry the electric charge or interact with each other (p.257-8). (If nature disagreed, we'd need a different theory.) Actually (note 8.6), there is very slight indirect photon-photon scattering, by way of spontaneous creation (from another photon) of electron-positron pairs, which do interact with the photons. (Can a photon scatter from a pair created from the vacuum? Seems like momentum wouldn't be conserved, but if the pairs appear, why doesn't it happen?)

The minimal interaction vertices are not just *diagrams* of possible interactions, they also include the strength, or probability of the interaction (p.258 bottom). For interactions between matter particles and force carriers, this strength must be measured. For interactions between two force carriers, the strength is determined by the specific nature of the failure of commutativity of the generators, specified by the Lie algebra (p.259-60). (Does anyone know how commutativity and the Lie algebra lead to a quantitative effect?) The section ends with the observation that any given particle's wave function must simultaneously satisfy *all* of the forces for which it carries the relevant charge, and all of their related symmetries.

P. 261-5 review and extend the idea of renormalization, first introduced on p. 80-84. (Renormalization is theoretically and historically important, but not deeply integrated into the narrative I am trying to support. It also takes a lot of words to explain, so I'll let Schumm stand on his own here. It has to do with accounting for all the possible combinations of virtual particles that might be part of any interaction.) Although renormalizable theories are very hard to construct, the lucky surprise is that gauge theories tend to be so, while other quantum field theories usually are not. Even the precise quantitative balance required for renormalizability is provided by the gauge principle.

On p. 265-7, we start another cycle through the logic of the gauge principle, this time describing the strong force with quantum chromodynamics (QCD). This name comes from the fact that the three "charges" of the strong force are named red, blue, and green. This time, we will see the strong force properties of asymptotic freedom and color confinement arise from the gauge principle.

As the eightfold way was being explored, it was discovered that the arbitrary introduction of three independent charges was needed. Each of these could be positive, negative, or zero. Any quark could be a superposition of colors, always positive for quarks, with negative values reserved for antiquarks. Composite particles only have neutral color combinations. Similar to the weak force's indifference to the electronness or neutrinoness of a particle resulting in a symmetry the same as rotation in two complex dimensions (weak isospin space), the strong force's indifference to the color content of a quark leads to the symmetry of rotation in three complex dimensions (color space) (p. 268-9). Although other Lie groups could do similar things in three complex dimensions, the color neutrality of equal amounts of R, G, and B charge, seen in nature, make SU(3) the preferred choice. I think there are some other possibilities, involving higher dimensional groups, which are not completely ruled out.

To make this into a gauge theory, we apply the gauge principle to the wave function and introduce additional terms that describe new force fields. Since SU(3) has eight generators, we need eight new fields and eight quanta (particles) to carry them. Each term has the same strength coefficient g_s, which must be measured experimentally. These terms completely describe the behavior of the strong force, mediated by the eight gluons, and are represented by another set of minimal interaction vertices (p. 269-71).

Just as with the weak interaction, our strong force symmetry group is non-Abelian. That means the gluons themselves carry color charge, they interact with each other, and have their own set of minimal interaction vertices. The strengths of these interactions are again set by the specifics of the non-commutativity of the generators as described by their Lie algebra (p. 272). When observed experimentally, nature again consents to behave as the theory predicts (p.273-4). The rest of the section discusses other consequences of SU(3) non-commutativity.

On p. 275, the last section of the chapter begins with noting the simplicity of the gauge principle (once you grasp its initially obscure mathematical form) and its overwhelmingly pervasive and precise modeling of the behavior of the physical world. Note 8.10 grants consciousness a place to the side of this, since it is not presently explainable as a physical phenomenon, although neural correlates are beginning to be observed. He also excludes gravity, since it is not presently explainable through relativistic quantum wave mechanics. With these exceptions, the global phase invariance of the wave nature of matter, combined with the principle of locality's demand that this be generalized to local phase invariance, mandate the gauge principle. This in turn leads to specific symmetry groups underlying the electromagnetic, weak, and strong forces, and the properties of these groups define the specific behavior of the forces, except for the numerical value of their charge quanta, which must be measured. Thus all phenomena governed by the weak, strong, or electromagnetic forces are described by the gauge principle.

From p. 277 middle to p. 281 middle, Schumm argues that a universe without causation, i.e. interaction, could not exist. This seems to hinge on the idea that the absence of interaction is the absence of existence (p.279, last par.). It seems to me that it is perfectly possible that nature could create a universe that is completely empty, or consisting only of particles that have exactly zero values of all four kinds of charge, and is thus completely sterile. So I conclude that the gauge principle is only responsible for existence of a universe that does something.

In the middle of p. 281, he returns to the question of the physical reality of the symmetry spaces, with no more definite conclusion than before.

Chapter 9

Chapter 9 has two main themes. The first is the transition from the incomplete gauge theory models of Ch. 8 to the current Glashow-Salam-Weinberg unified electroweak model introduced in 1967, which is now part of the Standard Model. The second theme is parity. Finally, these themes are combined to provide further quantitative description and evaluation of the electroweak theory.

The story of electroweak unification begins with the publication of Weinberg's U(1)×SU(2) theory. The SU(2) part is the weak isospin part of the interaction, with the W bosons. The U(1) part replaces electric charge with a new quantum number, weak hypercharge. This model predicts the existence of two unknown particles, the W^0 and the B^0, as well as the W^+ and W^-. (Actually, it predicts $W_{1,2,3}$, and B^0, but that's a fine point.) In this theory, the B^0 is the field quantum of weak hypercharge, but it does not exactly describe the observed electromagnetic phenomena. This is resolved by describing the photon as a superposition of about 77% B^0 and 23% W^0. The Z^0 (also unknown at the time) is the opposite superposition. This synthesis is called electroweak unification. The photon and the Z^0 are the observable particles in nature. Both particles are electrically neutral, and the properties of the photon were well understood. The discovery of weak neutral currents in 1973 confirmed the existence of the Z^0, and led to the Nobel Prize for Glashow, Salam, and Weinberg in 1979 (p. 283-7).

The first theme of the chapter continues with the innovations needed to explain how the field quanta of the SU(2) part of the interaction can have mass without destroying the local gauge symmetry of the wave equations. This begins with explaining why the weak force is "weak". As described before, weak means low probability, which is now equated (in the electroweak case anyway) to short-ranged, implying short wavelength and high energy. This high energy is the rest mass of the W and Z bosons, which gives a floor to their total energy, and therefore a ceiling to their wavelength and range. So we really need these quanta to have mass (p. 287-90). (Viewed from the perspective of potential energy, this phenomenon is called a Yukawa potential. See http://en.wikipedia.org/wiki/Yukawa_potential for more information.)

Adding the necessary mass-energy terms to the wave equations ruins the gauge symmetry and the renormalizability of the theory (p.290-1). So we need a way to introduce some sort of "effective mass" that doesn't do this. An analog of this effective mass is the electromagnetic screening effect of free electrons in a conductor. The presence of free electrons in the conductor reduces the range of the electromagnetic field, just as the mass of the W and Z bosons reduces their range. If the whole universe were filled with (undetectable) free electrons, there would be no way to tell the difference between this screening effect on photons and an effective photon mass. (P. 292-4. I do not know how physically accurate this analogy is.)

The next three sections develop the Higgs mechanism and show how it fits into the theory. Weinberg and Salam included this mechanism in their model. These sections describe *logical development* of the fixes applied to the model to allow the bosons to appear to have mass. After summarizing them, I will back up and describe how the Higgs mechanism works in nature.

A *very* brief summary of these sections begins with hypothesizing the existence of a background field, non-zero everywhere in space and time, which interacts with the W and Z bosons, but not with photons. Since it is uniform, it produces an analog of the screening effect of electrons on photons, which works like effective mass (p. 294-7). Remember that the W bosons arose from the introduction of additional terms in the wave equation to preserve its local gauge invariance. Adding terms for the W masses lost the invariance. Now, we introduce still more new terms, of a different nature, that again restore the invariance. The nature of these terms somehow allows the broken symmetry to be transformed into a "hidden symmetry", which allows the wave equation to retain its local gauge symmetry while incorporating the mass-energy of the W (p. 297-9).

> Schumm describes a logical analog of the difference between hidden and broken symmetries, but not a physical difference. One difference is that one member of the symmetric set becomes different, but it doesn't matter which one, so the symmetry is really still there, but hidden. Sometimes this is called "spontaneous" symmetry breaking. This does not seem to mean "unprompted". Rather, it means that the newly arising difference does not depend on any pre-existing difference between members of the set, but is an arbitrary choice among equal possibilities. Other examples are at http://en.wikipedia.org/wiki/Spontaneous_symmetry_breaking#Other_examples).

These new wave equation terms correspond to the introduction of a new pair of particles. They form another weak doublet, like the up/down quarks or the electron/electron neutrino, which is called the Higgs doublet. These look superficially like new *fermions*, but we don't want new fermions, we want a uniform, pervasive field (and ultimately a *boson*). We now hypothesize that the electroweak symmetry of this pair becomes hidden but not broken by allowing the field of one member to have a uniform non-zero value. (Higgs boson *particles* will be associated with deviations from this background value (p. 304). This is completely analogous to other particles, where deviations from the background value also indicate presence of the particle. The only difference is that for other particles, the background value is zero.) The non-zero field value, not actual Higgs particles, is what interacts with the W and Z bosons to cause the screening that looks like effective mass (p. 299-mid 302).

In order to provide a reason for this background field value to appear, we hypothesize the existence of the Higgs potential (p. 302-mid 303). Schumm does not say anything about what this is or how it works. (In other books, this is sometimes described as the reason that the "lowest energy" state of the Higgs field occurs at a non-zero value. This is still not very enlightening, but the real cause may not be known.) He then offers two interpretations of the two complex fields of the Higgs doublet. The more complete explanation is that these two complex fields can be taken as four real fields. One of these is at the background value, and represents the Higgs field. The other three are incorporated into the three W^{+-} and W^0/Z^0 wave functions (p. 303-4).

Schumm states that the massive, spin-1 W bosons have more "types of motion", so they need more components in their wave functions. This theory gives them three real components each (a spinor?). (Until now, we have only discussed massive, spin-½ fermions and massless, spin-1 photons, both of which only need two components, represented by a single complex field.) This leaves only one field for the massive, spin-0 Higgs. It is called a scalar boson because it has no spin. Does that also mean it only needs a real wave function? In various places (e.g. p. 305), Schumm says that the Higgs field/doublet/boson is unique. I guess this is an example. Another description of the Higgs mechanism is at http://en.wikipedia.org/wiki/Higgs_mechanism, including the following:

> The Higgs field, through the interactions specified (summarized, represented, or even simulated) by its potential, induces spontaneous breaking of three out of the four generators ("directions") [I think he means the four fields cease to be symmetric. -JG] of the gauge group SU(2) × U(1): three out of its four components would ordinarily amount to Goldstone bosons, if they were not coupled to gauge fields.
>
> However, after symmetry breaking, these three of the four degrees of freedom in the Higgs field mix with the three W and Z bosons (W+, W− and Z), and are only observable as spin components of these weak bosons, which are now massive; while the one remaining degree of freedom becomes the Higgs boson—a new scalar particle.

This Web page also references Ch. 9 of Schumm for further reading.

That's how the theoretical view comes together. This left me confused about what actually happens in nature. The last quoted paragraph above has a clue, "after symmetry breaking". Schumm doesn't say much about this, so here's my take, with some help from http://en.wikipedia.org/wiki/Electroweak_interaction and other sources. In the hot early universe, the electroweak interaction was unified. There were the W and B bosons and the Higgs doublet particles, and they were massless. Thermal energy kept things stirred up enough that the Higgs potential had no effect. As the universe cooled, the Higgs potential began to have an effect, which was to spontaneously break or hide the symmetry of the Higgs doublet. One of the four fields became different from the other three by taking on a non-zero average value. The symmetry wasn't really broken, because which of the four that was chosen was a matter of chance, not due to something inherent in the situation. The Higgs potential forced one field to do this because that was a lower energy state than all uniformly zero. (I didn't find a reason why.) The non-zero field causes screening of the W and B bosons, giving them effective mass. For some reason, the W^0 and B^0 also decide to reorganize themselves into the Z^0 and the photon at about (exactly?) the same temperature. The photon manages to escape the screening effect and remains massless. The other three Higgs doublet fields are now called on to be the third components in the wave functions of the W^+, W^-, and Z^0. They need a third component because they have spin 1, and they now have mass. (I don't know why or how these convenient excess of degrees of freedom manage to change from describing the Higgs doublet to describing the W and Z.)

This still leaves me with questions. At high enough energy (>1 TeV?) will we see W^0 and B^0 bosons that have separated themselves? Why did they reorganize into the photon and Z^0

anyway? Is it from the same hidden symmetry effect? Before the symmetry was hidden, were the Higgs particles fermions, or something like them? Do they have different charges like other weak fermion doublets? Is the spontaneous breaking of the Higgs doublet's symmetry limited to only four choices, or is some continuous combination allowed to become the non-zero component? Did this occur abruptly or gradually? What would be different if a different field or combination were selected? If it makes a difference, how did the universe manage to break the same way everywhere, or did it? Is it a coincidence that the energy at which the symmetry breaks only a couple of times the mass energy of the W and Z bosons? If the screening of electromagnetic fields is due to the presence of *charged particles* why does the Higgs field do its screening with a uniform non-zero value and no particles?, There's quite a bit of math behind this, and I'm sure some of these questions are covered. Maybe not all have known answers. And all of this only pertains to the simplest Higgs mechanism, which may not be the one that applies. Anyway, separating the logical view from the physical view was helpful, at least for me.

How the W/Z bosons acquire mass, and everything that follows from it, is (so far) all just a speculation, or with luck, a prediction which needs to be true for the Weinberg-Salam electroweak unification model to work (p. 305 mid). It is also possible that the Higgs mechanism could be responsible for the mass of all other particles, although this is not certain. If quark and lepton masses come from this effective mass, and quark composite object (e.g. neutron and proton) masses are mostly due to the internal energy that holds them together, does "real" mass exist at all? Maybe not (p. 305-6). With that, we set aside the first theme of the chapter, electroweak unification, for a discussion of parity. We will return to the electroweak interaction on page 315, when we can use parity to make quantitative statements about features of the model.

The second theme, parity, arises because the electroweak interaction does not conserve parity. It does in fact matter whether it is presented with a right- or left-handed situation. The only known handedness (I think) in fundamental physics is in the spin of particles, so we will be looking at particle spin and how it affects interactions. After a couple of sections of introduction, we will apply the ideas of parity to the electroweak interaction. This will take the form of comparing the observed and predicted amount of parity violation in various systems to try to confirm the electroweak model. Since the values don't match, we are led to adjust the predictions of the model by including expected but unobserved particles, with the resulting predictions of their masses. In the end, with the discovery of the top quark and the Higgs boson, it all fits.

To my untrained eye, there are lots of loose ends in this part of the chapter. I tried to pull a few out, but it just got worse. Mostly, we don't need to resolve them to follow the reasoning. I'll point out a few that jump out at me as physically interesting or potentially confusing.

First, some distinctions, which are not necessary if they don't bother you:

- "Handedness", or more properly "helicity", and "chirality" refer to the direction of spin of a particle, relative to its direction of motion. They are the same for massless particles, which must always travel at the speed of light. For massive particles, they are different because it is possible to move in the same direction as the particle, but faster, which reverses its apparent direction of travel without reversing the direction of spin. Schumm ignores this, and we can too. http://en.wikipedia.org/wiki/Chirality_%28physics%29

- Parity inversion and mirror reflection are conceptually different, but effectively the same in our 3-space (note 9.6). (This is not true in 2-space.)

- Parity is not the handedness of a situation or object, that's chirality. Something may have one of two kinds of chirality (right or left), or none at all if it's mirror symmetrical. Note 9.7 describes parity as the property of a system that tells whether a particle's wave function changes sign with parity inversion (odd), or not (even). I think parity is sometimes also used to describe particles or sets of particles, to mean whether or not they have chirality (or maybe even which kind?). This seems to extend the confusion. Maybe it's not so bad after you get used to it.

- Parity may also be the property of a process or law that describes whether it does or does not recognize chirality. By this logic, does parity inversion (a logical operation, not a physical one as far as we know) have odd parity, because it always reverses chirality, or even parity because it treats all situations the same? What about a process that produces an output that is right or left-handed, each 50% of the time, regardless of input? It is possible that questions such as this are irrelevant to physical interactions because conservation of angular momentum prevents the ambiguous cases, or because I've split some other hair too finely.

Schumm's treatment begins on p. 306, and notes that parity inversion is a discrete operation, in that you either invert or you don't. Almost everything in our universe is symmetrical under parity inversion, meaning that it would be the same either way. (You have to invert everything involved for this to be true.) P. 308 introduces the tau-theta puzzle, in which two apparently identical particles decay into systems with either odd or even parity. This was resolved by the realization (Lee and Yang, 1956, Nobel Prize, 1957) that nothing was known to contradict the possibility of weak parity violation (p. 309), and the rapid performance of an experiment showed that it did occur (p. 312-13). The tau and theta were subsequently found to be different decay modes of the K^+ meson, in one of which the weak interaction introduced a parity change.

P. 309-10 defines the handedness of spin, and gives an example of parity inversion of a spinning particle. I think the example is confusing, since it seems to mix reflection in a mirror with bouncing off the mirror. It's easier if you just imagine how the particle and its mirror image look to you, and whether it's spinning like *your* right hand (*not* your mirror image's) in each case.

On p. 312, we have a description of the Wu/Ambler experiment that showed that beta decay (a weak process) is in fact asymmetrical. The key to understanding this is in note 9.9, which points out that the decay is from a spin-5 system to a spin-4 system. Therefore a net spin of +1 must be removed. Since two particles of spin-½ (an electron and an anti-electron neutrino) come out, both must be oriented to remove positive (right-handed) spin. Thus, the electrons emitted in the direction of the cobalt-60 spin must be right handed to remove spin, and the ones emitted the other way must be left-handed. The result was that they mostly came out in the left-handed direction, so they were mostly left-handed. A cleaner and more precise experiment using muon decay was quickly performed, showing the preference for left-handedness of the emitted electrons was total (p.313-5). Note 9.12 points out that the electroweak preference is for right-handedness in antiparticles.

The section on p. 315 returns to the electroweak interaction and the Standard Model. The above experiments show a total chirality preference for the W^+ and W^-, although only the latter was described in the text. In the Glashow-Salam-Weinberg model, the W^0 must behave similarly. The photon has no known preference. Since the photon is a superposition of the W^0 and the B^0, the B^0 must have some degree of the opposite preference. Since the Z^0 is a different superposition of these, it must have a partial preference. If the degree of parity violation of the Z^0 can be both predicted and measured, this is an important test of the electroweak model.

P. 317-28 describe the convoluted flow of this attempt. It leads to predictions of the unknown masses of the top quark and the Higgs boson, and becomes a confirmation of the theory only when these masses are known.

The process begins with the proof by Veltman and 't Hooft in 1971 that the Standard Model, with this version of the electroweak interaction, is renormalizable. This means that calculations of interactions involving virtual particles are well-behaved and can be relied on. The three free variables in the model are weak isospin, weak hypercharge (or electric charge), and the mass of the Z^0. The formula on p. 318 gives the fraction of W^0 in the photon (or of B^0 in Z^0). (Schumm doesn't mention it here, but this calculation is for only the simplest case, with no extra virtual particles, shown in fig. 9.6.) The rest of the photon is B^0. Since the parity violation of the photon is zero, and the W^0 is presumed 100% left-preferring, that allows calculation of the degree of right-preference of the B^0. This in turn allows calculation of the degree of asymmetry in the Z^0 (p.318-19).

By 1998, that asymmetry had been measured. The result (here described in terms of the mixing fractions) did not agree with the prediction (p.321). Renormalizability comes to the rescue, allowing the prediction calculation to include more complex forms of the same reaction. (Page 321-323 and fig. 9.7. Note that the figure shows a virtual *fermion* pair.) For 12 fermions, we need 12 more diagrams. But the top quark had not been found yet. Eleven known fermion masses allowed prediction of the top quark mass. When it was found in 1995, it's mass made the predicted Z^0 asymmetry and mixing fraction closer to the measured value, but not close enough (p. 323-4).

By including one more kind of virtual particles, the Higgs, a prediction of its mass could be made. We're pretty far out in the tail of the effects now, so the prediction is not too sensitive to its mass. However, the calculation is very sensitive to the possible range of mixing fractions and net degree of preference, and the predicted range is still within the limits that make a confirming result possible. Any Higgs mass from 35 to 200 GeV will do (p. 324-6). It's apparent discovery in 2012 at 126 GeV counts as probable confirmation of the model.

This discussion raises questions in my mind. Figure 9.7 specifically calls for virtual *fermions* to be included. The Higgs particle is a *boson*, based on the spin of 0. It does seem to be an odd one, but why not include other bosons? Before electroweak symmetry breaking, it looks like some kind of a fermion doublet, but afterwards it looks like a scalar field with the three other scalar fields converted into components of the W and Z fields. In his "Dark Matter, Dark Energy" DVD lecture series (The Teaching Company, 2007), Sean Carroll explicitly calls the Higgs a carrier of a fifth kind of interaction, and a fifth force.

Another question I have is about the chirality preferences of the W bosons. Schumm states that experiment shows left-preference for both the W^+ and W^-, and that the model requires the same of the W^0 (p. 316 top). Note 9.12 says that the left-preference of the weak interaction really means right-preference for antiparticles. But the W^+ and W^- are antiparticles of each other, so why don't they have opposite preferences? If they had opposite (exactly?) preferences, would the entire left/right (and matter/antimatter?) asymmetry issue disappear? And if so, why isn't this simple and intuitive explanation mentioned in all the books?

Another question involves relativity and the presumed equality of inertial and gravitational mass. The Higgs mechanism (as described) creates a kind of synthetic mass to reduce the range of effect of the W and Z bosons. (It seems to give them mass-energy that reduces their maximum wavelength.) Schumm gives no explanation of how this relates to either inertial or gravitational mass. Which is it? Does this provide a theoretical reason why they should be the same? (Or is there already one that I've forgotten?)

On p. 317-8, Schumm states that everything in the electroweak model is determined by the two coupling strengths and the mass of W/Z bosons. It seems that the chirality preference of the Z^0 depends on its B and W^0 content. It seems odd to me that the mass of the Z^0 does not. Maybe the W^{+-} and Z^0 masses are both independent of the weak mixing angle.

On p. 328-30, we have a brief summary of the development of gauge theory and the Higgs mechanism. Lines like the first few on p. 329 appear in several places in the book. I like the version in the middle of p. 289 best. These describe the creation of a gauge theory by *identifying (guessing, postulating)* the underlying symmetry group and the nature of the charge of the associated interaction, and then *measuring* the value of the charges (electric/weak hypercharge, weak isospin, mass) for each particle that participates in the interaction. Then you work out the predictions of the theory and see how well they correspond to nature. In the case of the Standard Model, it was necessary to arbitrarily introduce the Higgs mechanism, with its particle and fields. These were needed solely to provide mass for the W and Z bosons, and to make the theory renormalizable. The Higgs potential serves only to provide as much rationale as possible for the mechanism itself, by providing some undefined physical reason for the background value of the Higgs field to be non-zero (p.330). If the Higgs mechanism turns out not to do the job, some other new physics must be found to rescue the Standard Model (p. 332). But with the apparent discovery of the Higgs boson in July, 2012, the Higgs mechanism is looking very promising.

Chapter 10

The first section is a broader overview of particle physics, some subjects omitted from the book, history, and plans for future accelerators. He also points out that one of the major assets in the particle physics program, the trained people to carry it out, is in place now. If unused, those skills are lost over time.

Then he moves to the benefits and funding of the study of particle physics. The root of the desire is ongoing human curiosity, the pursuit of ever deeper knowledge. At the present scale, particle physics must be a series of multinational efforts. He also mentions a variety of practical technological developments that came from advanced particle physics: Accelerator applications such as cancer treatments, research tools in various fields, and high resolution lithography for manufacturing ever denser integrated circuits. In addition to the accelerators themselves, the detection and data representation and distribution technologies (including the World Wide Web) have wide use in modern society.

On p. 338, Schumm begins a summary of the developments that led to the integration of the Standard Model. My summary is given here. Newton unified the earthly and celestial gravitational interactions. The first unification of overtly distinguishable forces was by Maxwell. His electromagnetic theory unified electricity and magnetism, and provided the starting point for special relativity. It was also the first gauge theory, although this was not recognized at the time. In the realm of particle physics, the particle zoo of the 1950s was organized by the eightfold way, which pointed to Lie group symmetries and the prediction of quarks and the lepton and quark generations. Schrödinger's equation led to other wave equations for various relativistic particles. These equations inherently contained global gauge symmetry, and the requirement of local causality made local gauge symmetry necessary. Including this in the equations was achieved by adding additional "cheating terms" to restore the symmetry, and these terms describe the behavior of the bosons that carry the force of the interaction. However, particles introduced in this fashion must be massless, and thus have long range effects. For the weak interaction, the weak isospin "charge" was not very different from the electric charge or weak hypercharge, yet the probability of the interaction was much less, i.e. weaker (p. 288, 290). It was speculated, and later shown, that the weak bosons' weakness i.e. short interaction range, was due to having mass. To allow this mass without breaking the global gauge symmetry, all of the features of the Higgs mechanism were introduced. Viewed as a whole, this theoretical synthesis, spanning over 85 years, is a huge achievement, and the gauge theory idea is at its core.

P. 340 mentions the generations of fermion doublets. Their appearance as doublets is well-explained by the electroweak theory. However, the number of generations, and even the equal number of quark and lepton generations, is not theoretically well-established. CP symmetry violation requires at least three generations. CP symmetry is a fundamental problem, the basis for the leftover excess of matter that now forms the universe. This is an important unresolved issue, with a somewhat complex hypothesis behind it. See http://en.wikipedia.org/wiki/Baryogenesis and the Sakharov conditions.

On p. 341, Schumm mentions renormalization, which shows up in various places along the way, providing a way to handle the cloud of virtual particles that can enter into interactions. Because of these virtual particles, the apparent charge of a real particle varies with the energy of a particle interacting with it, which causes an apparent change of interaction strength with energy. (This is sometimes called "running" of the charges or strengths.) P. 341 gives electric charge as an example, and p. 342-5 consider the other forces vs. energy. The U(1) strength must increase with energy, but the others are found to decrease.

This running raises the possibility that the three forces might be equal at some energy between 10^{13} and 10^{17} GeV, and therefore might be a single unified interaction. This idea is called grand unification. Fig. 10.1 shows how they might converge. P. 347 calls this unification "all but an assumption of modern science". I'm not sure why this is so strongly expected, except that similar unifications have been found. Note that small changes in slope of the U(1) or SU(2) curve can move the intersection point a lot. Schumm tells us that the introduction of supersymmetric (or other) particles can make the forces converge, with lighter ones causing the slope to change earlier. (Possibly there are some mass predictions for supersymmetric particles. Supersymmetry requires four Higgs bosons, one of which must be roughly near the mass of the one discovered in 2012. Supersymmetric particles are also candidates for dark matter (p. 348).)

In the Epilogue, Schumm points out that we are approaching the practical limit of accelerator energy. After the LHC generation of accelerators, or the next, we may see that there is nothing more to be found short of grand unification energy, ten or more orders of magnitude higher. At that point, fundamental experimentation must stop. Only theory can help, possibly by providing lower energy testable predictions of the hypothetical grand unification.

Conclusion

These notes have turned into a combination of a lengthy review and a study guide for a remarkable book. I have never seen such a thorough explanation of the Standard Model intended for non-physicists, even sophisticated and persistent ones. In particular, the detailed treatment of gauge theory and the Higgs mechanism were entirely new to me, and well worth the effort. I needed to do much of the work for this guide just to develop my own understanding. If it helps others work through this book, so much the better.