Perturbative Quantum Field Theory

Jack Dinsmore

November 30, 2023

I'm a PhD student studying physics at Stanford University. This is my work-in-progress quantum field theory textbook, intended for people who are curious about the phenomena, intuition, and mathematical methods behind quantum field theory and particle physics.

I am not an expert in quantum field theory. I was only a first year grad student at the time I started this book, and my field is astrophysics. I'm writing this book as an exercise to help me remember and understand quantum field theory better because I think it's a particularly unique and beautiful branch of physics. I hope this book evokes the same fascination in QFT in others that it did in me, but please remember that it is by no means authoritative.

Contents

Abstract

Quantum Field Theory, or QFT, is our most fundamental description of the universe. It combines the lessons of quantum mechanics with the laws of special relativity to create a theory — the only theory — consistent with both. Except for gravity, quantum field theory predicts every phenomenon yet observed, with only twenty or so free parameters to be determined by experiment. Below, we provide a short summary of QFT, with links to the chapters in which each topic is discussed. Reviews of relevant mathematical material are provided in appendices at the back of the book.

The need for QFT arose from a variety of sources. Theorists were bothered by the fact that quantum mechanics is inconsistent with relativity[1], while relativity is inconsistent with quantum mechanics[2]. Experimentally, new particles such as the positron were starting to be discovered and a theory that could create, destroy, and propagate these particles was necessary. In chapter 1, we reconcile these difficulties by presenting QFT as a principle of least action theory. We define action as a hybrid between the quantum mechanical and relativistic actions, and we show that this action can create and destroy particles.

The left hand side of this principle of least action is an n point cor-

---

[1] For example, the Schrödinger equation is not Lorentz invariant and allows particles to move faster than the speed of light.

[2] How can atoms be spacelike or timelike separated if they are fundamentally quantum mechanical wavefunctions with uncertain position?

relation functions of particles (nPCFs), where n is a positive integer. An nPCF may seem like an esoteric object at first, but we show that it is directly connected to scattering (chapter 2). For example, the 4PCF encodes the probability for two particles to scatter off each other[3]. The 2PCF is special; it encodes many of the characteristic properties of a particle, including its mass, its ability to scatter with and decay into other particles, and even the energies of its bound states (chapter 3).

Now that the nPCFs are understood, we tackle the right hand side of the principle of least action: an integral of e^{iS} over every path the n particles in the nPCF could take, where S is the action. The value of this integral depends on the action S used, so we start a simple action for a simple particle: a scalar (chapter 5)[4]. We solve for the nPCFs exactly with this action and interpret the results using our understanding of how nPCFs are related to scattering. This verifies that has the properties we expect a relativistic theory QFT to have, such as causality.

Next, we add a new term to the action which represents interactions between the particles. This renders the nPCFs impossible to compute exactly, but we use perturbation theory to find approximate answers as long as interactions are weak. The mathematics of this process are protracted but greatly simplified by the use of Feynman diagrams. After a few calculations, we are able to compute our first scattering amplitudes for an interacting theory. This process leads to a remarkable intuition which holds for any weakly interacting theory: all interactions come about by the creation of virtual particles, which are like a temporary excited quantum state used by the interacting particles to tunnel into another state. Just as an electron can escape a potential well by tunneling through an

---

[3]Two incoming particles plus two outgoing particles equals four particles, hence a 4PCF.

[4]Scalars are also called spin zero particles, or scalar bosons. The Higgs particle is the only fundamental scalar, though other particles such as the pion can be approximately modeled as scalars.

energy barrier in non-relativistic quantum mechanics, now a relativistic particle uses the relativistic equation $E = mc^2$ to create virtual particles as a means of tunneling through an energy barrier to a new external state.

We also show that a light scalar particle is able to carry information about the presence of other particles over long distances. To an observer who cannot see the light scalar particle, this may look like a long-range force. For a light enough scalar, we derive that this force is proportional to $1/r^2$, like Coulomb's force. However, we also demonstrate that it is unlikely that scalar particles can carry forces like this in Nature as it theories which have light enough scalars are rare.

Having treated the simplest case of a scalar particle, we consider fermions (chapter 6)[5]. Immediately, we notice that these fermions must behave differently from scalars because of the Pauli exclusion principle, which says that two fermions cannot exist in the same quantum state. but the spin statistics theorem allows us to encode this Pauli exclusion principle in the action, so we can derive nPCFs for fermions using the same techniques as with scalars while remaining consistent with the Pauli exclusion principle.

With these nPCFs, we propose a new model that binds the neutron to the proton in a nucleus. The model treats protons and neutrons as fermions and proposes a light scalar known as a pion to transfer force between them. We compute several predictions of this model which are consistent with experiment to first order, though experiments also tell us that protons, neutrons, and pions have substructure which this model does not contain. Emboldened, we suggest a similar model for the Coulomb force between a positron and an electron. However, this model cannot explain magnetic fields, so we reject it. **JTD: Symmetry breaking**

In order to pursue magnetic fields, we further consider vector par-

---

[5]Fermions are also called spin one-half particles. Most fundamental particles are fermions, including quarks as well as the electron, neutrino, and other leptons.

ticles (chapter 7)[6]. Like scalars, these particles are bosons so they do not obey the Pauli exclusion principle and can carry forces. However, a vector particle has too many degrees of freedom to be detectable in four dimensions. We propose the removal of one of these degrees of freedom by suggesting that the vector particle obeys some self-symmetry, in that a certain transformation of the vector field leaves the action invariant. This removes one physical degree of freedom from the particle, making it observable. It also limits the number of possible actions down to a very small number. We choose the simplest action, allow the vector particle to interact with a fermion, and we see that the vector particle has all the properties of a photon. This combined theory of the photon and electron is named Quantum Electrodynamics, or QED. We spend the rest of the chapter showing that it obeys Coulomb's force law, reproduces magnetism, and predicts that an electron has a magnetic dipole moment of about two which is consistent experiment. We also show that this theory does not suffer from the same rarity problems, so the photon can be massless.

We would like to compute the predictions of QED to see how well they hold. We start with magnetic moment of the electron g by calculating the scattering probability of an electron off the magnetic field. Unfortunately, our results blow up to infinity[7]! Similarly, the predicted electron charge and mass and the photon energy of QED are all infinite. Two simple physical arguments forestall panic, however. Firstly, the true values are probably merely very large, not infinite, because our equations include interactions in which enormous amounts of momentum are transferred. QFT has not been tested at these large energies and is probably wrong, so we shouldn't rely on its predictions for these interactions. Sec-

---

[6]Vector particles are also called vector bosons or spin-one particles. All the standard model's force carriers — the photon, gluon, and W and Z bosons — are vector bosons.

[7]The infinity is due to the sheer number of ways momentum can be transferred from initial states to final states via virtual particles.

ondly all experiments measure differences between the quantities, and it is possible for a large number minus another large number to be small. In chapter 8, we simulate these arguments mathematically using a process called "renormalization," by removing the infinite part of the equations for the nPCFs. This allows us to compute the magnetic moment of the electron and show that it is consistent with the experimental value.

We follow up this victory by using renormalization to compute other phenomena, such as electron-electron scattering probabilities, Compton scattering probabilities, and the Lamb shift in the spectrum of hydrogen. All of these predictions have been confirmed by experiment and marked a crucial step in the historical development of QFT. In the case of electron-electron scattering we notice another infinite prediction, this time arising from the fact that the photon can carry arbitrary amounts of momentum at low energy due to its massless nature. The answer is again that this infinity is canceled with another infinity, this time coming from the fact the same masslessness of the photon allows arbitrarily many very low energy photons to be created in a scattering process without violating energy conservation. When the probability that these photons appear is included, the infinity cancels out.

This concludes our discussion of QED, the theory of the electron and proton. But at the time that QED was wrapping up, a minefield of unrelated new particles were being discovered, demonstrating that QED was not the full story. In part 10.2, we discuss how these particles, called mesons, displayed properties with surprising symmetries among each other. We introduce group theory as a mathematical language to discuss these symmetries in chapter 11, and identify these symmetries as the group SU(3). The exact same symmetries would appear if each meson were made of two of three smaller particles, which were named quarks in the 1960s. Further investigation shows that hadrons such as the proton and neutron can be made of the same quarks, but this time coming in

triplets. We set ourselves the task of including these quarks into QED in a way which is consistent with experiment.

The first challenge is to explain why all the mesons appear to decay into other particles. QED provides no mechanism for light or electrons to cause this decay, so we require the existence of another vector particle called the W boson. **JTD: Z boson** . We say that the W and Z bosons together carry the weak force. However, the unique pattern of decays that the mesons show require that the W boson does not interact equally with all particles; it only interacts with particles of a certain spin, and it only allows some quarks to convert to others. In particular, the interactions of the up quark are identical to those of the down quark, and likewise for the strange and charm quarks and the up and down quarks. But this pairwise symmetry — SU(2) in group theory language — does not old fully because the quarks in each pair do not have the same masses. To solve this problem, we propose the existence of a new, scalar particle called the Higgs which could give rise to such a pattern of interactions even though the quark masses are different. This Higgs particle, discovered in 2012, was the last particle in the Standard Model to be observed.

A second experimental fact that quarks and electrons are attracted to each other electromagnetically. This implies that quarks are charged, but raises the problem that quarks should repel each other. If they do, then how can quarks group into pairs and triplets to make mesons and hadrons? Another new particle is required — one that carries an attractive force stronger than both electromagnetism and the weak force. We name this particle and force the gluon and strong force respectively. The gluon is another vector boson like the photon and W and Z bosons, and it attracts quarks into mesons and hadrons in the observed way if the equivalent of the electric charge for the gluon can take one of three values. We usually refer to this gluon-charge as color charge, and the three values as red, green, and blue. Then we require an SU(3) symmetry between the

three colors of quarks so that they have identical masses.

There is only one more problem to solve: why are quarks never observed by themselves? Unlike all other particles in the standard model, the always condense into pairs or triplets. No Feynman diagram theory can replicate this behavior, which requires us to accept that gluons interact so strongly with quarks that we cannot use perturbation theory and Feynman diagrams to understand them. We argue that a non-perturbative theory, however, could cause the quarks to condense into mesons and hadrons via an analogy to gases which can condense into liquids under the laws of statistical mechanics.

With the discussion of the gluon concluded, we can now understand all the fundamental particles. However, the action for all these particles is obscenely long and seemingly arbitrary. The Higgs mechanism for the weak theory inspires us to write a simpler Lagrangian under a larger symmetry and use the Higgs particle to break the symmetry and create the masses we observe. This larger symmetry is simply the combination of the symmetries we have already discussed: U(1) for the photon's self-symmetry and the electromagnetic force, SU(2) for the weak force, and SU(3) for the strong force[8]. The new action is much simpler and is almost always consistent with experiment. There are only a few vagueries to be cleared up, which we discuss in the final chapter.

Firstly, this action predicts that neutrinos are massless, which is not the case. It also predicts the existence of a fourth neutrino called the sterile neutrino which has not been detected. Possible solutions exist to both these problems but whether they are correct remains a mystery. Another problem for QFT at the current moment is that it does not predict the existence of a dark matter particle, though one could be added as a new particle which interacts only weakly with standard model particles and is therefore very difficult to detect. This new particle could be another

---

[8]This combination is named SU(3)$\times$SU(2)$\times$U(1), where $\times$ is pronounced "cross".

fermion, or a new kind of particle similar to a magnetic monopole called an axion.

Beyond the problem of whether the particle content of the standard model is correct, there is the question of verifying its predictions for hadrons. Due to the non-perturbative nature of gluons, it is much harder to estimate important properties such as hadron masses than it is to compute other QFT properties, but progress has been made in the field of Lattice QCD which numerically evaluates the nPCFs via the QFT principle of least action. With Lattice QCD, the experimental measurements of hadron masses have been confirmed among many other properties. Though there is some question of whether its predictions for the interaction strengths of hadrons are fully correct because of tension between Lattice QCD, a more traditional theory prediction, and experiment regarding the value of the magnetic moment of the muon.

Many theorists are also concerned with the remaining arbitrariness of the standard model. Why are there exactly six quarks? Why are the particle masses what we observe them to be? Why do the electromagnetic, weak, and strong forces differ so much in strength? A similar question stems from the fact that the standard model action is missing a few terms which could be added without violating any known physical principles. Why are these terms missing?

Finally, many theorists are working to create a version of QFT which is includes gravity. Many candidate theories have been proposed, but no decisive deviations from the standard model have been observed beyond those mentioned above, so these theories cannot yet be tested.

Notation

Units

In quantum mechanics, observable quantities have units of mass, length, time, and charge. This book collapses these four units down to one by defining Planck's reduced constant, the speed of light, and permittivity of free space to be $\hbar = c = \epsilon_0 = 1$. It's customary to keep mass as the remaining unit and define the other fundamental units in terms of mass:

$$[\text{length}] = [\text{mass}]^{-1}, \qquad [\text{time}] = [\text{mass}]^{-1}, \qquad [\text{charge}] = [\text{mass}]^0. \quad (1)$$

in four spacetime dimensions. Such a unit system is called *natural units*. It's also common just to quote the exponent of $[\text{mass}]$, saying for example that length has dimension -1.

To convert a number from natural units back to a more standard unit system, one need only multiply by the correct combination of $\hbar$, c, and ϵ_0. As an example, the fine-structure constant $\alpha = e^2/(4\pi)$ is $\alpha = e^2/(4\pi\epsilon_0\hbar c)$. Other important dimensions are

$$[\text{Action}] = 0, \qquad [\text{Lagrangian}] = d, \qquad [\partial_\mu] = 1,$$
$$[G_{\text{Newton}}] = -2, \qquad [G_{\text{Fermi}}] = -2 \quad (2)$$

where d is the dimension of spacetime.

Special Relativity

We will use the mostly-minus form of the metric tensor for special relativity:

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \quad \text{or} \quad \eta_{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{3}$$

which is equivalent to $\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. A result of this choice is that a timelike vector has positive norm:

$$\eta_{\mu\nu} p^\mu p^\nu = p \cdot p = p^\mu p_\mu = p^2 > 0. \tag{4}$$

If this momentum represents a particle, then the mass of the particle is defined by $p^2 = m^2$. This is known as the on-shell condition. When we refer to the spacial part of a four vector, we use boldface:

$$\boldsymbol{p} = (p_1, p_2, p_3) \tag{5}$$

so that we can also write a particle mass as

$$m^2 = p_0^2 - \boldsymbol{p}^2. \tag{6}$$

The energy-mass relation of special relativity, $E^2 = m^2 + p^2$, indicates that $|p_0|$ is the energy of the particle. We therefore often refer to p_0 as E_p and require that $p_0 = E_p > 0$.

Fourier Transforms

Our normalization for the Fourier transform and its inverse is as follows:

$$f(k) = \int d^d x \, e^{-ik \cdot x} f(x) \qquad f(x) = \int \frac{d^d k}{(2\pi)^d} \, e^{ik \cdot x} f(k) \tag{7}$$

where d is the number of space(time) dimensions used. To prove that the Fourier transform and inverse Fourier transform are indeed inverses, the following formula is useful:

$$(2\pi)^d \delta(k) = \int d^d x \, e^{ik \cdot x}. \tag{8}$$

Chapter 1

Principles of Quantum Field Theory

Historical Introduction

The first three decades of the 20th century marked an unparalleled advance in fundamental physics. Sparked by the observation of black body radiation — specifically its exponential frequency cutoff which suppresses high-energy radiation — Max Planck postulated in 1900 that the energy of light had to come in *quanta*. Inspired, Albert Einstein used this idea of light quanta to explain the photoelectric effect, and Niels Bohr and Ernest Rutherford generalized the idea of quanta by suggesting in 1913 that angular momentum was also quantized in order to explain atomic emission lines. With Louis de Broglie's proposal in 1924 that matter was also quantized, it became rare to find properties of nature that were *not* quantized — a statement made rigorous by the discovery of the Schrödinger equation in 1926, and the foundation of quantum mechanics.

Independently, Maxwell's prediction and the following experimental confirmation that the speed of light be constant had initiated its own cascade of discoveries. In 1905, Einstein reshaped our definitions space and time by uniting them into one entity — spacetime — in order to apply Maxwell's equations of electromagnetism in moving reference frames.

His new theory is now called special relativity. Ten years later, he unified special relativity with gravity to produce general relativity, which explained the slow precession of Mercury's orbit, and now explains much more exotic phenomena such as gravitational lensing, black holes, and the history of the universe.

Though these two theories of relativity and quantum mechanics were each incredibly successful, they were completely in conflict. The famous Schrödinger equation, though in agreement with experiment, contains one time derivative and two position derivatives:[1]

$$i\partial_t \psi = \left(-\frac{1}{2m} \partial_x^2 \right) \psi. \tag{1.1}$$

We've set potential energy equal to zero assuming that the particle is in vacuum. According to special relativity, a Lorentz boost to a new reference frame with velocity v must introduce the substitutions $\partial_t \to \partial_t + v\partial_x$ and $\partial_x \to \partial_x + v\partial_t$. But that would introduce additional space and time derivatives so that the result is no longer the Schrödinger equation. Thus, (1.1) equation is inconsistent with relativity.

A few experimental anomalies had also begun to appear by the late 1920s. It was becoming increasingly apparent that electron spin, largely pioneered by Pauli, needed to be introduced to quantum mechanics. The spectrum of hydrogen, once the poster child of the quantization of angular momentum, was showing a small discrepancy with the theoretical prediction of s orbital energy. Separately, Compton scattering experiments where high energy photons were scattered off electrons showed that light had important quantum mechanical properties which could not be understood with the Schrödinger equation.

Paul Dirac thought that the solution to some of these problems might be the introduction of relativity to quantum mechanics. There were two

---

[1]This is the first time we have used natural units, wherein $\hbar = c = 1$.

examples of equations of motion already introduced to physics that were consistent with relativity: one was Maxwell's equations (in vacuum), written in terms of the vector potential A^μ

$$\partial_\mu \left(\partial^\mu A^\nu - \partial^\nu A^\mu \right) = 0 \tag{1.2}$$

and the other was the wave equation, also called the Klein-Gordon equation, written for a wavefunction ϕ:

$$\partial^\mu \partial_\mu \phi^2 - m\phi^2 = 0. \tag{1.3}$$

Both of these equations are Lorentz invariant, though they are not good replacements for the Schrödinger equation since they do not have the same non-relativistic limit. For the same reason, they also don't resolve the new experimental anomalies. Instead, Dirac created a new relativistic equation. Like the Schrödinger equation, the Dirac equation is contains one derivative and consequentially reproduces the Schrödinger equation in the non-relativistic limit. In order to contract away the index of this single derivative ∂_μ, the Pauli matrices σ_1, σ_2, and σ_3 had to be combined with the identity matrix into a four-vector $\sigma^\mu = (\mathbb{1}, \sigma_1, \sigma_2, \sigma_3)$ which represents the spin operator.[2]

$$i \left(\partial_\mu \sigma^\mu \right) \psi = 0. \tag{1.4}$$

This was the first fundamental theory to describe particle spin. It also allowed the anomalous energy of the s orbitals to be understood as a rel-

---

[2](1.4) only applies to massless particles, which is why m is not present in the equation. It can be generalized to apply to massive particles such as electrons with some extra mathematical machinery, which is done in chapter 6.

The fact that spin and Lorentz invariance were resolved with the same blow in (1.4) is no accident; spin, also known as intrinsic angular momentum, is only a valid particle property because it is conserved, and angular momentum is only conserved because physics is invariant under rotations. Lorentz invariance implies invariance under rotations, so any theory which is Lorentz invariance should also come with a complementary interpretation of conserved angular momentum, or spin.

ativistic correction (the Darwin term), and it predicted Compton scattering cross-sections, with its relativistic electrons.

However, the Dirac equation had one glaring flaw. It predicted the existence of a new particle of equal mass to the electron but with opposite charge — a particle which somehow had never been seen despite the recent surge of particle physics experiments. Dirac wasn't so bold as to predict a new particle without experimental evidence, so he postulated that this particle was the proton, even though experiment had already shown that the proton was much more massive than the electron.

Four years later, it appeared Dirac should have stuck to his guns. The positron was discovered by Carl David Anderson and confirmed to have the same mass and opposite charge of the electron. The new particles didn't stop there; the neutron was discovered in the same year, and muon not long after in 1936. Together with these new particles was the ever-growing theory of nuclear decay, describing how protons, neutrons, and electrons conspired to turn into each other with fixed rates. No theory yet written allowed particles to change identity or split in such a way.

In order model decay, many physicists sought to embrace relativity more fully than Dirac's equation did by fundamentally changing our understanding of particles. Non-relativistic quantum mechanics models particles as described wavefunctions, which are in turn ruled by the Schrödinger equation or Dirac equation. But wavefunctions, which are defined as a function valid at all positions for one unique time t, are not a Lorentz-invariant concept. Fields, such as the electric or magnetic fields, are Lorentz invariant because they are defined for a unique spacetime coordinate. Thus, the search for "Quantum Field Theory," or QFT, began. Initially, progress was slowed by the temperamental behavior of energy; the vacuum energy predicted by the Schrödinger became infinite in a relativistic version, as did other quantities such as particle masses. Some physicists began to suggest abandoning the new QFT framework

altogether, but the occasional success kept many on the QFT track. For example, a new anomaly of the hydrogen s orbital known as the Lamb shift was discovered in 1947 and explained by Hans Bethe using QFT, and a correction to the magnetic moment of the electron was derived in 1948 by Julian Schwinger and experimentally confirmed.

The theoretical troubles of QFT were redressed over the next decade through a number of procedures largely developed by Julian Schwinger, Richard Feynman, and Shinichiro Tomonaga, for which they one the 1965 Nobel Prize in Physics. Their theory, named Quantum Electrodynamics (QED), explains all the phenomena of classical mechanics, non-relativistic quantum mechanics, and special relativity, as well as confirming the new discoveries of the positron, the Lamb shift, and the electron's anomalous magnetic moment. As of the 2000s, its predictions have been confirmed to as many as 12 decimal places in multiple experiments. The theory also addressed several philosophical questions, such as the origin of particle spin and the definitions of fermions and bosons. The first part of this book is dedicated to understanding this incredibly successful theory and how to reproduce its major predictions.

The Principle of Least Action

The three fundamental theories before QFT (classical mechanics, relativity, and quantum mechanics) can all be framed as principle of least action theories. It seems reasonable to pursue such a principle for QFT.

In classical mechanics, action S is defined as the integral of some function L with dimensions of energy, known as the Lagrangian, with respect to time. The integral is computed over the path C taken by a particle:

$$S = \int_C dt\, L. \tag{1.5}$$

We propose as a physical law that Nature choses the path C such that

the action is minimized. Thus, any small change to C does not lower the action:

$$\delta S = 0. \tag{1.6}$$

(1.6) is known as the principle of least action. This principle of least action can be shown to reduce to Newton's second law $F = ma$.

The advantages of the action formulation are many. Firstly, it reduces all the properties of a physical theory to one variable: the Lagrangian. All a physicist need do to create a theory is write down a new form of L. Furthermore, writing a new Lagrangian is made easier by the fact that all the symmetries of S are symmetries of the theory in general. For example, we have observed that physics is invariant under translation — that is, that experiments done in different locations yield the same results. Thus, S cannot depend explicitly on position. Other examples of symmetry are invariance under orientation and time translation. Relativity introduces Lorentz boosting as a symmetry, while QFT will introduce even more. All of these symmetries cut down so much on the number of forms S can take that Lagrangians tend to be quite simple in practice. In classical mechanics, the Lagrangian is

$$L_{\text{Classical}} = K - V \tag{1.7}$$

where K and V are kinetic and potential energy. All other terms are forbidden because they do not have the right units or the right symmetries.

In general relativity, the purpose of the Lagrangian is to dictate how the metric of spacetime $g_{\mu\nu}(x)$ interacts with matter. This requires us to change our thinking, since $g_{\mu\nu}(x)$ is a field and therefore fundamentally different from a particle path $\boldsymbol{x}(t)$ that one might solve for in classical mechanics. In the path case, $\boldsymbol{x}$ is a dynamical variable indicating the one location where the particle can be. In the field case, x is instead a label indicating which part of the field we're solving for; the dynamical variable is $g_{\mu\nu}$.

This distinction fuels a change in our definition of action. Instead of an integral over a path, we add up the individual Lagrangians of every $g_{\mu\nu}$ value in spacetime:

$$S = \int d^4x \, \sqrt{-g} \mathcal{L}(x) \tag{1.8}$$

where $\mathcal{L}(x)$ is called the *Lagrangian density*[3] In the name of causality, we also introduce the principle of *locality*, which means that the physics of one point in spacetime cannot be instantaneously affected by a that of a distant point. Since $\mathcal{L}(x)$ encapsulates the physics at a point, it must only depend on x and derivatives with respect to x rather than some distant point y. This principle of locality means that information can only communicated across spacetime by means of waves, such as the electromagnetic waves of Maxwell's equations, which must move at the speed of light.

Putting together the notion of a Lagrangian density with the requirement of Lorentz invariance and the principle of locality, the general relativity Lagrangian density is

$$\mathcal{L}_{\text{GR}} = \sqrt{-g} \, (R - 2\Lambda) + \mathcal{L}_{\text{Matter}}. \tag{1.9}$$

Here, R is the Ricci scalar (the simplest non-constant Lorentz-invariant scalar defined from the metric), and $\mathcal{L}_{\text{Matter}}$ is the Lagrangian density of matter that happens to be present. Λ is known as the cosmological constant which roughly represents the energy of the vacuum.

Since $g_{\mu\nu}$ is a field, the above Lagrangian is our first *field theory*. Another field theory is electromagnetism, whose Lagrangian is

$$\mathcal{L}_{\text{EM}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.10}$$

---

[3]The $\sqrt{-g}$ term is included in the volume element because to account for the possibility that the spacetime shrinks or expands in some regions. g represents the determinant of the metric.

where $F^{\mu\nu}$ is the Faraday tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. These field theories are still classical because they do not produce quantized particles., so covering the principle of least action for a quantum theory is the next step.

The success of quantum mechanics is largely due to the presence of the new constant $\hbar$, which has units of energy times time. $\hbar$ is necessary to quantize energy, angular momentum, and to write the Schrödinger equation, but $\hbar$ poses a problem for the principle of least action. Action has the same units as $\hbar$, so the principle of least action is less justified. Why can't $\delta S = \hbar$? Or $\delta S = 2\hbar$? Furthermore, how do we deal with the fact that quantum particles have wavefunctions and don't exist at well-defined points?

Paul Dirac developed a simple solution to both these problems in 1933, which was later formalized by Richard Feynman. To compute the probability that a quantum particle could go from position $|x(0)\rangle$ at time $t = 0$ to position $|y(t)\rangle$ at time t, one need only calculate

$$\langle \boldsymbol{x}(t)|\boldsymbol{y}(0)\rangle \propto \sum_{\text{paths}} e^{-iS/\hbar}. \tag{1.11}$$

and the probability is merely $|\langle \boldsymbol{x}(t)\boldsymbol{y}(0)\rangle|^2$. Here, S is the classical action (1.5) of the Lagrangian integrated over a path, and the sum indicates a sum over all possible paths between point $\boldsymbol{y}$ at time $t = 0$ to point $\boldsymbol{x}$ at time t. The proportionality merely indicates that this equation does not automatically normalize the probability; the user has to do this after computing $\langle \boldsymbol{x}(t)|\boldsymbol{y}(0)\rangle$.

Surprisingly, (1.11) is almost equivalent to the classical principle of least action $\delta S = 0$. Consider a the classical path C, which contributes a term of e^{iS} to $\langle x(0)|y(t)\rangle$, where S is the classical action. A nearby path will have almost the same action, since by definition $\delta S = 0$ for a classical path When added, e^{iS} values for both paths will have almost the same phase and will constructively interfere. But if we consider a path

not close to the classical path, its neighbors will have dissimilar actions and will destructively interfere. Thus, the paths that contribute the most to $\langle x(0)|y(t)\rangle$ are those that lie near the classical path.[4]

It is also easy to see how this definition might give rise to some quantum properties. If one shines an electron beam at two parallel slits, each electron may take one of two classical paths. The interference between the e^{iS} values of these paths corresponds exactly to the interference pattern made by the electron on a screen behind the slits. To give another example, when computing decay rates of a quantum system such as an atom, one finds that the way to compute a decay rate is to add up the rate of every possible decay path equally, just as the magnitude $|e^{iS}| = 1$ is equal for all paths.

Quantum Field Theory Lagrangian

As discussed in the abstract, the best way to guarantee a relativistic quantum theory is do away with the concept of wavefunctions and instead use fields, which automatically ensure that information is not transferred faster than the speed of light. So we create a field $\hat{\phi}(x)$ which is a quantum operator. It is defined so that the state $\phi(x)\,|0\rangle$ represents a single particle state with a particle existing at position x. In this field language, the $\langle \boldsymbol{x}(t)|\boldsymbol{y}(0)\rangle$ in (1.11) is written as $\langle 0|\hat{\phi}(x)^{\dagger}\hat{\phi}(y)|0\rangle$ where $x = (t, \boldsymbol{x})$ and $y = (0, \boldsymbol{y})$ are four-vectors.

To write an action for the operator $\hat{\phi}(x)$, we'll borrow the field theory Lagrangian of general relativity:

$$S = \int d^4 \, \mathcal{L}(x) \tag{1.12}$$

where we have dropped the $\sqrt{-g}$ term because we will always work in Minkowski space in this book, where $g = -1$.

---

[4]The rigorous version of this statement is known as the "stationary phase approximation."

The last step is to write a principle of least action and we'll borrow the quantum idea of adding contributions from all paths. However, in the non-relativistic quantum case x was a dynamical variable, so the relevant paths to sum over were paths through space — that is, through values of x. Now the dynamical variable is the value of a field, so the path should be all the values that the field can take at all points. We'll define a shorthand for summing over these paths

$$\int_{\text{paths}} \phi(\text{all } x) = \prod_{x \, \in \, \text{spacetime}} \int \phi(x) \equiv \int \mathcal{D}\phi \qquad (1.13)$$

and use it to make our first QFT principle of least action:

$$\langle 0|\hat{\phi}(x)^\dagger \hat{\phi}(y)|0\rangle \propto \int \mathcal{D}\phi \, \phi(x)^* \phi(y) e^{iS}. \qquad (1.14)$$

The absence of hats in the right hand side is not an accident; the $\phi(x)$ in the integral is a number (not an operator) representing essentially the magnitude of the operator $\hat{\phi}(x)$. That way we can integrate over ϕ to produce a number for the expectation value of $\hat{\phi}(x)^\dagger \hat{\phi}(y)$ as expected, rather than an operator-valued equation.

Again, the proportionality is present because we have not yet normalized the wavefunction, but we can easily do this by dividing by $\langle 0|0\rangle$. If we also generalize the result to arbitrarily many fields in the left hand side,

$$\langle 0|\hat{\phi}(x_1)^\dagger \ldots \hat{\phi}(x_n)^\dagger \hat{\phi}(y_1) \ldots \hat{\phi}(y_n)|0\rangle$$
$$= \frac{\int \mathcal{D}\phi \, \phi(x_1)^* \ldots \phi(x_n)^* \phi(y_1) \ldots \phi(y_n) e^{iS}}{\int \mathcal{D}\phi \, e^{iS}} \qquad (1.15)$$

This law, due to Feynman, is the heart of QFT. It can be used to compute the outcome of scattering experiments and particle properties such as magnetic moments Along with the quantitative predictions, (1.15) gives deep insight as to the nature of particles, since it can be used to derive the spin-statistics theorem which explains why spin-1/2 particles obey the Pauli exclusion principle.

Low Dimensionality Terms Dominate

Before we begin investigating this principle of least action, there are two crucial corollaries of the physics we have so far discussed. One is that the effect of complicated terms in a Lagrangian is suppressed, which can be seen with unit analysis. Since the action is unitless in a natural unit system, the Lagrangian must have units of 4. Consider a scattering experiment, where two particles ϕ with dimension 1 were collided with energy E. If $\mathcal{L}$ had a term like $\lambda\phi^6$, the constant λ would need to have dimension -2 to satisfy unit analysis. Whenever this constant appeared in the scattering cross section, it would need to be accompanied with an energy to allow the units to work: $\lambda/E^{-2} = \lambda E^2$. This would be measurable at high energies but suppressed at low energies. "High" energies here should be compared to the mass energy of a particle, so that E is high only for highly relativistic scenarios. Thus, these complicated terms do not influence our low-energy environment. Throughout this book, we will therefore limit ourselves to as simple Lagrangians as we can find.

Charge Parity Time Symmetry

Another crucial result of the QFT principle of least action is the CPT theorem, which states that if the charge of every particle, the parity of the coordinate system, and the direction of time are all reversed, then the predictions of the QFT principle of least action are unchanged, no matter what the Lagrangian is. This theorem was not understood to be important until well into the development of QFT, but it aids in understanding many of QFT's quirks, including the presence of antiparticles. A proof is not given here, but can be understood from other sources after first gaining some familiarity with particles and spin.

    The next part of this book is dedicated to understanding the n-point correlation functions, which are the left hand side of (1.15), while the second part focuses on evaluating the right hand side. The third and fourth

parts focus mostly on which Lagrangian Nature plugs into (1.15).

Part I

Non-Perturbative Techniques

This part is devoted to understanding n-point correlation functions, which are the output of QFT predictions. These functions have two major uses: one is to compute scattering cross-sections and decay probabilities (chapter 2) while another is to compute the spectrum of a quantum particle (chapter 3). In these two chapters we will introduce important tools, but save actual predictions for later parts, where we introduce new Lagrangians and use them to compute the n-point correlation functions.

Chapter 2

Scattering

As defined in chapter 1, the quantum state $\hat{\phi}(x)\,|0\rangle$ represents a ϕ particle created at spacetime position x. The Heisenberg uncertainty principle requires that this particle therefore have no definite momentum. If we wished, we could switch to the momentum basis by defining a new operator

$$\hat{\phi}(k) = \int d^4x\, e^{-ik\cdot x} \hat{\phi}(x) \tag{2.1}$$

which is the Fourier transform of $\hat{\phi}(x)$. Note that $k \cdot x$ is the dot product of four-vectors, or $k^0 x^0 - \boldsymbol{k}\cdot\boldsymbol{x}$. Under this definition, $\hat{\phi}(k)\,|0\rangle$ represents a particle with fixed momentum k, and this time with indefinite position.

This plane-wave configuration corresponds physically to a long beam of particles of fixed momentum, such as a laser. These beams are the incoming and outgoing states of a scattering experiment, which is the natural tool used to understand the quantum nature of particles. We will therefore use the momentum basis for most of this chapter.

2.1 S Matrix and Scattering Amplitudes

Suppose a scattering experiment's incoming beam contains n particles with four-momentum $p_1\ p_2, p_3, \ldots, p_n$. The above notation allows us to

write this incoming state as

$$|\text{in}\rangle = \hat{\phi}(p_1) \cdots \hat{\phi}(p_n) |0\rangle . \tag{2.2}$$

and likewise we can name the outgoing state

$$|\text{out}\rangle = \hat{\phi}(q_1) \cdots \hat{\phi}(q_m) |0\rangle . \tag{2.3}$$

Note that the number of incoming particles n need not equal the number of outgoing particles because in a relativistic theory, particles can be converted to energy unlike in non-relativistic quantum mechanics.

The overlap between the input and output states is usually called the S matrix, defined with an extra factor of i due to convention.

$$iS = \langle \text{out}|\text{in}\rangle = \langle 0|\phi(q_m)^\dagger \cdots \hat{\phi}(q_1)^\dagger \hat{\phi}(p_1) \cdots \hat{\phi}(p_n)|0\rangle . \tag{2.4}$$

The magnitude of the S matrix is an observable quantity because the probability of a scattering event with these initial and final momenta is $|S|^2$. The right hand side is also important because it is an $(n + m)$PCF, which the QFT principle of least action tells us how to compute.

If no interactions occur, we expect no momentum transfer, so $n = m$ and $p_j = q_j$ for all j. In other words, $iS \propto \prod_i \delta(p_j - q_j)$. Even when interactions are possible, many particles will still not interact and will pass through the beam unchanged. We therefore define a new quantity T such that

$$iS_{p \to q} = \prod_i \delta(p_j - q_j) + iT_{p \to q} \tag{2.5}$$

where the T-matrix is the "off-diagonal" part of the S-matrix, which quantifies the scattering interactions.

Even when interactions are present, $\sum_i p_i = \sum_j q_j$ always holds due to momentum and energy conservation. To extract this fact we make one more definition:

$$iT_{p \to q} = (2\pi)^4 \delta \left(\sum_i p_i - \sum_j q_j \right) iM_{p \to q} \tag{2.6}$$

where M is called the *scattering amplitude*[1]. M is the interesting physical quantity behind scattering experiments; its value is set by the fundamental physics rather than mere conservation laws or unscattered particles. Most of the rest of this book will therefore compute the scattering amplitude and stop there.

There are some additional complexities which we have glossed over but will address later. One is that momentum is not the other measurable property of a particle. Other conserved quantities, such as fermion spin or photon polarization, can be observed and should therefore be included in the initial and final states. We neglected these additional quantities for the sake of clarity and will add them back in when fermions and photons are discussed.

The reader may also be curious about the properties of this $\hat{\phi}$ operator. It's definition and its action against $|0\rangle$ have been given, but other details such as the commutator $[\hat{\phi}(x), \hat{\phi}(y)]$ and the structure of the Hilbert space it belongs to have been omitted. This is because we will rarely use the $\hat{\phi}$ operators themselves. They were useful to derive the QFT principle of least action and to understand why $|S|^2$ is the probability of a scattering experiment, but now we can forget about the operators and think in terms of the S matrix and scattering experiments. They will only appear again when we discuss Fermi's golden rule and second quantization in section 9.1.

2.2   Decay Widths and Cross Sections

It's worth making the connection between scattering amplitudes and experiment more explicit. The connection is quite simple except for the fact that scattering amplitudes are Lorentz invariant and experiments, which

---

[1]It's worth noting that M is Lorentz invariant because the original definition of the S matrix is invariant as long as the states $|in\rangle$ and $|out\rangle$ are invariant, which is ensured as long as the states are normalized correctly. The factors we subtracted off the S matrix to form M are also Lorentz invariant because they are delta functions of four-vectors

are done at fixed velocities and orientations, are not. Scattering amplitudes and the outcome of scattering experiments can therefore differ by a energy-dependent coefficient, which we will work out by normalizing the incoming and outgoing plane waves.

Let's consider a single particle of mass m scattering into n final particles, with momentum q_j. A single-particle scattering event is also known as particle decay, and the measurable quantity is the decay width Γ which is the probability of scattering per unit time in the rest frame of the particle. The lifetime τ of the particle is related to Γ by $\tau = 1/\Gamma$.

Converting the momentum-space fields $\hat{\phi}(p)$ back to position-space via the inverse Fourier transform gives us

$$\hat{\phi}(x) = \int \frac{d^4 k}{(2\pi)^4} \, e^{-ik \cdot x} \delta(k^2 - m^2) \hat{\phi}(k). \tag{2.7}$$

The delta function can be inserted with no penalty because of the relativistic constraint that $k^2 = m^2$, where m is the particle mass. This constraint is named the *on-shell constraint*, so called because the surface $k^2 = m^2$ looks like a hyperboloid of two sheets in k-space. That is, a "shell".

If we limit our focus to the incoming and outgoing beams, we can replace $\hat{\phi}(x)$ with the wavefunction $\psi(\boldsymbol{x})$. This is because we already understand the state $\hat{\phi}(x)\,|0\rangle$ — the single particle state — in non-relativistic quantum mechanics by describing its wavefunction as a plane wave. During a scattering experiment, we cannot rely on plane wave knowledge from non-relativistic quantum mechanics and must instead rely on the principle of least action, which converts the operators to a path integral of functions ϕ rather than a single wavefunction.

Performing this replacement,

$$\psi(t, \boldsymbol{x}) = \int \frac{d^4 k}{(2\pi)^4} \, e^{i\boldsymbol{k} \cdot \boldsymbol{x}} e^{-ik^0 t} \delta(k^2 - m^2) \psi(\boldsymbol{k}). \tag{2.8}$$

Since experiments measure the spatial part of momentum, we'll integrate out the k_0 component. The fact that

$$\int dx\,\delta(f(x) - f(z)) = \frac{2\pi}{|f'(z)|}\delta(x - z), \tag{2.9}$$

leads to

$$\psi(\boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\,e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\frac{e^{-iE_{\boldsymbol{k}}t}}{\sqrt{2E_{\boldsymbol{k}}}}\psi(\boldsymbol{k}) \tag{2.10}$$

where $E_{\boldsymbol{k}} = k_0 = \sqrt{m^2 + \boldsymbol{k}^2}$ is the energy of the particle at momentum $\boldsymbol{k}$.

Wavefunctions must be normalized, meaning

$$1 = \int d^3\boldsymbol{x}\,|\psi(\boldsymbol{x})|^2 = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2E_{\boldsymbol{k}}}|\psi(\boldsymbol{k})|^2 \tag{2.11}$$

The lesson to take from this is that the observed wavefunction $|\psi(\boldsymbol{k})|^2$ differs from Lorentz invariant operators $\hat{\phi}(k)$ by a factor of $1/2E_{\boldsymbol{k}}$. Thus, the probability to decay into the exact state of $q_1, \ldots, q_m$ per unit time is

$$d\Gamma = \left[\frac{1}{2E_p}\right]\left[\prod_j \frac{d^3q_j}{(2\pi)^3}\frac{1}{2E_{q_j}}\right]\left[|M_{p\to q}|^2(2\pi)^4\delta\left(p - \sum_j q_j\right)\right] \tag{2.12}$$

The full decay rate integrates over all possible final states to give the true lifetime of the particle:

$$\Gamma = \int d\Gamma$$
$$= \frac{1}{2E_p}\int \frac{d^3q_1}{(2\pi)^3(2E_{q_1})}\cdots\int \frac{d^3q_m}{(2\pi)^3(2E_{q_m})}|M_{p\to q}|^2\delta\left(p - \sum_j q_j\right) \tag{2.13}$$

Similar techniques can be used to find the probability of interaction for multiple to multiple particle scattering; in this case, the probability is called a cross section σ. The differential cross section $d\sigma$ is very similar to

the equation for $d\Gamma$, except that there may be additional coefficients due to the fact that multiple momenta $\boldsymbol{p}_i$ are now unknown. For example, for two incoming particles with velocity v_1 and v_2, the differential cross section is

$$d\sigma = \frac{1}{|\boldsymbol{v}_1 - \boldsymbol{v}_2|} \left(\prod_i \frac{d^3 p_i}{(2\pi)^3} \frac{1}{2E_{\boldsymbol{p}_i}} \right) \left(\prod_j \frac{d^3 q_j}{(2\pi)^3} \frac{1}{2E_{\boldsymbol{q}_j}} \right)$$

$$\times |M_{p \to q}|^2 (2\pi)^4 \delta \left(\sum_u p_i - \sum_j q_j \right).$$

(2.14)

These cases of 1 and 2 particles cover most scattering events. To evaluate the total probability of scattering, $d\Gamma$ or $d\sigma$ may be integrated over all momenta. If identical particles A and B are present in the initial and final states, the states where A and B are switched should count only once. Thus, for k types of m_j identical particles, we should divide the cross section by $\prod_{j=1}^{k} m_j!$.

Chapter 3

Two Point Correlation Function Spectrum

The nPCFs were connected to scattering in the previous chapter, but the 2PCF in particular has an intriguing property known as the Källén-Lehmann spectral representation which we will remark in this chapter. Namely, the 2PCF $\langle \hat{\psi}(p) \hat{\psi}^\dagger(q) \rangle$ summarizes all of the important properties of the particle ψ, including its mass, decay rate, bound states, and scattering.

Let's build intuition for $\langle 0 | \hat{\psi}(p) \hat{\psi}^\dagger(q) | 0 \rangle$ by treating it in a scattering scenario as an S-matrix element. It would represent the probability for a particle to scatter from momentum p to momentum q, meaning $q = p$ by conservation of momentum. The fact that only one particle is involved makes this 2PCF an intrinsic measurement of the properties of a particle. It does not depend on external particles coming close and scattering; it only measures a particle's only self-interactions.

Due to Lorentz invariance, $\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle$ cannot depend on the four-vector p^μ. It can only depend on the Lorentz invariant quantity p^2 and constants — that is, $\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle$ is a function of one variable p^2. We expect the on-shell constraint to force $p^2 = m^2$, meaning that $\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle$ should blow up at $p^2 = m^2$ in order to suppress $p^2 \neq m^2$ effects. The

simplest way to blow up in this way is

$$\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle = \frac{i}{p^2 - m^2}. \tag{3.1}$$

To measure deviations from this expected blow-up, we define the *spectral-density* $\rho(\mu^2)$ such that

$$\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle = \int d(\mu^2) \rho(\mu^2) \frac{i}{p^2 - \mu^2}. \tag{3.2}$$

The on-shell constraint requires $\rho(\mu^2)$ to be infinite at $\mu = m$, and if $\langle \hat{\psi}(p) \hat{\psi}^\dagger(p) \rangle$ blows up in the simple way we also expect ρ to be zero elsewhere, so $\rho(\mu^2) = \delta(\mu^2 - m^2)$.

We have produced this equation only using our interpretation of $\langle \hat{\psi}(p) \hat{\psi}^\dagger(q) \rangle$ as an S matrix element and requiring the on-shell constraint. It turns out that (3.1) is exactly true for a theory with no interactions, as we will see in the next chapter. Thus, $\rho(\mu^2) = \delta(\mu^2 - m^2)$ is a prediction of that theory.

Interactions add extra detail to the spectral density. For example, interactions between the electron and the electric field allow bound states such as hydrogen to exist. In a purely electron theory, electrons and positrons can still combine into a hydrogen-like bound state called positronium, which has energy $E = 2m_e - \delta$, where $\delta = 6.8\,\text{eV}$ is a small energy shift due to the binding energy of the state. We would expect a small Dirac spike in $\rho(\mu^2)$ representing that this state can occur. Excited states of positronium would also appear with smaller values of δ.

As we will discuss, positronium decays usually into two gamma rays with a lifetime of $\tau = 0.12\,\text{ns}$. The sometimes-discussed energy-time uncertainty principle[1] implies that this limited lifetime causes the energy of

---

[1]If you haven't heard of the energy-time uncertainty principle, it arises from the same place as the Heisenberg momentum-position uncertainty principle. Momentum and position are intrinsically related because the fact that physics is invariant under translation is what causes momentum to be conserved. Similarly, the fact that physics is invariant under time translation causes energy to be conserved, giving rise to another uncertainty relation for energy and time.

the positronium ground state to be uncertain. It is difficult to understand this statement precisely in the context of non-relativistic quantum mechanics, but in QFT it manifests as blurring out the Dirac spike in $\rho(\mu^2)$ into a peaked function known as the *Breit-Wigner distribution*

$$\rho(p^2) \propto \frac{1}{(p^2 - E^2)^2 + E^2\Gamma^2} \tag{3.3}$$

where $\Gamma = \frac{1}{\tau} = \frac{1}{0.12\,\text{ns}}$ is the state decay width.

Some particles in the standard model are themselves unstable, such as the neutron which has a lifetime of 15 minutes. For these particles all the Dirac spikes are blurred into the Breit-Wigner distribution, including the $p^2 = m^2$ pole. When detecting particles in a collider, this Breit-Wigner distribution is observed and its peak is sometimes used to define the mass of very short-lived particles.

JTD: Scattering and spectrum beyond $2m$

Chapter 4

Lattice Quantum Field Theory

The past few chapters dealt with connecting observables such as scattering amplitudes and particle creation to the n-point correlation functions. The most common way to actually evaluate these functions is to use the perturbative techniques discussed in the rest of this book, but in this chapter we briefly mention the alternative methods in case the perturbative approach should fail.

Part II

Quantum Electrodynamics & Abelian Gauge Theories

Quantum Electrodynamics, or QED, is the theory of the electron and the photon, and can also be extended to approximately describe the proton. It is the first example of a successful quantum field theory and remains one the most accurate theory measured.

QED contains two types of particles: fermions and gauge bosons. This part will be dedicated to discussing both types of particle, as well as a simpler particle known as a scalar boson. The difference between these particles is codified in what *quantum numbers* they hold.

A quantum number is a quantity which is conserved as the particle propagates. From Newtonian mechanics, we are aware of momentum as a conserved property — and therefore a quantum number — that particles should have. However, quantum mechanics adds other possibilities such as spin for electrons, polarization for photons, helicity for massive particles, and more. Any particle theory must provide a complete description for the evolution of each of these quantum numbers.

Chapter 5

Spin Zero

Our particle will be the simplest kind of particle: a *scalar boson* or *spin-zero particle*, whose only quantum number is momentum. The standard model — the complete theory of particle physics — contains one of these scalars in the form of the Higgs particle, though there are additional particles such as mesons which can be approximated as scalars, as we will later discuss.

The first step in creating this scalar is to write a Lagrangian for a field ϕ defined at every spacetime point x. Let's assume for now that the field takes on real values. Symmetries place the following severe limit the terms this Lagrangian may contain:

1. Translations symmetry forbids terms that contain coordinates like x or t

2. Lorentz symmetry forces forbids uncontracted spacetime indices

3. The Hamiltonian (and Lagrangian) must have a minimum energy so the field does not fly off to infinity

4. If ϕ were complex, the fact that action is real would forbid any complex-valued term, which includes any odd power of ϕ

Furthermore, in the low energy regime where we might expect ϕ to be small (few particles present) and $\partial_\mu \phi$ to be small (low momentum), we should not include terms which involve too many powers of ϕ. The exact cutoff for how many powers is "too many" will be discussed later.

Given the above constraints on the Lagrangian, the only available terms to the Lagrange density $\mathcal{L}$ are

$$\phi \partial_\mu \partial^\mu \phi, \qquad (\partial_\mu \phi)(\partial^\mu \phi), \qquad \phi^2, \qquad \phi^3, \qquad \phi^4,$$

and higher order terms. However, the first two terms are identical. This is because the QFT principle of least action (1.15) only involves $\int d^4x \mathcal{L}$ rather than $\mathcal{L}$ directly. When under an integral, integration by parts guarantees that $\phi \partial_\mu \partial^\mu \phi = -(\partial_\mu \phi)(\partial^\mu \phi)$. For this reason, we drop the first term with no loss of generality. We also usually drop the last term, but we will add it back in for a later section.

This discussion leaves us with only one possible action at low order:

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 - \frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4. \tag{5.1}$$

The constants m^2 and λ were inserted to preserve generality, but no constant was inserted before the first term because such a constant can always be set to $1/2$ by rescaling ϕ. Each term in (5.1) as an intuitive meaning which we will briefly describe now, and the later sections will elucidate them in much greater detail.

1. $-\frac{1}{2}(\partial_\mu \phi)^2$ is known as the *kinetic term*, defined to be negative because, as will be seen later, if it were positive the n-point correlation functions would all diverge. This term encapsulates the kinetic energy of the particle because derivatives of fields represent momentum in much the same way that the quantum mechanical definition of the momentum operator $\hat{p}$ is $i\nabla$. The presence of two derivatives is akin to the p^2 of kinetic energy.

2. $\frac{m^2}{2}\phi^2$ is known as the *mass term*. Since ϕ^2 is positive for any nonzero phi and is larger than any other power of ϕ, this term controls the amount of energy that a particle contains merely by existing, and thereby increasing the local value of ϕ. This inherent energy is defined as its mass. These first two terms of the Lagrangian are very similar to the Hamiltonian quantum harmonic oscillator, which has energies of $E = m(n + \frac{1}{2})$, where n is an integer. Thus, energy can only increase in integer multiples of mass, which is an intuitive result. We will see that the same result occurs in quantum field theory.

3. $\frac{\lambda}{4!}\phi^4$ is an *interaction term*. Since it is proportional to the square of the mass term, it represents two-body interactions, whereas ϕ^5 would represent three-body interactions, etc. The presence of this term ruins the harmonic oscillator-like spectrum of $E = m(n + \frac{1}{2})$ because it generates potential energy when two particles are present in the same system. This is reminiscent one of the $1/r$ Coulomb potential of a charged particle, and indeed we will see that $\lambda\phi^4$ produces a $1/r$ potential between two ϕ particles.

Encouragement to pursue (5.1) comes from the fact that, if we set $\lambda = 0$ and treat $\mathcal{L}$ as ϕ as a classical field theory, it produces the equation of motion

$$\partial^\mu \partial_\mu \phi^2 - m\phi^2 = 0 \tag{5.2}$$

which is the famous Klein-Gordon equation mentioned in the introduction (1.3). It was already known to satisfy the laws of special relativity, and if $m = 0$, it reduces to the familiar wave equation. The scalar Lagrangian (5.1) is therefore a largely familiar beast.

5.1   Solving the Scalar Path Integral Without Interactions

In this section, we will solve for n-point correlation functions $\langle \hat{\phi}(x_1) \cdots \hat{\phi}(x_n) \rangle$ using the QFT principle of least action (1.15). We neglect the ϕ^4 term by setting $\lambda = 0$ in this section, but we will restore λ in the following section. As such, the result of this section will be correlation functions for a non-interacting scalar theory.

Even with $\lambda = 0$, evaluating the path integral is an intricate process that requires several new mathematical techniques. We introduce these techniques in this section, which ends with a formula for the nPCFs of non-interacting scalar theory. The physical meaning of this formula is discussed in the following section.

Plugging the scalar Lagrangian into (1.15),

$$\langle 0 | \hat{\phi}(x_1) \cdots \hat{\phi}(x_n) | 0 \rangle = \frac{\int \mathcal{D}\phi \, \phi(x_1) \cdots \phi(x_n) e^{-\frac{i}{2} \int d^4x \left[-(\partial_\mu \phi)^2 - m\phi^2 \right]}}{\int \mathcal{D}\phi \, e^{-\frac{i}{2} \int d^4x \left[-(\partial_\mu \phi)^2 - m\phi^2 \right]}}.$$

(5.3)

The ϕ fields in the exponent are functions of x, but the (x) has been suppressed for clarity.

It appears that if we can calculate the numerator for any $x_1 \ldots x_n$, then we can calculate the denominator too and therefore any nPCF. We'll start with the case where $n = 0$:

$$\int \mathcal{D}\phi \, e^{-\frac{i}{2} \int d^4x \left[-(\partial_\mu \phi)^2 - m\phi^2 \right]}$$

(5.4)

Note that both terms in the exponent are quadratic in ϕ. We can therefore separate the ϕs as follows:

$$\int \mathcal{D}\phi \, e^{-\frac{i}{2} \int d^4x \, \phi^* \left[-\partial^2 - m \right] \phi}.$$

(5.5)

(5.5) is a Gaussian integral because the exponent is quadratic in ϕ. This makes it one of the only path integrals which can be solved ana-

lytically with perfect accuracy, and we will therefore spend considerable time on it. The next subsection is dedicated to evaluating it, while the following generalizes the result to $n \neq 0$. Following this, we will return to computing nPCFs.

Infinite Dimensional Gaussian Integrals

The Gaussianity of (5.5) can be seen more clearly if we consider a one-dimensional spacetime, broken up into a lattice with N sites ans spacing a. In this case, $\phi(x_i)$ becomes ϕ_i, $\int d^4x$ becomes $\sum_i a^4$ and $\partial^2\phi(x_i)$ becomes $(\phi_{i-1} + \phi_{i+1} - 2\phi_i)/a^2$. Furthermore, the path integral breaks down as an integral over each ϕ_i, so that the entire function becomes

$$\prod_i^N \int d\phi_i \, e^{-\frac{i}{2}\sum_{jk}\phi_j^* M_{jk}\phi_k} = \prod_i^N \int d\phi_i \, e^{-\frac{i}{2}\phi^\dagger M \phi} \tag{5.6}$$

where

$$M_{ij} = -ma^4\delta_{ij} + a^2(\delta_{i,j+1} + \delta_{i,j-1} - 2\delta_{ij}) \tag{5.7}$$

is a matrix. In the second equality (5.6), we have switched to matrix notation, where ϕ is a vector of ϕ_i. Inspection reveals that M is Hermitian and therefore can be diagonalized as $M = U^\dagger \Lambda U$, where U is a unitary matrix and Λ is a diagonal matrix of eigenvalues λ_i. Thus, the path integral becomes

$$\int d^N\phi \, e^{-\frac{i}{2}(U\phi)^\dagger \Lambda(U\phi)} = \int d^N\phi \, e^{-\frac{i}{2}\phi^\dagger \Lambda \phi} = \prod_i^N \left[\int d\phi_i \, e^{-\frac{i}{2}\phi_i^*\lambda_i\phi_i} \right]. \tag{5.8}$$

In the first equality we changed variables $\phi \to U\phi$, but since $\det U = 1$, doing so does not change the value of the integral. In the second equality we took advantage of the fact that Λ is diagonal to integrate each component separately, leaving us with N single-variable Gaussian integrals. It's

a common fact that

$$\int dx\, e^{-\alpha x^2} = \sqrt{\pi \alpha^{-1}} \qquad \text{whenever} \qquad \text{Re}\,\alpha > 0. \qquad (5.9)$$

Here, $\alpha = i\lambda_i/2$, and since M is Hermitian, λ is real and α is purely imaginary. We are therefore stuck on the edge of the validity of (5.9). A common choice is to add a small imaginary number $i\epsilon$ to every eigenvalue so that Re α is always slightly positive. Then our integral becomes

$$\prod_i^N \int d\phi_i\, e^{-\frac{i}{2} \sum_{jk} \phi_j^* M_{jk} \phi_k} = \frac{\pi^{N/2}}{\prod_i^N \sqrt{\lambda_i}} = \frac{\pi^{N/2}}{\sqrt{\det M}}. \qquad (5.10)$$

The next step is to remove the discretization we performed at the beginning, by dividing spacetime into a lattice. This involves sending $N \to \infty$ and also changing $\det M$ such that the value of the above integral may become infinite. This is worrying, but not catastrophic, since the nPCFs of (5.3) are all *ratios* of gaussian integrals. If the $\pi^{N/2}/\sqrt{\det M}$ is common to both the numerator and the denominator, it will cancel out regardless of whether it is infinite.

The next step is therefore to generalize our Gaussian integral result to $n \neq 0$ so that we can confirm this cancellation in the nPCFs.

Wick's Theorem

The last barrier preventing us from evaluating the properties of non-interacting scalar theory is the integral

$$\int \mathcal{D}\phi\, \phi(x_1) \ldots \phi(x_n) e^{-\frac{i}{2} \int d^4 x\, \phi^* \left[-\partial^2 - m \right] \phi}. \qquad (5.11)$$

We'll use a trick to solve this problem. First, we'll discretize spacetime as we did in the previous section:

$$\prod_i^N \int d\phi_i\, \phi_{j_1} \ldots \phi_{j_n} e^{-i\, \phi^\dagger M \phi}. \qquad (5.12)$$

where ϕ_{j_1} is the discretized $\phi(x_1)$. This integral can be solved more easily via the *generating function*

$$Z(\boldsymbol{J}) = \prod_i^N \int d\phi_i \, e^{-i\,\phi^\dagger M\phi + \boldsymbol{J}\cdot\phi} \tag{5.13}$$

where $\boldsymbol{J}$ is an arbitrary vector. This function exists only to be differentiated, since

$$\frac{d^2 Z}{dJ_{j_1} dJ_{j_2}}\bigg|_{\boldsymbol{J}=0} = \prod_i^N \int d\phi_i \, \phi_{j_1} \phi_{j_2} e^{-i\,\phi^\dagger M\phi} \tag{5.14}$$

which can be seen by explicitly differentiating (5.13). This derivative is clearly useful, since it is the integral need for a 2PCF. We therefore go about finding $Z(\boldsymbol{J})$ via the same method as in the previous section:

$$
\begin{aligned}
Z(\boldsymbol{J}) &= \int d\phi \, e^{-i\left[(U\phi)^\dagger \Lambda (U\phi) + (U\boldsymbol{J})^\dagger (U\phi)\right]} \\
&= \prod_i \left[\int d\phi_i \, e^{-i\left[\phi_i^* \lambda_i \phi_i + (U\boldsymbol{J})_i^* \phi_i\right]} \right] \\
&= \prod_i \left[\int d\phi_i \, e^{-i\left|\phi_i \sqrt{\lambda_i} + \frac{(U\boldsymbol{J})_i^*}{2\sqrt{\lambda_i}}\right|^2} e^{i\frac{|U\boldsymbol{J}|_i^2}{4\lambda_i}} \right] \\
&= \prod_i \left[\sqrt{\frac{\pi}{\lambda_i}} e^{i\frac{|U\boldsymbol{J}|_i^2}{4\lambda_i}} \right] \\
&= \frac{\pi^{N/2}}{\det M} e^{\frac{i}{4}\sum_i \Lambda_i^{-1}|U\boldsymbol{J}|_i^2} = \frac{\pi^{N/2}}{\det M} e^{\frac{i}{4}\boldsymbol{J}^\dagger U^\dagger \Lambda^{-1} U\boldsymbol{J}} \\
&= \frac{\pi^{N/2}}{\det M} e^{\frac{i}{4}\boldsymbol{J}^\dagger M^{-1}\boldsymbol{J}}
\end{aligned}
\tag{5.15}
$$

where the first equality is true by diagonalizing the Hermitian matrix M, the second by substitution of variables, the third by completing the square, the fourth by substitution of variables again and evaluating the Gaussian integral, and the rest by rearranging the equation. This calculation revealed two crucial properties: (1) that $Z(\boldsymbol{J})$ has the same infinite

prefactor that the original Gaussian integral in the last section had, and (2) this prefactor is multiplied by a simple term containing M^{-1}. This term is simple enough that its derivative can be easily computed:

$$\frac{d^2 Z}{dJ_{j_1} dJ_{j_2}}\bigg|_{\boldsymbol{J}=0} = Z(0) M^{-1}_{j_1 j_2}. \qquad (5.16)$$

Higher derivatives are also simple: two more derivatives multiplies the above by $M^{-1}_{j_3 j_4}$. However, since derivatives commute, $Z(0) M^{-1}_{j_1 j_2} M^{-1}_{j_3 j_4}$ cannot be the only term. The value of the integral must be symmetric under all permutations of the indices. Therefore,

$$\frac{d^2 Z}{dJ_{j_1} dJ_{j_2} dJ_{j_3} dJ_{j_4}}\bigg|_{\boldsymbol{J}=0} = Z(0) \left(M^{-1}_{j_1 j_2} M^{-1}_{j_3 j_4} + M^{-1}_{j_1 j_3} M^{-1}_{j_2 j_4} + M^{-1}_{j_1 j_4} M^{-1}_{j_2 j_3} \right).$$
$$(5.17)$$

For n-many derivatives with n even, there are $(n)!$ terms where each term represents one possible way to pair all the indices up. This rule for evaluating Gaussian integrals is called *Wick's theorem*.[1] If n is odd, there are no ways to pair up the indices and the value of the Gaussian integral is zero. This can also be seen by noticing that the Gaussian is an even function, and an odd number of derivatives applied to an even function is an odd function which will take the value zero at $\boldsymbol{J} = 0$.

Remembering the connection between Z and the PCFs, as in (5.14), Wick's theorem states that

$$\langle 0|\hat{\phi}(x_{j_1}) \cdots \hat{\phi}(x_{j_{2n}})|0\rangle = \sum_{\text{pairs } (a,b)_i \text{ of indices}} M^{-1}_{a_1 b_1} M^{-1}_{a_2 b_2} \ldots M^{-1}_{a_n b_n}. \quad (5.18)$$

We have successfully computed our first nPCF! The next section is devoted to returning to a continuous, four-dimensional spacetime, after which we will discuss the consequences of this nPCF.

---

[1]Problem 1 is to prove this fact by induction.

Continuous Spacetime Gaussian Integrals

When spacetime is continuous, the matrix M used profusely in the preceding sections does not make sense, meaning that the M^{-1} of Wick's theorem is not defined. For scalar the free theory, we had $\mathcal{L} = \phi^*(-\partial^2 - m^2)\phi/2$, so the analog of the matrix M will be the differential operator $O_x = (-\partial^2 - m^2)/2$, where the subscript x indicates that O contains derivatives with respect to x. The analog of M^{-1} is the function $G(x, y)$ which satisfies

$$O_x G(x, y) = \delta(x - y). \tag{5.19}$$

This is analogous to the definition of a matrix inverse, which is that

$$M_{ij} M_{jk}^{-1} = \delta_{ik}. \tag{5.20}$$

The function $G(x, y)$ is called the Green's function for operator O and must be computed for each O that may appear. Computing a Green's function for an arbitrary operator is generally difficult, but our operator $O_x = (-\partial^2 - m^2)$ has translation symmetry, which makes its Green's function easier to compute. Translation symmetry is manifested in the fact that O_x not a function of x, but we could have also guessed it would appear from the fact that the Lagrangian density from which O_x is derived had to be translation invariant. This symmetry requires that G must also be translation invariant, so that $G(x, y) = G(x - y_0, y - y_0)$ for any y_0. Choosing $y_0 = y$ allows us to write $G(0, x - y)$, or $G(x - y)$ if we drop the first argument. Thus, the new definition of the Green's function can be simplified to

$$O_z G(z) = \delta(z) \tag{5.21}$$

for $z = x - y$.

Translational symmetry allows us to solve for G_x via the Fourier transform. Let's define the Fourier transform of $G(z)$ as

$$G(k) = \int d^4z \, e^{-ik \cdot z} G(z), \qquad G(z) = \int \frac{d^4k}{(2\pi)^4} e^{ik \cdot z} G(k), . \tag{5.22}$$

We must also transform O_x such that it can operate on $G(k)$. Constants such as m are unaffected by a Fourier transform, but derivatives such as ∂_μ do transform:

$$\partial_\mu f(x) = \int \frac{d^4 k}{(2\pi)^4} \partial_\mu e^{ik \cdot x} f(k) = \int \frac{d^4 k}{(2\pi)^4} ik_\mu e^{ik \cdot x} f(k). \qquad (5.23)$$

These two substitutions for $G(x)$ and ∂_μ transform (5.21) to

$$\int \frac{d^4 k}{(2\pi)^4} \left(k^2 - m^2\right) G(k) = \delta(z). \qquad (5.24)$$

A property of the δ function is

$$\delta(z) = \int \frac{d^4 k}{(2\pi)^4} e^{ik \cdot z}. \qquad (5.25)$$

Substituting this property into (5.24) causes the k integral to appear on both sides of the equation. Removing it, we get

$$(k^2 - m^2) G(k) = 1 \implies G(k) = \frac{1}{k^2 - m^2}. \qquad (5.26)$$

This concludes our calculation of the Green's function. The analog of Wick's theorem (5.18) for Green's functions is

$$\langle 0|\hat{\phi}(x_1) \cdots \hat{\phi}(x_{2n})|0\rangle = \sum_{\text{pairs } (a,b)_i \text{ of indices}} G(x_{a_1} - x_{b_1}) \ldots G(x_{a_n} - x_{b_n}).$$

$$(5.27)$$

and nPCFs with n odd are zero.

Putting the Pieces Together

In order to use our momentum-space Green's function (5.26) we must Fourier transform the nPCF:

$$
\begin{aligned}
\langle 0|\hat{\phi}(k_1)\cdots\hat{\phi}(k_{2n})|0\rangle &= \int d^4x_1\ldots d^4x_{2n}\left(e^{ik_1\cdot x_1}\ldots e^{ik_{2n}\cdot x_{2n}}\right)\\
&\quad\times\langle 0|\hat{\phi}(x_1)\cdots\hat{\phi}(x_{2n})|0\rangle\\
&= \int d^4x_1\ldots d^4x_{2n}\left(e^{ik_1\cdot x_1}\ldots e^{ik_{2n}\cdot x_{2n}}\right)\\
&\quad\times\sum_{\text{pairs }(a,b)_i}G(x_{a_1}-x_{b_1})\ldots G(x_{a_n}-x_{b_n})\\
&= \sum_{\text{pairs }(a,b)_i}\int d^4x_1\ldots d^4x_{2n}\left(e^{ik_1\cdot x_1}\ldots e^{ik_{2n}\cdot x_{2n}}\right)\\
&\quad\times G(x_{a_1}-x_{b_1})\ldots G(x_{a_n}-x_{b_n})
\end{aligned}
$$

$$(5.28)$$

where the second equality came from applying Wick's theorem and the third commuted the sum over pairs to the front of the equation. We can now perform a unitary change of variables. For each Wick pairs of indices $(a,b)_i$, we define $y_i = (x_{a_i}+x_{b_i})/2$ and $z_i = x_{a_i}-x_{b_i}$, so that $d^4x_{a_i}d^4x_{b_i} = d^4y_id^4z_i$. The advantage of this transformation is that $G(x_{a_i}-x_{b_i})$ is now merely $G(z_i)$ and does not depend on y_i. The exponents also transform:

$$
e^{ik_{a_i}\cdot x_{a_i}}e^{ik_{b_i}\cdot x_{b_i}} = e^{ip_i\cdot y_i}e^{i\ell_i\cdot z_i} \tag{5.29}
$$

where $p_i = k_{a_i}+k_{b_i}$ and $\ell_i = (k_{a_i}-k_{b_i})/2$. The result is

$$
\begin{aligned}
\langle 0|\hat{\phi}(k_1)\cdots\hat{\phi}(k_{2n})|0\rangle &= \sum_{\text{pairs }(a,b)_i}\prod_{i=1}^{n}\int d^4y_id^4z_i\,e^{ip_i\cdot y_i}e^{i\ell_i\cdot z_i}G(z_i)\\
&= \sum_{\text{pairs }(a,b)_i}\prod_{i=1}^{n}(2\pi)^4\delta(p_i)G(\ell_i).
\end{aligned}
$$

$$(5.30)$$

The $\delta(p_i)$ term is nonzero only when $k_{a_i} = -k_{b_i}$, in which case $\ell_i = (k_{a_i} - k_{a_i})/2 = k_{a_i}$. Thus,

$$\langle 0|\hat{\phi}(k_1)\cdots\hat{\phi}(k_{2n})|0\rangle = \sum_{\text{pairs } (a,b)_i} \prod_{i=1}^{n}(2\pi)^4\delta(k_{a_i} + k_{b_i})G(k_{a_i}). \quad (5.31)$$

This result is actually valid for any quadratic Lagrangian, even ones which are not the free scalar lagrangian. Plugging in the Green's function for a scalar field,

$$\langle 0|\hat{\phi}(k_1)\cdots\hat{\phi}(k_{2n})|0\rangle = \sum_{\text{pairs } (a,b)_i} \prod_{i=1}^{n}(2\pi)^4\frac{\delta(k_{a_i} + k_{b_i})}{k_{a_i}^2 - m^2}. \quad (5.32)$$

This is our first ever nPCF, computed exactly assuming the QFT principle of least action (1.15) and the scalar Lagrange density (5.1). It depends only on the mass of the ϕ particle m (included in the form of $G(k)$) and the k values chosen to evaluate the nPCF at. In the next section, we will interpret the physical implications of this nPCF formula and show that it has both quantum mechanical and relativistic properties, confirming our hope that the QFT principle of least action would lead to a relativistic quantum theory. We will also discuss the spectrum of the 2PCF using the framework discussed in chapter 3, showing that the 2PCF indeed contains arbitrarily many particles of mass m.

5.2   The Spin Zero Free Theory Is Quantum Mechanical and Relativistic

The nPCFs of (5.32) have several crucial properties which are necessary for a relativistic quantum theory, which we enumerate below.

Mass Energy Equivalence   (5.32) enforces the relativistic $k^2 = m^2$ constraint due to the $k^2 - m^2$ in the denominator of the Green's function. This causes the nPCFs to be much greatest near $k^2 = m^2$ — that is, the probability for a particle with momentum k to be produced is greatest for $k^2 = m^2$ and much smaller elsewhere.

Quantized Amplitude The nPCFs also show that a field $\hat{\phi}(k)$ at a given momentum k cannot be arbitrarily rescaled. Since m is constant and $k^2 = m^2$ as discussed above, there is no freedom to choose an arbitrary amplitude for ϕ and obey the (5.32). That is, the amplitude of ϕ is quantized. This is unlike a classical field theory such as electromagnetism, where the $\boldsymbol{E}$ field may take on whatever amplitude it wishes. But in any quantum mechanical theory, we expect the properties of particles to be quantized, with values such as field amplitude taking on only discrete values.

No Scattering in Free Theory Chapter 2 outline a toolkit to use nPCFs to compute scattering cross-sections. It interpreted

$$\langle 0|\hat{\phi}(p_1)\hat{\phi}(p_n)\cdots\hat{\phi}(-q_1)\hat{\phi}(-q_m)|0\rangle$$

as the amplitude of scattering n particles from momentum p_i to m particles of momentum q_i. (5.32) shows that this amplitude is zero except when the momentum can be paired up as $p_i = q_i$. Each pair adds the same factor to the correlation function due to the quantization of amplitude. This $p_i = q_i$ requirement shows that the free theory nPCFs do not allow particles to change momentum, even when other particles are present. This is equivalent to requiring no interactions, which is the definition of a free theory.

Mass Spectrum of a Particle In chapter 3, we discussed the 2PCF in particular and showed that it is expected to have a pole at $k^2 = m^2$ where m is the mass of any particle in the theory. We also showed that branch cuts where the 2PCF is infinite represent interaction cross-sections. Our 2PCF contains a pole at $k^2 = m^2$, where m is the mass of the ϕ particle, exactly as expected. This is the only pole and there are no branch cuts, indicating again that this theory contains no interactions.

Causality   A founding principle of relativity is that information cannot be translated faster than the speed of light. If (5.32) is to be taken seriously as a theory of physics, we should expect $\hat{\phi}(x)$ to be uncorrelated with $\hat{\phi}(y)$ as long as x and y are spacelike-separated. Thus,

$$\langle 0|\hat{\phi}(x)\hat{\phi}(y)|0\rangle = 0 \qquad \text{whenever} \qquad |\boldsymbol{x} - \boldsymbol{y}| > |x_0 - y_0|. \qquad (5.33)$$

To check this, we will Fourier transform the 2PCF

$$\langle 0|\hat{\phi}(k)\hat{\phi}(-k)|0\rangle = (2\pi)^4 \frac{\delta(0)}{k^2 - m^2}$$

back into position space and check this fact. If we translate to $y = 0$ and rotate until $x^\mu = (t, x, 0, 0)$, then

$$\begin{aligned}
\langle 0|\hat{\phi}(x)\hat{\phi}(0)|0\rangle &= \int \frac{d^4 k}{(2\pi)^4} e^{ik\cdot x} \frac{(2\pi)^4}{k^2 - m^2} \\
&= \int d^3\boldsymbol{k} \int dk^0 \frac{e^{-ik_0 t + ik_1 x}}{k_0^2 - \boldsymbol{k}^2 - m^2}.
\end{aligned} \qquad (5.34)$$

The integrand contains a pole at $k_0^2 = \boldsymbol{k}^2 + m^2$, making the Fourier transform undefined. However, if we add a small $i\epsilon$ term into the denominator where $\epsilon > 0$ is small, we can move the pole off the real axis and out of the way of the integration without changing the value of $G(k)$ very much. Defining $E^2 = \boldsymbol{k}^2 + m^2 > 0$, we must now solve[2]

$$I = \int dk^0 \frac{e^{-ik_0 t}}{k_0^2 - E^2 - i\epsilon} = \int dk^0 \frac{e^{-ik_0 t}}{2E} \left(\frac{1}{k_0 - E + i\epsilon} - \frac{1}{k_0 + E - i\epsilon} \right). \qquad (5.35)$$

This integral is performed over $z \in (-\infty, \infty)$, but we can always add more to the contour of integration (even extending it into the complex

---

[2] $E^2 = \boldsymbol{k}^2 + m^2$ is actually the energy of a particle with mass m and momentum $\boldsymbol{k}$, but we will not need this fact.

Furthermore, in this equation we redefine ϵ on the right hand side to absorb its co-efficients, which is allowed since ϵ is assumed to be small enough that it remains small even when multiplied by a finite number.

plane) as long as the value of the integrand on that contour is zero. For $t > 0$, $e^{-ik_0 t}$ is small whenever $\operatorname{Im} k^0 \ll 0$ and for $t > 0$ the condition is $\operatorname{Im} z \gg 0$. We therefore choose to extend the contour in a large arc around the complex plane, below the real axis for $t > 0$ and above it for $t < 0$ (Figure 5.1). Since this new contour is closed, we may use the Residue Theorem (see chapter 19) to evaluate it, picking up the $-E$ pole when closing the curve above the real line, and the $+E$ pole when closing the curve below the real line. Thus,

$$I = \begin{cases} \pi i \dfrac{e^{-itE}}{E} & \text{for } t > 0 \\[2mm] -\pi i \dfrac{e^{itE}}{E} & \text{for } t < 0 \end{cases} \tag{5.36}$$

The only difference between the $t > 0$ and $t < 0$ cases is an overall sign, which we ignore.

We can now use I to solve (5.34). Dropping proportionality constants which do not matter for this demonstration, and assuming $x > 0$ and $t > 0$,

$$\langle 0 | \hat{\phi}(x) \hat{\phi}(0) | 0 \rangle \propto \int d^3 \boldsymbol{k} \, \frac{e^{i(xk_1 - t\sqrt{\boldsymbol{k}^2 + m^2})}}{\sqrt{\boldsymbol{k}^2 + m^2}}$$

$$= \int dk_2 dk_3 \int dk_1 \, \frac{e^{i(xk_1 - t\sqrt{k_1^2 + F^2})}}{\sqrt{k_1^2 + F^2}} \tag{5.37}$$

where in the second line we defined $F^2 = k_2^2 + k_3^2 + m^2$. We now work on the k_1 integral in a similar fashion to the k_0 integral. Again, we have poles when the denominator is zero, at $k_1^2 = F^2$, or $k_1 = \pm iF$. But this time there is also a branch cut for all $k_1^2 + F^2 < 0$ due to the difficulty of defining the square root operator on negative real values. The complex plane of this function is depicted in Figure 5.2.

To work out whether to close the contour above or below the real line, we substitute $k_1 = Re^{i\theta}$ into the exponent, leaving a real part of $-R \sin\theta (x - t)$. We close the contour in the region where this real part
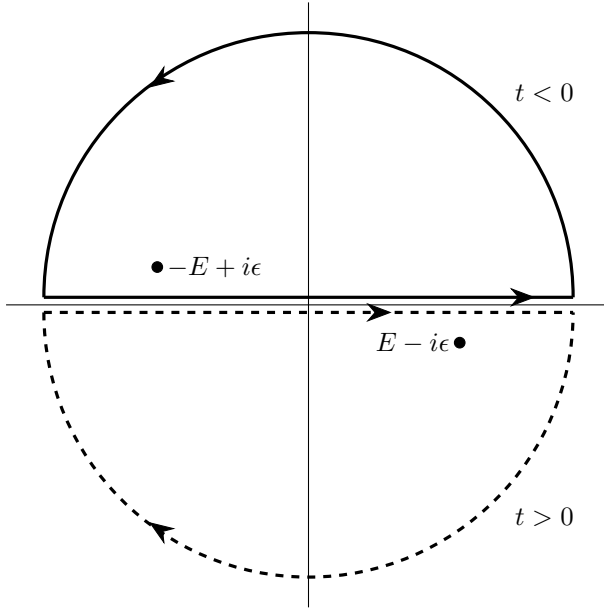
Figure 5.1: The contour to be used for $\int dk^0 \frac{e^{-ik_0 t}}{k_0^2 - A^2 - i\epsilon}$ depending on whether $t > 0$ or $t < 0$.

is small, which is above the real line for $x > t$ and below it for $x < t$. However, we must avoid the branch cut by looping the contour around it, which adds nonzero values where the contour nears the origin. The scale of this added amount is controlled by the value of the integrand near the end of the branch cut, where the integrand is largest. For $x > t$, this occurs for $k_1 = i(F - \delta)$ with δ small, and the integrand is

$$\sim \frac{e^{-xF}}{\sqrt{2F\delta}}. \tag{5.38}$$

Meanwhile, for $x < t$, the integrand near $k_1 = -i(F - \delta)$ is considerably
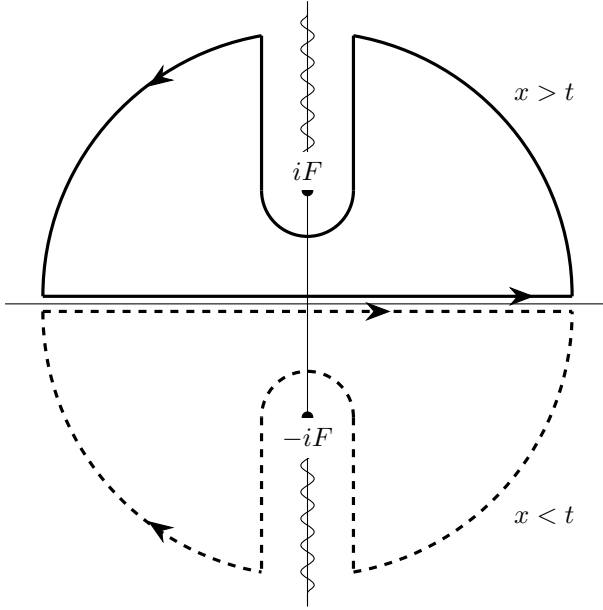
Figure 5.2: The contour to be used for $\int dk_1 \, \frac{e^{i(xk_1 - t\sqrt{k_1^2 + F^2})}}{\sqrt{k_1^2 + F^2}}$, with the branch cut.

larger:

$$\sim \frac{e^{xF}}{\sqrt{2F\delta}}. \tag{5.39}$$

It follows that the 2PCF is nonzero for $x < t$, which is inside the light-cone. But when the lightcone and correlations are forbidden by causality, the 2PCF decays exponentially. This is reminiscent of how a wavefunction decreases exponentially in a region where potential energy is greater than the energy of the particle in non-relativistic quantum mechanics, or of how a light wave exponentially decays in a medium which cannot support its frequency. Thus, the QFT principle of least action not only obeys causality, but does so in a manner similar to both quantum mechanical

and classical field theories.

The exponential decrease outside the lightcone came from our choice to add $i\epsilon$ to the denominator of the Green's function. This shifted the $\pm E$ poles above and below the real line as shown in Figure 5.1. If we had shifted the poles in another way, we would not have gotten the cancellation we achieved, and the 2PCF would not have fallen outside the light cone. This choice of pole-shifting was made by Richard Feynman and is known as the *Feynman Propagator*, and we will nearly always use it in the future.

Due to the success of the QFT principle of least action in this free scalar field case, we are inspired to continue on to the $\lambda \neq 0$ case of an interacting theory, which contains a vast variety of novel phenomena.

5.3   Solving the Spin Zero Path Integral With Interactions

One of the key results of the previous section was that for a non-interacting theory where the Lagrangian density is Gaussian, particles cannot transfer momenta and cannot scatter. The proof of this is (5.30), where we separated an $2n$PCF attained by Wick's theorem into n multiplied factors, each of which enforced momentum conservation.

However, real particles scatter, so to make a true particle theory we will have to break the chain of logic that led us to the no-scattering conclusion. The solution is to add a small, non-quadratic term to the Lagrangian density, which adds small additional terms to the $2n$PCF which are not separable. These small terms allow a small amount of momentum transfer between the n particles, causing scattering. This is the motivation for us associating the $\frac{\lambda}{4!}\phi^4$ term in $\mathcal{L}$ with scattering. In this section, we'll compute the effect of this new term on the momentum-space nPCFs.

Our strategy will be to apply the momentum-transferring ϕ^4 term

as a perturbation to the Gaussian free theory discussed above. In other words, we'll write

$$S = S_0 + \lambda S_1,$$

$$S_0 = \int d^4x \left(\frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 \right), \qquad S_1 = -\int d^4x \frac{1}{4!}\phi^4$$

$$(5.40)$$

Our goal will be to produce a power series in λ for nPCFs, of which the zeroth order term is (5.32) and the following terms are small momentum-mixing terms under the assumption that $\lambda \ll 1$. To do this, we start with the QFT principle of least action:

$$\langle 0|\hat{\phi}(x_1)\cdots\hat{\phi}(x_n)|0\rangle = \frac{\int \mathcal{D}\phi\, \phi(x_1)\cdots\phi(x_n)e^{iS_0+i\lambda S_1}}{\int \mathcal{D}\phi\, e^{iS_0+i\lambda S_1}}. \qquad (5.41)$$

Obtaining a power series in λ is just as simple as expanding the exponent as a power series:

$$\langle 0|\hat{\phi}(x_1)\cdots\hat{\phi}(x_n)|0\rangle = \frac{\int \mathcal{D}\phi\, \phi(x_1)\cdots\phi(x_n)e^{iS_0}\left(1 + i\lambda S_1 - \frac{\lambda^2}{2}S_1^2 + \dots\right)}{\int \mathcal{D}\phi\, e^{iS_0}\left(1 + i\lambda S_1 - \frac{\lambda^2}{2}S_1^2 + \dots\right)}.$$

$$(5.42)$$

Now that only the quadratic term S_0 remains in the exponent of the integrand, (5.42) can be completely solved to arbitrary order with Wick's theorem! For example, when S_1 as defined in (5.40) is substituted into the above equation, the numerator expands to

$$\int \mathcal{D}\phi\, \phi(x_1)\cdots\phi(x_n)e^{iS_0}$$

$$-\frac{i\lambda}{4!}\int d^4x \int \mathcal{D}\phi\, \phi(x_1)\cdots\phi(x_n)e^{iS_0}\phi^*(x)\phi(x)\phi^*(x)\phi(x) + \mathcal{O}(\lambda^2).$$

$$(5.43)$$

Both of these terms are Gaussian correlation functions. The first line is the same nPCF we found for the non-interacting theory, confirming our expectation that adding the small ϕ^4 perturbation did not change the

first-order structure of the theory. The second is an $n + 4$PCFs which will allow these n particles to interact and exchange momentum.

To see this momentum exchange clearly, we will switch to the momentum basis. Fortunately, the hard work of determining how nPCFs Fourier transform was done in (5.32), which reduced a momentum-space nPCF to products of Greens functions evaluated on Wick contractions of ϕs. The only trickiness introduced by interactions is the fact that the $\phi(x)$s that come from the perturbation S_1 are integrated over, as in the second line of (5.43). This position integral becomes

$$
\begin{aligned}
\int d^4x \, \langle \hat{\phi}(x)^4 \rangle &= \int d^4x \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_4}{(2\pi)^4} e^{ik_1 \cdot x} \ldots e^{ik_4 \cdot x} \, \langle \hat{\phi}(k_1) \ldots \phi(k_4) \rangle \\
&= \int d^4x \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_4}{(2\pi)^4} e^{i(k_1 + \cdots + k_4) \cdot x} \, \langle \hat{\phi}(k_1) \ldots \phi(k_4) \rangle \\
&= \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_4}{(2\pi)^4} \delta(k_1 + \cdots + k_4) \, \langle \hat{\phi}(k_1) \ldots \phi(k_4) \rangle .
\end{aligned}
$$

$$(5.44)$$

In other words, when some of the x values of an nPCF are the same in position space, then the corresponding k values in the momentum space have to sum to zero. This mathematical fact looks similar to momentum conservation, and in fact we'll see in the next section that (5.44) does enforce momentum conservation in scattering events.

In principle, we have now computed an interacting theory nPCF. (5.42) describes how to use perturbation theory to write an interacting nPCF in terms of non-interacting mPCFs We computed these mPCFs in (5.32) using Wick's theorem, and (5.44) describes how to integrate over the ϕ values present in the perturbing action S_1. Actually using these formulas however is a complicated and error-prone task. Fortunately, there is a graphical technique to isolate the Wick pairs and reduce these three equations to one simple form. This technique is called a Feynman dia-

gram, introduced by Richard Feynman in the 1960s, and is widely used across high energy physics.

Feynman Diagrams

To compute an nPCF in momentum space, (5.32) dictates that we must contract the $\hat{\phi}$ operators into pairs. The goal of a Feynman diagram is to do this visually. Let's represent every point in spacetime as a vertex in a graph, and Wick contractions are represented as edges (called propagators) between the vertices. In an interacting theory, some of the vertices come from the original nPCF to be computed. These vertices were denoted as k_i and are called *external vertices*. Others, called *internal vertices*, come from the perturbation S_1 and will be integrated over. Because each external vertex only has one ϕ operator attached to it, they will only connect to one edge. Each internal vertex will connect to four edges because every S_1 term contains four ϕs each, due to the ϕ^4 of the Lagrangian.

After the propagators are drawn in according to these rules, we can assign momentum values to the ϕ terms. Each Wick contraction consists of two operators $\phi(k_1)\phi(k_2)$ which may have different momenta. However, (5.32) showed that if $k_1 \neq k_2$, then the entire diagram will evaluate to zero. Thus, we can assume that $k_1 = k_2 = k$. As shorthand, we assign a momentum of k to the propagator that represents this Wick contraction.

For a 4PCF in the scalar interacting theory there are four external vertices. There is one internal vertex to first order in λ. Following the rules for drawing edges, there are ten ways of pairing the ϕs to evaluate the numerator of (5.42) which are shown in Figure 5.3. Diagrams A1-A6 are similar, as are B1-B3, so we will only calculate one diagram from each of these classes. For this reason, only the momenta of this diagram are labeled. The direction in which momentum is chosen to flow is a matter of preference.

To calculate the 4PCF $\langle \hat{\phi}(p_1)\hat{\phi}(p_2)\hat{\phi}(q_1)\hat{\phi}(q_2) \rangle$ from a diagram, we

Figure 5.3: Feynman diagrams for a 4PCF $\langle \hat{\phi}(p_1)\hat{\phi}(p_2)\hat{\phi}(q_1)\hat{\phi}(q_2) \rangle$ in the interacting scalar theory. Each line is a Wick pair and each vertex is a point in spacetime.

follow (5.32) by writing $G(k)$ for every propagator with momentum k. For internal vertices, (5.44) tells us to enforce momentum conservation by integrating over all momenta subject to the constraint that the total momentum of all the propagators connected to an internal vertex is zero. Thus diagram A1 diagrams evaluates to

$$\text{(A1)} \propto G(p_1)G(q_1)G(p_2)\delta(p_2 - q_2) \int \frac{d^4k}{(2\pi)^4} \delta(p_1 - q_1 + k - k)G(k)$$

$$\propto G(p_1)G(q_1)\delta(p_1 - q_1)\delta(p_2 - q_2) \int \frac{d^4k}{(2\pi)^4} G(k).$$

$$(5.45)$$

The proportionality statements acknowledge that we are focusing on the structure of the Green's functions and have not yet put in the correct powers of λ or other coefficients that might appear.

The first delta function in line 1 is the delta function of (5.32), enforcing that momentum does not change along a propagator. The second delta function is the delta function of (5.44), enforcing that momentum cannot change at a vertex. For this diagram, these constraints together enforce that momentum of each must be conserved, and the double δ-function in the above equation falls out to prevent the momenta from being mixed. That is, the free theory did not lead to momentum mixing in this case. It only changed the value of the 4PCF by some value $\int \frac{d^4k}{(2\pi)^4} G(k)$ which we have not computed yet.

We can compute the diagrams (B1) and (C) in a similar way:

$$\text{(B1)} \propto G(p_1)\delta(p_1 - q_1)G(p_2)\delta(p_2 - q_2) \int \frac{d^4k_1}{(2\pi)^4} \frac{d^4k_2}{(2\pi)^4} G(k_1)G(k_2)$$

$$\text{(C)} \propto G(p_1)G(p_2)G(q_1)G(q_2)\delta(p_1 + p_2 + q_1 + q_2).$$

$$(5.46)$$

Diagram (B1) is like A1 in that it conserves momentum for each pair of particles. However, Diagram (C) allows the much-heralded momen-

tum mixing. Even if p_1, p_2, q_1, and q_2 are all different, (C) will be nonzero as long as the momenta sum to zero. This diagram (and only this diagram) is the leading order contribution to $\phi\phi \to \phi\phi$ scattering, and we will use it a great deal throughout the rest of this chapter.

We have now verified that this ϕ^4 theory allows particles to interact. Before moving on to quantify this interaction by computing scattering cross sections, it's worth covering two more details. The first is to replace the above two proportionality statements with actual equations. The second is to work out the denominator of (5.42) and divide the sum of the (A), (B), and (C) diagrams by that to have a full equation for the interacting theory nPCF.

Symmetry Factor and Feynman Rules

All of the diagrams drawn in figure 5.3 were first order, since they contained only one power of S_1 and therefore only one power of λ. If we had written the diagrams of order λ^2, they would have had two internal vertices (one for each S_1). Thus, every diagram should be multiplied by the constant $i\lambda$ for every internal vertex.

Another potential worry is the possibility of under-counting. S_1 contains four ϕs, which can each be contracted with another ϕ in the nPCF. Contracting with the first ϕ and contracting with the second represent different contractions and should be added individually. But this is not reflected in our Feynman diagram, where we draw a contraction with any one of the ϕs as a propagator running into the internal vertex and make no note of which ϕ it was contracted with. Another source of error enters when we compute second order diagrams: these involve two identical S_1 values connected in different ways to the exterior vertices. We should double each diagram to reflect the fact that the two S_1 values could be swapped without changing the diagram structure. There are more numerical factors we haven't taken care of as well, which are all represented

Contribution	Source
$\dfrac{(-i\lambda)^n}{n!}$	Taylor series expansion of e^{-iS_1}
$\dfrac{1}{4!}$	Definition of the interacting term in the Lagrangian
$n!$	The number of ways to swap internal vertices
$\dfrac{1}{(\text{identical vertices})!}$	In the above row, we should not have swapped vertices which were already identical in the diagram
$4!$	The number of ways to connect 4 ϕs to 4 incoming propagators for each internal vertex
$\dfrac{1}{2^{\text{reversible propagators}}}$	In the above row, we should not have swapped propagators that both came from and led to the same vertex

Table 5.1: Factors we should multiply each Feynman diagram by due to expansion coefficients, over- or under-counting, and other sources. Here, n represents the number of internal vertices.

in table 5.3.

To get the total correction to each diagram, we multiply every entry in the left hand column of table 5.3. Most of the entries cancel, but to keep track of those that don't, we define the *symmetry factor*

$$S_f = (1 + \text{identical vertices})!(1 + \text{reversible propagators})! \qquad (5.47)$$
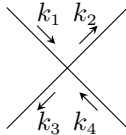
where identical vertices are internal vertices which are connected to the same vertices as each other, and reversible propagators are those that lead

to and out of the same internal vertex. in which case we should multiply each diagram by the factor $(-i\lambda)^n/S_f$.

As an example, in figure 5.3, all the A diagrams have symmetry factor 2 due to the reversible propagator with momentum k. The B diagrams have symmetry factor 4 because they have two reversible propagators, and the C diagram has symmetry factor 1 because there are no reversible propagators or identical vertices.

The definition of the symmetry factor allows us to write down the value of any diagram exactly by following the following rules, called the *Feynman rules*:

1. The edge $\xrightarrow{\ k\ }$ contributes $G(k)$

2. A vertex ⤫ contributes $-i\lambda\delta(k_1 - k_2 - k_3 + k_4)$

3. Every internal propagator k is integrated over $\int \frac{d^4k}{(2\pi)^4}$

4. Diagrams with exchangeable vertices or reversible edges contribute a symmetry factor S_f.

Once the diagrams are computed, they may be added and the sum divided by the denominator of (5.42), discussed in the next section, to get the nPCF

The Denominator: Vacuum Diagram Cancellation

This section deals with the denominator of (5.42):

$$\int \mathcal{D}\phi\, e^{-iS_0} e^{-iS_1}$$

which we must divide the sum of all Feynman diagrams by to get an nPCF. This factor came from requiring the nPCF to be normalized. Fortunately, this integral will cancel perfectly with terms in the numerator.

To see how, consider the diagrammatic expansion of this denominator. It is essentially a 0PCF, if such a thing exists. For scalar field theory, the zeroth order term is 1 and the the first order diagram is

where we have left out the circle to indicate the internal vertex (this is often done since any intersection of propagators is clearly a vertex).

This kind of diagram which is unconnected to external vertices called a *vacuum diagram*. You may recognize it diagram as a part of all the (B) diagrams in figure 5.3. In fact, the vacuum diagrams in the denominator will cancel with all of the diagrams in the numerator like (B) that have vacuum sub-diagrams. The mathematical explanation for this fact is that for every diagram like (A) or (C) that has no vacuum sub-diagrams, you can just multiply by a vacuum sub-diagram to get a diagram like (B), where both are combined. The reverse of this statement is that to divide by the vacuum diagrams, we need to remove all the diagrams like (B) that have vacuum sub-diagrams.

There is also physical intuition for removing by diagrams like (B) that have vacuum sub-diagrams. The vacuum diagrams have no external vertices, and external vertices generally represent particles. Thus, vacuum diagrams represent some change in energy of the vacuum, hence the name. Changes to the vacuum energy are generally undetectable, since we can only measure energy differences between systems and the vacuum is always present. For this reason, we should not count diagrams like (B) that measure the energy of the vacuum.[3]

---

[3]This is fortunate for us, since the vacuum diagrams can be infinite. The internal

Now that we know not to draw diagrams with vacuum sub-diagrams, we can fully compute any nPCF in the interacting theory with the Feynman rules. Our first and primary application for this newfound power will be to compute scattering amplitudes, which the next section is devoted to.

5.4   Spin Zero Scattering

In this section, we'll compute scattering amplitudes for $\phi\phi \rightarrow \phi\phi$ and $\phi \rightarrow \phi\phi\phi$ to get used to ϕ^4 theory and to practice with the Feynman rules. We'll follow up with a theory of two scalar particles and compute the probability of decay between the two.

Chapter 2 informed us that the scattering probability is the square of the S matrix, where the S matrix entry for m particles with momenta p_i scattering to n particles with momenta $-q_i$ is given by the $(m + n)$PCF for momenta p_i and q_i. We noted however that the S matrix is mostly zero; only when the incoming total momentum equals the outgoing momentum can scattering occur. This appeared in our calculation in the previous section of the 4PCF as well: every diagram in (5.45) and (5.46) was multiplied by $\delta(p_1 + p_2 + q_1 + q_2)$.

We stripped off this delta function part by defining the scattering amplitude $-iM$ as the part of the S matrix with $p_1 + p_2 + q_1 + q_2 = 0$, but excluding the no-scattering, diagonal entries where $p_1 = p_2$ and $q_1 = q_2$. The (A) diagrams for the 4PCF shown in figure 5.3 had delta functions $\delta(p_1 - p_2)$ and $\delta(q_1 - q_2)$, revealing these to be no-scattering diagrams. The only diagram that represents actual $\phi\phi \rightarrow \phi\phi$ scattering to first order is diagram C.

To extract a scattering amplitude from C, we must divide the corre-

---

vertices come with integrals over momenta due to the Feynman rules, but the lack of external vertices gives fewer delta functions. For many internal vertices, the diagrams can blow up.

sponding nPCF by the propagator for each incoming and outgoing field, as **JTD: Equation number** dictates. The motivation for this was that we are not interested in the way the particles move before or after they interact, which is represented by the propagators connected to the external legs. We are interested in the actual interaction, which occurs in the middle of diagram (C). Once this division is complete, we get our first ever scattering amplitude:

$$-iM_{\phi\phi\to\phi\phi} = -i\lambda + \mathcal{O}(\lambda^2).\tag{5.48}$$

This equation is very simple, perhaps disappointingly so. Let's calculate the second-order contributions to the scattering amplitude to see if they are more interesting.

The first step is to draw new second-order diagrams. Many of these diagrams will look like this:
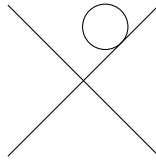
but these diagrams actually do not contribute to the scattering amplitude. Recall that the point of dividing by $G(p)$ for every external propagator p was to account for the motion of a particle before it interacts. The loop in the above diagram represents particle activity after it has interacted with the rest of the diagram. This loop represents self-interaction and is interesting in its own right. In fact, we'll study it in section 5.8. However, it still represents a particle propagating on its own and we must divide it out. Therefore, this diagram is just the same $-i\lambda$ first-order diagram we computed above.

This is an example of a generalized rule for calculating scattering amplitudes: the only diagrams that contribute to amplitudes are *truncated*. The definition of truncated is as follows:

In a truncated diagram, no internal propagators can be cut that sever exactly one external vertex from the diagram.

An internal propagator is one that does not connect directly to an external vertex. The above diagram was not truncated because the propagator between the central vertex and the loop could have been cut, and this would have severed the upper right external vertex from the diagram. However, the following diagrams *are* truncated, and do contribute to the matrix amplitude:

$$(5.49)$$
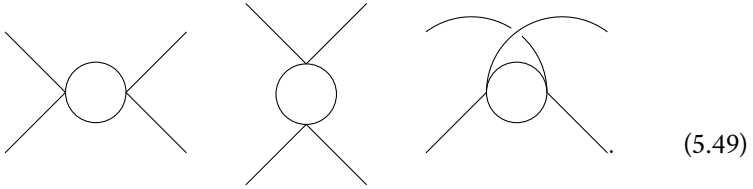
JTD: Put in momentum arrows

These represent the first non-trivial effect of QFT we have seen. Everything we have done so far has confirmed classical expectations, from $M = \lambda$ above to momentum conservation. These diagrams are fundamentally quantum and relativistic phenomena and represent a a true prediction of QFT. Let's compute them using the Feynman rules. Two propagators in each diagram are identical and can be exchanged, so they have symmetry factors of 2. Thus,

$$\text{Left} = -\frac{\lambda^2}{2} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2} \frac{1}{(k - p_1 - p_2)^2 - m^2}$$
$$\text{Center} = -\frac{\lambda^2}{2} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2} \frac{1}{(k - p_1 - q_1)^2 - m^2} \qquad (5.50)$$
$$\text{Right} = -\frac{\lambda^2}{2} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2} \frac{1}{(k - p_1 - q_2)^2 - m^2}.$$

where we have removed the external propagators in the spirit of truncation. It's common to deal with these diagrams by defining *Mandelstam variables*, which are a reparametrization for any two-to-two-particle interaction:

$$s = (p_1 + p_2)^2 \qquad t = (p_1 - q_1)^2 \qquad u = (p_1 - q_2)^2. \qquad (5.51)$$

There are three variables because momentum conservation makes a fourth one obsolete. They have the interpretation that s is the incoming momentum, equivalently the energy in the center of mass frame, and t and u represent the amount of momentum transfer. Calculation will reveal that each diagram is sensitive only to s, t, or u respectively, so all three diagrams can be added up as follows:

$$-iM = -i\lambda - \lambda^2 \left(V(s) + V(t) + V(u) \right) + \mathcal{O}(\lambda^3) \qquad (5.52)$$

where

$$V(s) = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2} \frac{1}{(k - p_1 - p_2)^2 - m^2}. \qquad (5.53)$$

This integral may strike you as a bit peculiar. Let's pretend for a moment that k^2 is the Euclidean norm $k_0^2 + \boldsymbol{k}^2$ rather than the Minkowski norm $k_0^2 - \boldsymbol{k}^2$. Then we could switch to spherical coordinates and d^4k would become $k^3 dk$ times a radial volume element $d\Omega$. For large k, this k^3 cancels with the integrand, which goes like k^{-4}, leading to an overall trend of $\int^\infty dk/k$. This integral is divergent.

The techniques for handling this divergence took a great deal of time to be developed historically. They necessitate a change in thinking which we will discuss in chapter 8, so we'll postpone this particular integral until then. In three spacetime dimensions (two spatial dimensions), the integral trends to $\int^\infty dk/k^2$ which is a perfectly reasonable number. Figure 5.4 shows the plot of $V(s)$ in three dimensions.[4]

---

[4]You might reasonably object to us computing predictions for an unphysical uni-
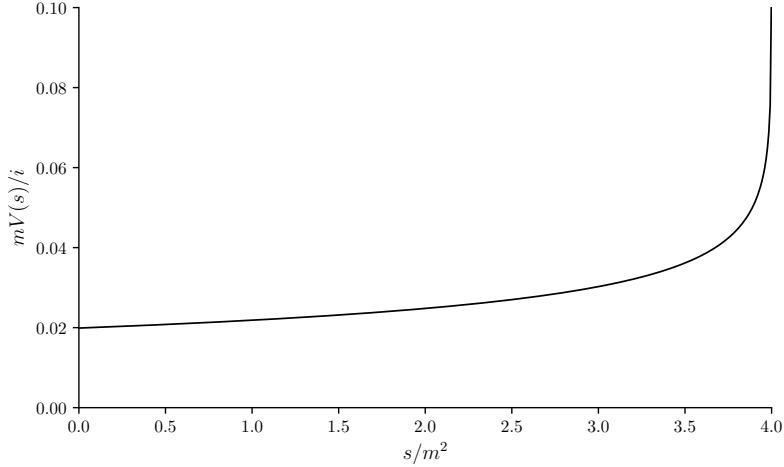
Figure 5.4: $mV(s)/i$, the unitless second order component of the scattering amplitude. This is computed for two scalar particles in 3 spacetime dimensions of center-of-mass energy s.

Something immediately apparent about Figure 5.4 is that the values are small. Multiplied by a small number λ^2 as 5.52 requires, and this $V(s)$ correction will be tiny except near the maximum value of $s = 4m^2$, where the function blows up. This is the hyper-relativistic limit, where the particles are colliding at enormous speeds. In this limit, the quantum corrections introduced by the second-order diagrams can become arbitrarily large. (5.53) is actually analytical computable:

$$V(s) = \frac{i}{16\pi\sqrt{s}} \ln\left(\frac{m + \sqrt{s/4}}{m - \sqrt{s/4}}\right). \tag{5.54}$$

verse, like one with only three spacetime dimensions, in a book about the phenomena of this universe. However, many of the phenomena we'll see are qualitatively similar between dimensions, and the characteristics of dimension four specifically will be covered later.

We will discuss how to arrive at this value for the integral in chapter 8.

Taking the limit as $s \to 4$ reveals that $V(s)$ blows up and scattering becomes extremely intense. This is a QFT correction to the classical expectation that $M = \lambda$; i.e., scattering amplitude is independent of energy.

Intuition for Feynman Diagrams

In the derivation of this chapter, each propagator in a Feynman diagram represents a Wick contraction, but there is an intuitive picture for Feynman diagrams beyond the mathematical structure of Wick's theorem. We already know an external propagator — a particle that links to an external vertex — represents a particle entering the scattering experiment. Why not view *every* propagator as a particle? Then the ϕ^4 term of the Lagrangian, which corresponds to four propagators meeting at a vertex x in a diagram, really does represent an interaction at x. A diagram with a loop, such as those in (5.49), represent two particles colliding, only for the outgoing particles to collide with each other again. If one collision is rare, then this secondary collision should be rarer than one collision only, which explains why these diagrams are small. On the other hand, if a collision is common, we might expect every particle pair to collide not just once or twice but many times, in which case we would have to draw infinitely many diagrams. It is at this point that QFT is no longer perturbative, and alternative methods to Feynman diagrams must be taken to compute scattering amplitudes.

There is a small problem with this picture, however. The two internal propagators in (5.49) are not real particles in a key sense: they are not on-shell. This can be seen in (5.50): we integrate k over all four-dimensional space without requiring $k^2 = m^2$. Instead, we call these internal propagators *virtual particles*, which interact like normal particles except that they can be off-shell (meaning that k^2 needn't equal m^2) and are never observed.

These virtual particles could have been predicted even before we saw the QFT principle of least action. One of the major results of special relativity was that mass is a form of energy, and a result of quantum mechanics is that a quantum particle may tunnel through a high-energy barrier in order to reach a low energy final state. Any theory of quantum mechanics and relativity, such as QFT, should combine these two well-understood phenomena. What better way to combine them than to allow particles to "tunnel" into an unphysical, perhaps higher-energy state before reaching a final state? The loop propagators represent this higher-energy state, and are unphysical in the same way that the electron should not classically exist in a high-energy barrier in quantum mechanics. Nevertheless, a particle can pass through the high-energy state in order to attain a different final state.

We could even have predicted that virtual particles can be off-shell by another example of hybrid quantum mechanical and relativistic reasoning: the energy-time uncertainty principle predicts that quantum states with a short lifetime have uncertain energies. In a relativistic context, the equivalent statement is that particles with a short lifetime have uncertain mass. Virtual particles are never observed, so they have effectively zero lifetime and infinitely uncertain mass, preventing the enforcement of the on-shell constraint.

We'll often treat any diagram with a loop, such as those of (5.49), as a "quantum correction" — a small correction quantum mechanics imposes on a classical theory. The loop brings all the quantum effects mentioned in the previous paragraphs into the picture; it is comprised of virtual particles and the $\int d^4k$ represents the fact that these virtual particles are off-shell. Figure 5.4 for the second-order $\phi\phi \to \phi\phi$ amplitude was our first example of a quantum corrections, but more will appear such as mass, energy, magnetic and electric dipole moment corrections.

The diagrams without loops, often called *tree-level diagrams*, gener-

ally represent the properties of a classical theory that could be attained by merely minimizing the action via the Euler-Lagrange equations. For example, a tree-level diagram delivered us $M_{\phi\phi\to\phi\phi} = \lambda$ as classical thinking would suggest.

5.5   Complex Scalar Theory

So far, we've considered only real scalar fields where $\phi^* = \phi$, but one may well ask if complex scalar fields exist. After all, the fermions in Nature can all take on complex values, and the only fundamental scalar particle in Nature (the Higgs) is complex. Fortunately for us, generalizing the above results to a complex field will be mathematically easy. However, it will bring up a critical new physics concept which we will need to understand: the antiparticle.

Antiparticles

The Lagrangian of a complex scalar field is nearly the same as that of a real field. The only differences are coefficients to the term and the presence of complex conjugates:

$$\mathcal{L} = -|\partial_\mu \phi|^2 - m^2|\phi|^2 + \lambda|\phi|^4. \tag{5.55}$$

The reason why the coefficients have changed is so that the counting arguments of table 5.3 which allow computation of the symmetry factor still apply even though now $\phi^* \neq \phi$.

Immediately, we suspect that ϕ^* may be the antiparticle of ϕ. This is because, as mentioned in chapter 1, an antiparticle is the mirror image of a particle under charge reversal. Charge reversal is manifested in complex conjugation for reasons that we will eventually see in chapter 7, so that charge reversal takes ϕ to ϕ^*. Thus, ϕ^* is the antiparticle of ϕ.

The antiparticle explanation may not be fully satisfying, however, since (5.55) still looks like the Lagrangian of only one particle. Let's there-

fore break up the ϕ field into components. Just as a complex number $z = (a + ib)\sqrt{2}$ can be broken into two real components (the $\sqrt{2}$ is for normalization), the complex field $\phi = (\phi_a + i\phi_b)/\sqrt{2}$ can be broken into real fields ϕ_a and ϕ_b. Plugging these into the Lagrangian, we get

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi_a)^2 - \frac{m^2}{2}\phi_a^2 + \frac{\lambda}{4}\phi_a^4$$
$$-\frac{1}{2}(\partial_\mu \phi_b)^2 - \frac{m^2}{2}\phi_b^2 + \frac{\lambda}{4}\phi_b^4 \qquad (5.56)$$
$$+\frac{\lambda}{2}\phi_a^2\phi_b^2.$$

We can now see that the complex scalar Lagrangian is just two real fields with the same mass m, same self-interaction λ, and an $a - b$ interaction term which is the last line. These properties were required for (5.55) to be written so simply.

The complex scalar Lagrangian (5.55) looks elegant, but when broken into its components it seems oddly contrived. Why should the real and imaginary parts have the same mass and interaction strengths? Why should they interact with each other in the exact way shown by (5.56)? These constraints are imposed by CPT symmetry. A particle's antiparticle must have the same mass and interactions. Mathematically speaking, they obey a *symmetry*, in that swapping ϕ_a and ϕ_b does not change the Lagrangian.

In fact, the symmetry is even deeper than this; any (normalized) linear combination of the two fields is also indistinguishable from ϕ_a and ϕ_b. For example, if we define $\chi_a = (3\phi_a + 4i\phi_b)/5$ and $\chi_b = (4\phi_a - 3i\phi_b)/5$, we could rewrite (5.56) in terms of χ_a and χ_b and we would get the exact same form, just with ϕ swapped out for χ. If it weren't for the odd $\frac{\lambda}{2}\phi_a^2\phi_b^2$ interaction term, this would not be true. **JTD: Unitary group too**

This particular symmetry is called SO(2), or *the Special Orthogonal Group of Order 2*. The name is motivated by the kinds of linear combinations which do not change the Lagrangian; "special" indicates that the

new linear combinations χ must remain normalized, "orthogonal" indicates that the two fields χ_a and χ_b must have the same product as the original $\phi_a\phi_b$, "group" indicates that any linear combination that satisfies these two properties is allowed as a symmetry of the Lagrangian, and "order 2" means that the linear combination acts on two fields ϕ_a and ϕ_b. This is our first example of a *continuous symmetry* because the set of available linear combinations is continuous. Nature adores continuous symmetries. The Standard Model which is the best known description of the fundamental particles has three of these symmetries, as will be discussed in the next part of this book.

You might wonder where ϕ's antiparticle was for the real scalar field case. For real scalars, the particle is its own antiparticle. This also occurs for some Standard Model particles, such as the photon.
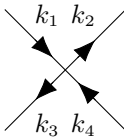
Feynman Rules for Complex Scalars

Now that we have shown that a complex scalar contains an antiparticle, it is time to discuss the Feynman rules of the new theory. Fortunately, we do not have to repeat the many pages of derivations necessary for the real field theory. The Lagrangian is essentially the same, so the Green's function of a complex scalar field is identical to that of a real field. Wick's theorem and its application to the free theory and the interacting theory did not rely on anything other than that the perturbation λ is small. Thus, we can skip straight to the rules for how to draw Feynman diagrams.

The new Lagrangian restricts diagrams slightly. The ϕ^4 term became $\phi^*\phi\phi^*\phi$, meaning that only two antiparticles and two particles can meet at a vertex, not three and one, for instance. We'll have to enforce this in Feynman diagrams with a graphical rule. Suppose we draw an arrow on a propagator, where the arrow points one way for a particle and another way for an antiparticle. Then the $\phi^*\phi\phi^*\phi$ term means every vertex must have two arrows leading in and two leading out. (Since ϕ and $\overline{\phi}$ have

opposite charge, you can think of these arrows as charge arrows.) Thus, the Feynman rules are

1. The edge $\xrightarrow{\quad k \quad}$ contributes $G(k)$

2. A vertex (diagram with k_1 k_2 k_3 k_4) contributes $-i\lambda\delta(k_1 - k_2 - k_3 + k_4)$

3. Every internal propagator k is integrated over $\int \frac{d^4k}{(2\pi)^4}$

4. Diagrams with exchangeable vertices or reversible edges contribute a symmetry factor S_f.[5]

We stopped drawing the momentum arrows in the diagrams to reverse clutter. There's no harm in just defining momentum to point in the same direction as the charge arrows.

As an example, here are the diagrams of (5.49) for a complex scalar field, representing $\phi\bar{\phi} \to \phi\bar{\phi}$ scattering:

$$(5.57)$$
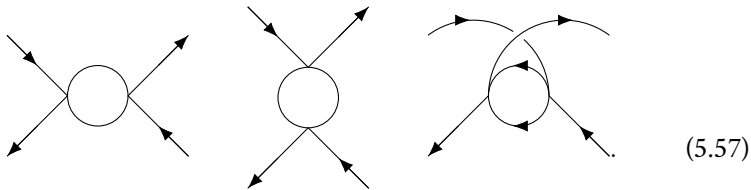
There are similar diagrams for $\phi\phi \to \phi\phi$ scattering:

---

[5]You might recall that the definition of S_f derived from a counting problem for how much each diagram should be weighted by, and that problem depended on ϕ being real. The coefficients of the new Lagrangian (5.55) have been carefully chosen so that the definition of the symmetry factor is the same, so we don't have to worry about redoing that argument.

$$\tag{5.58}$$
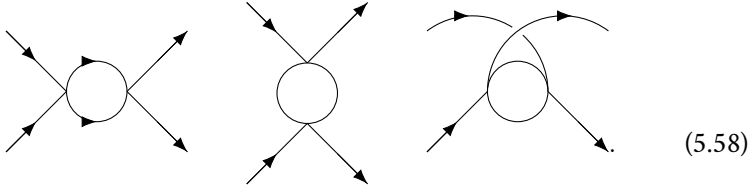
When a loop is drawn without charge arrows, the charge arrows can go either way so long as they are antiparallel. This does *not* count as a reversible edge in the symmetry factor sense; a symmetry factor represents that we have overcounted a diagram, whereas if a loop can run either of two directions we are undercounting the diagrams. Thus, we should double each diagram where the loop can run either day, not half it.

Thus, a difference between complex scalar scattering and real scalar scattering is that the quantum corrections to complex scattering is more intense. Contributions from both particles and antiparticles can occur, increasing the total interaction strength. The exception is in some channels which forbid particle-antiparticle interactions. For example, for $\phi\phi \to \phi\phi$, the s channel is not amplified because only ϕ particles may mediate that interaction. In $\phi\bar\phi \to \phi\bar\phi$, the u channel is not amplified.

One of the most famous properties of matter and antimatter is that they explode on contact with each other. Electrons and positrons annihilate to two or three photons, as do neutral pions. Charged pions decay to muons and neutrinos, et cetera. In the next section, we'll introduce a real scalar field as a photon-like particle and compute amplitudes for a complex scalar field to decay in this theory.

5.6   Scalar Decay

Let's create a theory with two particles: a complex scalar ϕ and a real scalar χ and an interaction between them of $\phi\phi^*\chi$. This corresponds to

a Lagrangian of

$$\mathcal{L} = -|\partial_\mu \phi|^2 - m_\phi|\phi|^2 - \frac{1}{2}(\partial_\mu \chi)^2 - \frac{1}{2}m_\chi \chi^2 - g\phi\phi^*\chi. \qquad (5.59)$$

In order to draw Feynman diagrams for a system with two particles, we need two types of lines — a dashed line for χ and a solid line for ϕ. The Feynman rules are as you might guess; the ϕ and χ propagators are still $G(k)$, momentum conservation is still applied at every vertex, and internal propagators are still integrated over. The only difference comes from the new interaction terms, which allows vertices connecting two ϕ particles to one χ particle
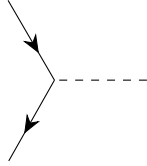
which contributes an amplitude of $-ig$. Note that there is no charge arrow on the dashed χ line because χ is real and therefore uncharged.

Many interesting phenomena can be studied in this theory, but in this section we will only focus on one: the decay of χ into two ϕ particles. Due to energy conservation, this decay can only occur if $m_\chi > 2m_\phi$, so we assume this inequality.

The leading order interaction for decay is just the above Feynman diagram with some added labels
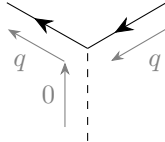
We've also rotated the diagram because, when drawing diagrams for scattering experiments, it is traditional to put the initial state at the bottom

and the final state at the top so that time flows upwards as in embedding diagrams in relativity. We assumed that the χ particle is at rest, which we can always undo by boosting into a different reference frame if needed. The only truncated next order diagram is
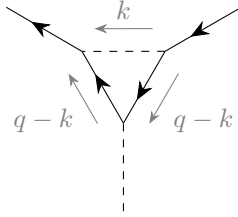
Together, these diagrams produce a decay amplitude of

$$iM = -ig + (-ig)^3 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m_\chi^2} \frac{1}{[(q-k)^2 - m_\phi^2]^2} \qquad (5.60)$$

which yields a decay rate given by (2.13). For the rest of this discussion, we'll neglect the second order term and just use $M = -g$.

Energy conservation gives $2q^0 = m_\chi$ and the on-shell constraint dictates that $(q^0)^2 - \boldsymbol{q}^2 = m_\phi^2$, so the momentum of the outgoing particles is $\boldsymbol{q}^2 = m_\phi^2 - m_\chi^2/4$. Thus, the three-dimensional integral over outgoing momenta is actually a two-dimensional integral on a sphere, and the spherical symmetry of this decay problem implies that the value is 4π times the integrand. Thus,

$$\Gamma = \frac{|M_{p \to q}|^2}{(2\pi)^5 m_\chi^3} \left(m_\phi^2 - \frac{m_\chi^2}{4} \right) \qquad (5.61)$$

It's more interesting to compute the unitless constant Γm_χ^2 — physically, the probability to decay within one Compton wavelength of the particle, as a function of the ratio of $\mu = 2m_\phi/m_\chi < 1$. This is

$$m_\chi \Gamma = \frac{1 - \mu^2}{4(2\pi)^5} g^2 \qquad (5.62)$$

One can see that as the decay products get lighter and μ gets smaller, the lifetime of χ becomes shorter. This is intuitively what we expect of our first every decay width calculation. Decay is the relativistic version of quantum tunneling into a new state, and quantum tunneling is generally more likely to happen if the potential energy of the end state is lower. Here, the mass energy of the end state is proportional to μ, so we expect the lifetime to decrease with μ.

Another intuitive fact is that, for $\mu = 0$, Γ reaches a finite value. This is something we should have expected because many particles in Nature, such as the pion, decay into massless photons. Yet the neutral pion has a relatively long lifetime of seconds, corresponding to a travel distance of 25 nanometers when moving at the speed of light (many particles live even shorter lives). Ignoring the fact that photons are not complex scalar particles, it would be concerning if our current calculations predicted zero lifetime for such decays.

If $m_\chi < m_\phi$, this decay is energetically forbidden, and completely new physics arises from the same Lagrangian which is discussed in the next section.

5.7   Coulomb Potential for Scalars

One of the foundational assumptions of QFT was that the theory is local, meaning that every particle is influenced only by its immediate environment. This might seem obviously incorrect at first because of long-range forces like Coulomb's law, which allow distant objects to affect one another. But in electromagnetism we explained these forces via a field which propagates information between objects by the means of waves, and in QFT we graduate those waves to particles. In this section we use the same Lagrangian as in the previous to demonstrate a long-range force carrier.

As an experiment to look for forces, we could scatter a complex scalar

ϕ particle off of an equally charged ϕ particle and measure the cross section. The light, real scalar χ will carry the force between the two particles. The leading order diagram for this interaction is
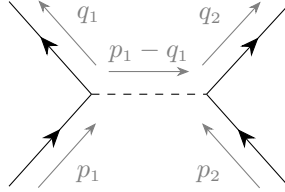
If the outgoing particles were distinguishable then there would be a second diagram wherein the particles are swapped, but in this case the outgoing particles are equally charged and are indistinguishable. Setting $m_\chi = 0$, the scattering amplitude is

$$iM = \frac{-ig^2}{(p_1 - q_1)^2} \tag{5.63}$$

which is just a constant times the Green's function for the one internal propagator.

We can use this scattering amplitude to compute the equivalent of the potential between the particles if we make a few approximations. Firstly, we'll work in the non-relativistic limit, where the energies of all particles is approximately their masses, in which case $p_1 - q_1$ has zero temporal component. The second approximation is to compare this amplitude with that predicted by non-relativistic quantum mechanics under the Born approximation, as discussed in appendix 18. That amplitude is $M(\boldsymbol{p_1} - \boldsymbol{q_1}) = -V(\boldsymbol{p_1} - \boldsymbol{q_1})$ where V is the Fourier transform of the potential between particles. Plugging in our amplitude calculation, we have

$$V(\boldsymbol{x}) = -g^2 \int \frac{d^3\boldsymbol{\ell}}{(2\pi)^3} \frac{e^{i\boldsymbol{\ell}\cdot\boldsymbol{x}}}{\boldsymbol{\ell}^2} \tag{5.64}$$

where $\boldsymbol{\ell} = \boldsymbol{p}_1 - \boldsymbol{p}_2$. By inspection, the integral cannot depend on the orientation of $\boldsymbol{x}$, so we write $\boldsymbol{x} = r\hat{\boldsymbol{z}}$ and switch to spherical coordinates:

$$
\begin{aligned}
V(r) &= -\frac{g^2}{4\pi^2} \int_{-1}^{1} d\cos\theta \int_{0}^{\infty} d\ell\, e^{i\ell r \cos\theta} \\
&= -\frac{g^2}{8\pi^2} \int_{-1}^{1} d\cos\theta \int_{-\infty}^{\infty} d\ell\, e^{i\ell r \cos\theta} \\
&= -\frac{g^2}{4\pi} \int_{-1}^{1} d\cos\theta\, \delta(r\cos\theta) \\
&= -\frac{g^2}{4\pi r}
\end{aligned}
\tag{5.65}
$$

The second line came about by recognizing that the real part of the integrand was symmetric under $\ell \to -\ell$, so we could extend the ℓ integral. The third line applied the theorem connecting the integral of an oscillating exponent to the delta function. The last line came from the fact that $\delta(f(x)) = \delta(x)/f'(x)$.

If we interpret the g as the "charge" of the scalar particle, then the result is the Coulomb potential! In case the QFT principle of least action were still in doubt, this is an excellent sign that the theory is pointing us in the right direction. The only strangeness is that the sign is wrong; these like-charged particles are attracted to each other whereas in electromagnetism they would be repelled. This is a result of our choice to use scalar particles. In the next two chapters we will introduce new particles which are capable of flipping this sign to become more electromagnetism-like.

Another motivation for us to pursue more particles to describe electromagnetism is that this theory contains no magnetic field. There aren't enough degrees of freedom in the particles provided to store the direction of an additional field, and adding new "magnetic field particles" would not be able to interact so as to guarantee Maxwell's equations. We will need a particle capable of encoding the direction of the magnetic field within its quantum state.

5.8   Massless Scalars are Rare

One final problem with the Coulomb-like potential derived in the previous section is that massless scalars are unlikely to occur. To see why, recall from chapter 3 that the 2PCF of a particle contains a pole at the particle's mass. We can compute the 2PCF of the χ particle using the Lagrangian above to see where its pole is. The first two orders of Feynman diagrams are
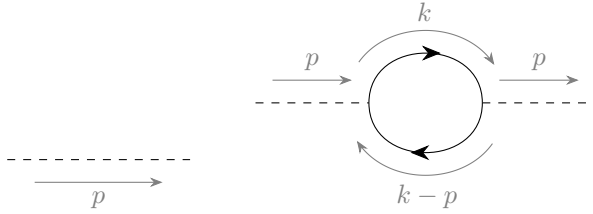
which correspond to a 2PCF of

$$\int d(\mu^2)\, \rho(\mu^2) \frac{1}{p^2 - \mu^2} = \frac{1}{p^2}\left[1 + \frac{\Sigma}{p^2}\right].\qquad(5.66)$$

where we used the definition of ρ from (3.2), and

$$\Sigma = (-ig)^2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m_\phi^2} \frac{1}{(p-k)^2 - m_\phi^2}.\qquad(5.67)$$

Note that the external legs are included because we are computing a 2PCF, not a scattering amplitude.

We could have continued adding loops, leading to diagrams such as
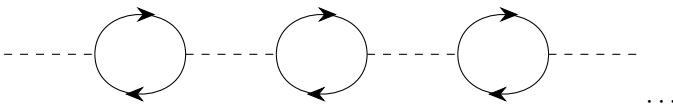
Adding together all the contributions from all numbers of loops gives

$$\int d(\mu^2)\,\rho(\mu^2)\frac{1}{p^2-\mu^2} = \frac{1}{p^2}\left[1 + \frac{\Sigma}{p^2} + \frac{\Sigma^2}{p^4} + \cdots\right]$$
$$= \frac{1}{p^2}\frac{1}{1-\Sigma/p^2} = \frac{1}{p^2-\Sigma}.$$

(5.68)

Surprisingly, the loops move the spike in the spectral density $\rho(\mu^2)$ from zero to $\mu^2 = \Sigma$, in violation of our expectation that the mass of the χ particle should be m_χ, which is zero[6].

This fact leaves an interesting question. If the spike in the spectral density ρ need not equal the particle mass in the Lagrangian, which should we interpret as mass? It is tempting to continue to quote the m_χ in the Lagrangian as the mass of the particle because it is an intuitive quantity, but this number is not actually observable while the spike in the spectral density is. As we mentioned earlier in this chapter, the spike also plays the key role of ensuring that any particle's momentum always obeys the relativistic $p^2 = m^2$ constraint. The square of the 2PCF gives the probability of a state being observed, so if the 2PCF blows up for p^2 equal to a certain amount it makes sense to define that amount as the mass.

This realization — one might say betrayal — that the Lagrangian mass is not the true mass of the particle has a physical explanation, which also explains why we only noticed the difference between the Lagrangian mass and the spectral mass (the true mass) when we looked at a loop diagram. When a χ particle is moving through space, this loop represents the creation of virtual ϕ-antiϕ pairs that immediately annihilate. This happens at all times at all points in space and is an inevitable prediction of a quantum theory that allows tunneling. We never observe these virtual particles, but their presence temporarily increase the rest energy of the χ particle and is responsible for the shift in the spike of $\rho(\mu^2)$. Thus, when

---

[6]We have implicitly assumed that $\Sigma \neq 0$. This assumption is correct, but we will not have the mathematical tools to evaluate Σ until chapter 8.

we observe the χ particle we also observe the shadow of the ϕ particles it is coupled to.

One might be suspicious of this rather hand-wavy argument, but some experiments confirm that this particle-antiparticle creation and annihilation is exactly what is happening. Specifically, when measurements are made in an extremely small volume or extremely close to another particle, these particle-antiparticle pairs have less room to form and are suppressed. The Lagrangian mass being exposed a little, and the observed properties of particles start to change. For example, when an electron is put very close to a proton, the charge of both appears to increase a little and this results in a small shift in the energy of the s orbitals of hydrogen known as the Lamb shift. Another example is that when you put two plates very close to each other, the particle-antiparticle pairs that occur naturally in a vacuum are suppressed between them, and the pressure of these pairs outside the plates forces them together in a phenomenon known as the Casimir effect. The mathematical description for these effects will be given in chapter 8.

This mass shift is why the title of this section is that massless scalars are rare. In order for χ to be observed as massless — that is, the spike of its spectrum is at zero — its m_χ value in the Lagrangian must be "fine tuned" such that the loop diagram and all higher order loop diagrams cancel exactly with m_χ and move the spike of the spectrum to zero. Historically, a theory that requires parameters to be finely tuned to reproduce observations has been ill-trusted because it suggests the existence of some unexplained structure to the universe. Fortunately, we do not observe any massless scalars in Nature, so there is no fine tuning problem. The only massless particles are vector bosons, which we will see in chapter 7 cannot be made massive by loop diagrams.

Problems

5.1   Wick's Theorem. In section 5.1, we explained Wick's theorem, which is that $\frac{d^n Z}{dJ_{j_1}\ldots dJ_{j_n}}$ is equal the sum of M_{ab}^{-1}, where a, b are pairs of $j_1, \ldots j_n$. This was shown for $n = 2$ in (5.16). Prove via induction that this fact holds for $n > 2$.

5.2   Momentum Conservation. In this question, we'll determine why momentum conservation came so naturally out of the QFT principle of least action.

1. Show that all Feynman diagrams, of any order and with any number of external vertices, conserve momentum in that if the external momenta do not sum to zero, then the nPCF is zero.

2. **JTD: Connect this to the real-ness of the Lagrangian via Fourier transform logic**

3. Show that the operator e^{iH} is unitary if and only if H is Hermitian. Since the integrand of the QFT principle of least action is $e^{i\mathcal{L}}$, where $\mathcal{L}$ is real, the conservation of momentum structure we got is just the requirement that $e^{i\mathcal{L}}$ be unitary. Unitarity is a property of non-relativistic quantum mechanics that was necessary for the probability to detect a particle not to change over time. The QFT principle of least action is written to preserve unitarity in the sense of momentum conservation, even though this theory can create and destroy particles.

5.3   Symmetry Factors. **JTD: Evaluate some diagrams and find some symmetry factors**

5.4   New Lagrangians. Suppose we designed a new scalar theory in which ϕ were real. This allows us to write a cubic term in the Lagrangian instead of a quartic term:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - m^2 \phi^2 + \frac{g}{3!}\phi^3.$$

Write out the Feynman rules for this theory. What would be the Feynman rules of the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - m^2\phi^2 + \frac{g'}{4}\phi^2(\partial_\mu \phi)^2?$$

Hint: this question requires very little math.

5.5   Coulomb potentials. Using the same Lagrangian as section 5.7, determine whether oppositely charged ϕ particles repel or attract.

5.6   Coulomb potentials. Section 5.7 showed that scalar particles reproduce Coulomb's force.

1. Convince yourself that the reason why Coulomb's force was reproduced was that the diagram we were using evaluated to a single particle propagator, and that the form of this propagator came from the kinetic term in the Lagrangian

2. The Coulomb potential is sometimes justified via Gauss's law, that $\nabla^2 V(x) \propto \rho(x)$. Like the Lagrangian, this is a local equation. Argue that it is unsurprising that the Lagrangian kinetic term and Gauss's law should have both led to the same conclusion.

Chapter 6

Spin One Half

This chapter deals with extending QFT to cover particles with spin, also known as *fermions*. Spin is covered in non-relativistic quantum mechanics, though its properties seem sometimes mysterious and arbitrary. In particular, the fact that fermions obey the Pauli exclusion principle and bosons do not cannot be explained in non-relativistic quantum mechanics. This explanation is provided by relativity in QFT in the form of the spin-statistics theorem, which we explain in this chapter. Before arriving at this key result, however, we discuss spin in a generalized sense (section 6.1) and show that when relativity is taken into account, spin implies the existence of antiparticles (section 6.2).

Let's start with a simple definition: spin is the intrinsic angular momentum of a particle. In order to store a particle's spin state in QFT we'll need to graduate our scalar field particle operator $\hat{\phi}$ to a vector of operators $\hat{\psi} = (\hat{\psi}_0, \hat{\psi}_1, \dots)^T$. This vector is called a *spinor*.

The next section uses the definition of spin as intrinsic angular momentum to dictate how many entries the spinor must have and what operators can act on it. Our discussion will start with the non-relativistic case, where we do not consider boosts, and then we will add boosts to see how they change the discussion. Following this, we'll describe how to

similarly graduate the ψ in the path integral of the QFT principle of least action to a spinor. With this complete, we will repeat the analysis of the previous chapter and derive Feynman rules for fermions. Finally, we will use these new fermions to understand electrons, the Coulomb potential, and atomic nuclei.

6.1   Spinors and Spin Operators

Spin is a quantum state, and in any quantum theory, a quantum state is acted on by operators. It will require the description of new mathematical tools which come from the field of group theory, but by the end we will have a complete description of any particle with spin.

To measure the spin of a spinor ψ, we act on ψ with a vector of operators $\boldsymbol{S}$. In order to be interpreted as an angular momentum vector, $\boldsymbol{S}$ must satisfy the defining feature of angular momentum:

Angular momentum generates rotations.

The notion of a *generator* is our first group theoretic term and we will define it in the following paragraph. But first, this definition of angular momentum is different from the classical definition of $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$. It is nevertheless a correct definition due to Noether's theorem, which states that an operator's generator is conserved if that operator is a symmetry of the Hamiltonian (appendix 20.4). Rotations are a symmetry of spacetime, meaning that if angular momentum generates rotations then angular momentum is conserved. The cross product definition of the angular momentum will return later in this section.

Generators

Now to define what a generator is. Consider a rotation operation R which acts on a three-component vector. We know that R can be written

as products of the three-dimensional rotation matrices:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \qquad R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(6.1)

This fact shows that there are many possible rotation matrices, but that they can each be traced down to the three matrices above. In fact, the above rotation matrices can be simplified even more by writing $R_j(\theta) = e^{i\theta S_j}$, where

$$S_x = \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad S_y = \frac{i}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

(6.2)

$$S_z = \frac{i}{2}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

You can check for yourself that exponentiating these matrices gives back the rotation matrices. An arbitrary rotation can therefore be written as

$$R = e^{i\boldsymbol{\theta}\cdot\boldsymbol{S}}$$

(6.3)

where we have arranged S_x, S_y, and S_z into a vector called $\boldsymbol{S}$. The three components of the vector $\boldsymbol{\theta}$ are called Euler angles and they parameterize the rotation matrix R.

In mathematical language, we say that $\boldsymbol{S}$ *generates* the rotations because all rotations can be achieved by plugging different values of $\boldsymbol{\theta}$ into (6.3). The fact that angular momentum is generated by rotation could

mean that we should just take S as the spin operator S, and write the spinor as a three dimensional vector which S acts on.

Angular Momentum Algebra

This procedure does not work in practice because there are other ways to generate rotations. One can rotate a complex vector for instance, in which case we need a new set of S matrices which are themselves complex. There is an even finer point which must be made: it is necessary that S generate a three-dimensional rotation since spacetime is invariant under three-dimensional rotations. However, the matrix R need not act on a three-dimensional vector. R could act on a five-dimensional vector, for example, and rotate only in a three-dimensional subspace. In this case S would still contain three generators, but each generator would be a 5×5 matrix rather than 3×3.

We therefore need to write a more general definition of rotations than just the rotation matrices (6.1). One property rotations must have is that they must preserve the length of vectors. That is, R must be unitary. Problem **JTD: cite** asks the reader to show that for e^{iM} to be unitary, M must be hermitian and determinant 1. Thus, the entries of S are hermitian.

The other required property of rotations comes from the way in which they anticommute. Consider rotating a vector aligned with the z-axis. If we rotate in the x direction by a small angle δ first and then y by the same $delta$, we get a different vector than rotating in the y direction first then y. This is because the rotation matrices anticommute. Using (6.1) or drawing pictures that the two rotations differ by a rotation about the z axis by angle δ^2. We can turn this fact into a constraint on S by using (6.3) to define the x, y, and z rotations. They show that $S_x S_y - S_y S_x = i S_z$.

In principle, we could have aligned our initial vector along any axis

z, so that the more general version of this constraint is

$$[S_i, S_j] = i\epsilon_{ijk}S_k \tag{6.4}$$

where ϵ_{ijk} is the Levi-Civita symbol. Any set of hermitian, determinant one matrices $\boldsymbol{S} = (S_x, S_y, S_z)$ which satisfies this equation is a valid angular momentum operator. We usually refer to (6.4) as an *angular momentum algebra*, and matrices that satisfy it are *representations* of the angular momentum algebra

Using Lie groups to find representations of the angular momentum algebras

The task of finding matrices $\boldsymbol{S}$ which satisfy (6.4) (or in mathematical language, finding representations of the angular momentum algebra) has fortunately been solved using group theory. A Lie group is a closed, continuous set of operators, where closed signifies that two operators in the group U and V multiply to produce another operator W which is also in the group. We have unknowingly been working with the group of rotations in three dimensions, which is called the "special orthogonal group of order three," or SO(3). "Special" means that the group operators do not change the length of vectors, "orthogonal" means that the group operators are real and unitary, and the three indicates that the rotation occurs in three dimensions.

A Lie group satisfies the general result that any group element G can be written as

$$G = e^{i\boldsymbol{\theta}\cdot\boldsymbol{S}} \tag{6.5}$$

where $\boldsymbol{S}$ are a special set of matrices known as the generators of the Lie group. We used this property in (6.3), but we did not know that it was valid for any Lie groups. The matrices $\boldsymbol{S}$ must satisfy special relations to each other which are dependent on the Lie group the generate; these

relations are called a Lie algebra. An example is the angular momentum algebra we discussed above, which we now know as the Lie algebra of SO(3). Mathematicians sometimes label this algebra as $\mathfrak{so}(3)$ but physicists usually use SO(3) to refer to the algebra in addition to the group. We will therefore use the $\mathfrak{so}(3)$ notation in this section only.

Once we specify the dimensions of the spinor which the group elements act on — say the spinor has dimension n, then we can write group elements as $n \times n$ matrices. This is called a representation. Our three-dimensional rotation matrices (6.1) were a representation of SO(3), and our three-dimensional angular momentum operators (6.2) were representations of its Lie algebra. Matrix exponentiating the generators as in (6.5) can be used to produce the representation of the group if a representation of the Lie Algebra is known.

Physical consequences of the group theory explanation for spin

This new formalism connects directly to the definition used at the beginning of this section — "Angular momentum generates rotations." The connection is that "rotations" refers to the Lie group SO(3), which has the Lie algebra given in (6.4). The angular momentum operators we're looking for are the rotation generators, which are representations of this Lie algebra.

One reason why this group theory framework is necessary is because (6.4) is not just the Lie algebra of SO(3). It is also the Lie algebra of SU(2), which is the set of unitary matrices which rotate a two-dimensional complex vector, rather than a three-dimensional real vector. Valid angular momentum operators could therefore be generators of SO(3) or SU(2).

Another use of group theory is the mathematical result that all the

faithful[1] representations of SO(3) have odd dimensions, and all the representations of SU(2) have odd[2] dimensions. Therefore, the spinor which the spin operators act on can have any number of dimensions n. If n is even, then one uses representations of $\mathfrak{su}(2)$ as spin operators. If n is odd, one uses representations of $\mathfrak{so}(3)$.

This fact that the parity of n changes what group the spin operators come from has several fascinating properties. Firstly, it means that the SU(2) spins may have different properties than the SO(3) spins. This is manifested in the spin-statistics theorem, which is described in the next section. It also means that since SU(2) acts on complex vectors and SO(3) acts on real vectors. This is crucial because, as noted in the previous chapter, complex particles have an electric charge. Thus, all SU(2) particles are automatically charged. SO(3) particles can be neutral or charged, because one can multiply them by a complex phase.

Another result of group theory is that $\boldsymbol{S}^2 = \frac{n^2-1}{4}\mathbb{1}$, where $\boldsymbol{S}^2 = S_x^2 + S_y^2 + S_z^2$. This is interesting because we define the spin of a particle s such that the eigenvalue of the angular momentum operator $\boldsymbol{S}^2$ is $s(s+1)$ so that $s = \frac{n-1}{2}$. Thus, particles with half-integer spin s have spinors of even dimension and are therefore represented by SU(2) and we refer to them as fermions. Integer spin particles, called bosons, have odd dimension spinors and are represented by SO(3).

One final result of group theory is a list of what the spin operators actually are. For zero spin ($s = 0$), the spinor has dimension 1, and the operators are all the 1×1 matrices (0). This is an incredibly boring case from the perspective of spin, and it is the reason why we were able to complete all of the previous chapter on spin-zero particles.

---

[1] A faithful representation is one where every group element corresponds to exactly one matrix. For odd dimensions, SU(2) is not faithful because every group element corresponds to two matrices which can be generated using $e^{i\boldsymbol{\theta}\cdot\boldsymbol{S}}$.

[2] Except, that is, for the trivial representation, where all the generators are the 1×1 matrix with zero as the only element.

Spin one-half particles have $n = 2$, so that the spin operators are 2×2 representations of $\mathfrak{su}(2)$ which are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (6.6)$$

Non-relativistic quantum mechanics courses usually skip the derivation of spin operators and merely state that they are equal to these Pauli matrices, partially because most of the particles dealt with in non-relativistic quantum mechanics are spin one-half so that the operators for other particles are unnecessary.

Spin one particles have $n = 3$, so that the spin operators are 3×3 representations of $\mathfrak{so}(3)$. We have already written these matrices in (6.2) because $n = 3$ is the representation used to rotate normal three-dimensional geometric vectors. This means that for spin-one particles (and only spin-one particles), spin is a true geometric vector which rotates normally and need not be understood in a quantum mechanical way. Photons are spin-one particles whose spin corresponds to their polarization, and classical electromagnetism only works to describe photons because of this fact — their spin behaves like a classical vector and can be treated as such. For this reason, we sometimes call spin-one particles "vector bosons." Vector bosons are able to play a special role in Lagrangians because their spin can be used as a Lorentz index to turn vector-valued quantities into scalar-valued quantities that can go into the Lagrangian. This allows them to act as "gauge bosons," as described in the next chapter.

Throughout the rest of this chapter, we will explore only the spin one-half case, saving spin-one for the next chapter.

The Lagrangian of a Massless Fermion

Since we are specializing to the spin one-half case, the spin operators are the Pauli matrices, which can be combined into a three-dimensional vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. To make a relativistic version of the spin operator, we need to graduate $\boldsymbol{\sigma}$ to a four-vector σ^μ by adding a time component, which corresponds to Lorentz boost operator for spin.

This section has so far considered only rotations, so we do not know how operators transform under spin. As a simplification, we'll consider a massless fermion, which can only move at the speed of light. Thus, boosting cannot change its state, and σ^0 should just be the identity matrix:

$$\sigma^\mu = (\mathbb{1}, \boldsymbol{\sigma}). \tag{6.7}$$

Now that we have relativistic spin operators, we can derive the Lagrangian of this massless Fermion by writing down all terms that satisfy the symmetries of QFT. Firstly, the Lagrangian must be a real scalar rather than spinor-valued, so that ψ cannot exist alone. It must be paired with another spinor as in $\psi^\dagger \psi$, which indicates the dot product between two spinors[3].

Another symmetry of the Lagrangian is Lorentz transformations, which implies that all Lorentz indices must be contracted. In the scalar field theory, the only object with Lorentz indices we had access to was the derivative ∂_μ, but now we also have the spin operators σ^μ. Thus, the available terms are of the form

$$\psi^\dagger \psi \qquad \psi^\dagger \sigma^\mu \partial_\mu \psi \qquad \psi^\dagger \partial^2 \psi.$$

We did not include $\psi^\dagger \sigma^2 \psi$ because $\sigma^2 = 4\mathbb{1}$, so this term is redundant with the first.

---

[3]Traditionally, we use matrix notation for spinors and index notation for spacetime indices. Thus, a spinor dot product is $\psi^\dagger \psi$ but a four-vector dot product is $k^\mu k_\mu$.

The Lagrangian must also be invariant under "Internal" Lorentz transformations, which transform spinors rather than spacetime. These transformations are of the form $\psi \to \Lambda\psi$ where $\Lambda = \exp\left[\frac{1}{2}\sigma^\mu\omega_\mu\right]$ where ω_μ are arbitrarily chosen constants. Since σ^μ are all hermitian, Λ is unitary. Thus, $\psi^\dagger\psi \to \psi^\dagger\Lambda^\dagger\Lambda\psi = \psi^\dagger\psi$ and all of the above terms are invariant under these internal transformations.

A final symmetry to consider is CPT symmetry, which actually forbids the first term — specifically, due to parity (P). Spinors are invariant under time (T) symmetry, and $\psi \to \psi^\dagger$ under charge (C) symmetry, so that $\psi^\dagger\psi$ is invariant under CT. But under parity, angular momentum changes direction because it is a cross product, so the spinor cannot be invariant under P.

The last two terms are allowed by all the symmetries of the Lagrangian, but we usually do not include the last term because its dimension is too high. To compute its dimension, consider the middle term which must have dimension 4 so that $S = \int \mathcal{L}d^4x$ is dimensionless (recall that d^4x has dimension -4). Since σ^μ is dimensionless and ∂_μ has dimension 1, this implies that the spinor ψ has dimension $3/2$, in contrast to a scalar field's dimension $[\phi] = 1$. Therefore, the last term $\psi^\dagger\partial^2\psi$ has dimension 5. To bring it to the correct dimension, we must multiply by a number g which has dimension -1, and therefore all the scattering cross-sections that use this term will contain a factor of $(gE)^n$, where $n > 0$ is an integer and E is the energy of the scattering. But scattering is generally done at low energies, so that gE is small. Therefore, this term has a small impact on the physics of fermions and we do not include it.

The final Lagrangian for a massless fermion is thus

$$\mathcal{L} = \psi^\dagger\sigma^\mu\partial_\mu\psi. \tag{6.8}$$

This gives the famous Dirac equation (1.4), when variational calculus is used to compute the corresponding differential equation.

6.2   **Massive Fermions and the γatrices**

The Lagrangian we just derived does not contain a mass term because we had assumed a massless fermion in the above discussion. We could try to add a mass term of the form $\psi^\dagger \psi$ — inspired by the $\phi^* \phi$ mass term of a scalar field — but this is incorrect because $\psi^\dagger \psi$ is not CPT invariant as we discussed above.

The solution is to introduce another particle. We call the original spinor ψ_L and we define its transformation under parity as $P\psi_L = \psi_R$. Likewise, $P\psi_R = \psi_L$ since $P^2 = \mathbb{1}$. We say that ψ_L has "left helicity" and ψ_R has "right helicity." A more practical definition of *helicity* is that it is the projection of spin against the velocity of the particle[4].

The $\psi^\dagger \psi$ (now $\psi_L^\dagger \psi_L$) can be made parity invariant by converting it to $\psi_R^\dagger \psi_L + \psi_L^\dagger \psi_R$. The new Lagrangian is

$$\mathcal{L} = \psi_L^\dagger \sigma^\mu \partial_\mu \psi_L + \psi_R^\dagger \overline{\sigma}^\mu \partial_\mu \psi_R + m(\psi_R^\dagger \psi_L + \psi_L^\dagger \psi_R) \tag{6.9}$$

where $\overline{\sigma}^\mu = (\mathbb{1}, -\boldsymbol{\sigma})$ is the parity reflection of σ^μ and m is the mass.

It is convenient to combine ψ_L and ψ_R into one spinor called the Dirac spinor and the spin matrices into four-by four matrices called the Dirac γ matrices:

$$\hat{\psi} = \begin{pmatrix} \hat{\psi}_L \\ \hat{\psi}_R^\dagger \end{pmatrix}. \tag{6.10}$$

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix}. \tag{6.11}$$

---

[4]This definition comes from the fact that we know helicity must flip under parity transformations, and that spin also flips under parity transformations while normal four-vectors do not. So the projection of spin against a four-vector — of which velocity is the only available one — flips under helicity

The ψ_L and ψ_R components can be extracted back out of ψ via another matrix

$$\gamma^5 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}. \tag{6.12}$$

because the eigenvectors of this matrix are $\hat{\psi}_L$ and $\hat{\psi}_R$, with eigenvalues $+1$ and -1 respectively. So this matrix separates the left and right components. In particular, $\frac{1+\gamma_5}{2}$ selects out the left-handed component and removes the right-handed component, which will be crucial when we discuss the weak force.

With these changes in notation, the Lagrangian is now written as

$$\mathcal{L} = \psi^\dagger \gamma^0 (\gamma^\mu \partial_\mu - m)\psi. \tag{6.13}$$

In an effort to simplify things even further, a common approach is to define the new symbols $\overline{\psi} = \psi^\dagger \gamma^0$ and $\gamma^\mu \partial_\mu = \slashed{\partial}$. Then

$$\mathcal{L} = \overline{\psi}(i\slashed{\partial} - m)\psi. \tag{6.14}$$

This is called the *Dirac Lagrangian* and it is a fully consistent relativistic theory of a fermion. **JTD: Some i factors missing.**

The remainder of this book can be understood with just this information, but the reader might be curious as to whether introducing a new particle ψ_R was the only resolution to the problem of CPT invariance. A fascinating way to answer this question is to derive the Dirac Lagrangian and the γ matrices from a different perspective: we will find the generators of the Lorentz transformations. Lorentz transformations take the form

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \text{Time dilation} & \text{Boosts} \\ \hline \text{Boosts} & \text{Rotations} \end{pmatrix} \tag{6.15}$$

so that these generators will include the generators of rotation, which are the Pauli spin matrices. However, the additional generators due to boosts will force us to use the γ^μ matrices instead of the σ^μ matrices. The use of four-component spinors is therefore required for massive fermions.

Along the way, we will discover many useful properties of the γ^μ matrices, find how spinors transform under Lorentz transformations, and show that many forms of the γ^μ matrices exist beyond (6.11).

Generators of the Lorentz Group

A full four-dimensional sketch of boosts is given in (6.15), and an example of a boost in two dimensions is

$$\Lambda = \begin{pmatrix} \cosh\gamma & \sinh\gamma \\ \sinh\gamma & \cosh\gamma \end{pmatrix} \tag{6.16}$$

Our first task is to add generators for boosts to our rotation generators in order to create the full set of generators of the Lorentz boosts Λ. A glance at the above example shows that the boosts are almost rotations, but they are symmetric and the trigonometric functions are replaced with hyperbolic trigonometric functions. A little more thought reveals that the additional generators for the boosts are

$$B_x = i \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad B_y = i \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B_z = i \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \tag{6.17}$$

which can be used to verify the example. To write an algebra that contains both the boost generators $\boldsymbol{B}$ and the rotation generators $\boldsymbol{S}$ we define an

antisymmetric tensor of matrices Σ where the entries are generators. The timelike entries are boosts $\Sigma^{0i} = B^i$ and the spacelike components are rotations $\Sigma^{12} = S_z$, $\Sigma_{13} = -S_y$, and $\Sigma_{23} = S_x$[5]. The Σ tensor then satisfies the algebra

$$[\Sigma^{\mu\nu}, \Sigma^{\rho\sigma}] = i\left(\eta^{\nu\sigma}\Sigma^{\mu\rho} + \eta^{\mu\rho}\Sigma^{\nu\sigma} - \eta^{\mu\sigma}\Sigma^{\nu\rho} - \eta^{\nu\rho}\Sigma^{\mu\sigma}\right) \tag{6.18}$$

where $\eta^{\mu\nu}$ is the spacetime metric. This equation is called the *Lorentz algebra*.

We have done nothing except put the spin operators in relativistic language by adding in boost generators so far. Because there are now six generators, the Σ tensor can no longer be considered angular momentum. However, we can use a trick due to Dirac to reduce the tensor to four components by defining a four-vector of matrices γ^μ such that

$$\Sigma^{\mu\nu} = -\frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right]. \tag{6.19}$$

Then the Lie algebra of $\Sigma^{\mu\nu}$ implies the following Lie algebra for γ^μ:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \tag{6.20}$$

where $\{\cdot, \cdot\}$ is the anticommutator. (6.20) is called the *Clifford algebra* or the *Dirac algebra*, and its representation γ^μ are called the Dirac γ^μ matrices. One can manually check that (6.11) is a valid representation of the γ^μ matrices — i.e., it satisfies th Clifford algebra. Any other representation that satisfies the Clifford algebra is also valid but we will not need them for this book. Likewise, there is no point in giving the additional 6×6 and higher-dimensional representations of the Clifford algebra because we will not use them, but they do exist.

---

[5]This structure of the Σ tensor may seem odd, but it is very similar to the layout of the electromagnetic tensor, also known as the Faraday tensor, where the electric field is laid along the timelike components and the magnetic field along the spacelike components.

Lorentz Boosting Spinors

We have defined the γ^μ matrices with a spacetime index, so far just because there were four γ^μ matrices and spacetime is four-dimensional. However, this index is not just for show. It is a true spacetime index in that it indicates that the γ^μ matrices transform as vectors under Lorentz transformations.

To see this, we will first act on the γ^μ matrices with the spin representation of a boost Λ. This is sometimes called an "internal boost" because it affects only the spin. We can write Λ as an exponential using the $\Sigma^{\mu\nu}$ generators of the Lorentz algebra which we just worked out:

$$\Lambda = e^{-i\theta\mu\nu\Sigma^{\mu\nu}}. \tag{6.21}$$

In principle, $\theta_{\mu\nu}$ can be any tensor of constants. However, because $\Sigma^{\mu\nu}$ is antisymmetric, only the antisymmetric part of $\theta_{\mu\nu}$ will matter. Acting with Λ on the γ^μ matrices gives the transformation

$$\gamma^\mu \to \Lambda^{-1}\gamma^\mu\Lambda. \tag{6.22}$$

Remember that the $\Sigma^{\mu\nu}$ matrices do not commute with γ^μ.

If we were to transform the γ^μ as a spacetime vector rather than as a set of matrices, then we would get

$$\gamma^\mu \to \Lambda^\mu{}_\nu\gamma^\nu. \tag{6.23}$$

where this time, $\Lambda^\mu{}_\nu$ is a boost of spacetime, meaning that its entries are numbers and not matrices. This is called an "external boost" because it explicitly does not affect spin.

It helps to write the $\Lambda^\mu{}_\nu$ matrix as an infinitesimal boost, in which case we can use the generators of the Lorentz algebra that we derived above in (6.17), which we now call $J^{\mu\nu}$ to distinguish them from the generalized spin operators $\Sigma^{\mu\nu}$. Then $\Lambda^\mu{}_\nu = \delta^\mu_\nu + \left(\frac{1}{2}\omega_{\alpha\beta}J^{\alpha\beta}\right)^\mu{}_\nu$. A helpful

identity of γ^μ matrices is that

$$[\gamma^\alpha, \Sigma^{\mu\nu}] = (J^{\mu\nu})^\alpha{}_\beta \gamma^\beta \tag{6.24}$$

Using this fact, a few calculations using the Lorentz algebra (6.18) and Clifford algebra (6.20) give that the internal and external boosts are identical: that is,

$$\Lambda^{-1}\gamma^\mu\Lambda = \Lambda^\mu{}_\nu\gamma^\nu. \tag{6.25}$$

This confirms that γ^μ transforms like a four-vector.

Finding a Lorentz Invariant Mass Term

We have now derived the γ^μ matrices and shown that they can be contracted with ∂_μ, reproducing the massless Dirac Lagrangian. In our previous discussion, we needed to define $\overline{\psi} = \psi^\dagger\gamma^0$ in order to include a mass term so that $\overline{\psi}\psi$ is Lorentz invariant. The reason why $\psi^\dagger\psi$ is not Lorentz invariant can now be seen mathematically from the fact that γ^μ are not all hermitian so that Λ is not unitary. Thus, $\psi^\dagger\psi \to \psi^\dagger\Lambda^\dagger\Lambda\psi \neq \psi^\dagger\psi$ under a Lorentz transformation.

Inserting a γ^0 matrix, the true mass term transforms as $\overline{\psi}\psi \to \psi^\dagger\Lambda^\dagger\gamma^0\Lambda\psi$. We'll show that $\Lambda^\dagger\gamma^0\Lambda = \mathbb{1}$ with (6.24). Writing out an infinitesimal Lorentz transformation,

$$\overline{\psi}\psi \to \overline{\psi}\psi + i\theta_{\mu\nu}\psi^\dagger\left(\Sigma^{\mu\nu}\gamma^0 - \gamma^0(\Sigma^{\mu\nu})^\dagger\right)\psi \tag{6.26}$$

The goal is to show that

$$\Delta^{\mu\nu} = i\psi^\dagger\left(\Sigma^{\mu\nu}\gamma^0 - \gamma^0(\Sigma^{\mu\nu})^\dagger\right)\psi \tag{6.27}$$

is zero.

If μ and ν are both spacelike, then $(\Sigma^{\mu\nu})^\dagger = \Sigma^{\mu\nu}$ and the term in parentheses is just the commutator given in (6.24). $J^{\mu\nu}$ for spacelike indices are the rotation generators and never have timelike components, so

that $\Delta^{\mu\nu} = 0$. If one is spacelike and the other is timelike, then $(\Sigma^{\mu\nu})^\dagger = -\Sigma^{\mu\nu}$ for boosts, but (6.24) implies that $\Sigma^{\mu\nu}$ and γ^0 anticommute, so again $\Delta^{\mu\nu} = 0$. Thus, $\overline{\psi}\psi = 0$.

6.3   Spin-Statistics Theorem: Fermions Anticommute

In most quantum mechanics courses, one of two definitions is given for a fermion. The first is the definition given in the last section — that fermions are particles with half-integer spin. The second is that fermions obey the Pauli exclusion principle, meaning that no more than one can occupy a single quantum state.

Why these two definitions are equivalent is a mystery until the presence of relativity and CPT symmetry, in which case Pauli exclusion and half-integer spins are tied together in the *spin-statistics theorem*. In this section, we prove the spin-statistics theorem using a method due to Julian Schwinger. This gives us insight not only into the deep nature of Fermions, but also how to think of the Pauli exclusion principle for a free particle. We will use this insight to clarify the path integral definition in the case of fermions before deriving the fermionic Feynman rules in the next section.

The proof of the spin statistics theorem is as follows. Half-integer spin particles transform via SU(2) under a rotation whereas whole-integer spins transform via SO(3) as discussed in section 6.1. A key difference between these two groups is that SU(2) is "twice as big," meaning that a rotation in SO(3) is contained twice in SU(2). This is essentially because a rotation R in SO(3) can be represented with real matrices, whereas the same rotation R in SU(2) is represented with complex matrices. The complex conjugate of the complex matrices is another member of SU(2) but is the same rotation, so that R is contained twice.**JTD: Is this true?**

A consequence of this fact is that if we apply a full SO(3) rotation to

spacetime, the spin (transforming under the same rotation via 6.3) completes only a half-turn. We would need another full turn to cover the entire SU(2) spin group.

Suppose we have two spinors which are on opposite sides of the origin — $\hat{\psi}(x)R\hat{\psi}^\dagger(-x)$ — where one is rotated by a half-turn R. Another rotation by R switches their location while also rotating the spinors: $R\hat{\psi}(-x)R^2\hat{\psi}^\dagger(x) - R\hat{\psi}(-x)\hat{\psi}^\dagger(x)$. CPT inverting this expression gives $-R\hat{\psi}^\dagger(-x)\hat{\psi}(x)$. Since QFT is CPT and rotation invariant, the first and the last expressions are equal:

$$\hat{\psi}(x)R\hat{\psi}^\dagger(-x) = -R\hat{\psi}^\dagger(-x)\hat{\psi}(x). \tag{6.28}$$

Due to translation invariance, we could have placed these spinors at any pair of points x and y. It therefore follows that $\hat{\psi}(x)\hat{\psi}(y) + \hat{\psi}(y)\hat{\psi}(x) = 0$, or

$$\left\{\hat{\psi}(x), \hat{\psi}(y)\right\} = 0. \tag{6.29}$$

In other words, fermions *anti-commute*. This is the opposite of bosons, where $R^2\hat{\psi} = \psi$ and we have

$$\left[\hat{\phi}(x), \hat{\phi}(y)\right] = 0 \tag{6.30}$$

so that bosons *commute*. The anticommuting nature of fermions is directly responsible for the Pauli exclusion principle since, for $x = y$, it implies $\phi(x)\phi(x)^2 = 0$. Therefore, the operator that creates two particles at position x sends the vacuum to zero and these particles cannot be created.

6.4   Feynman Rules for Fermions

Now that we have the Lagrangian of free fermions (6.14) , we can solve the path integral to find nPCFs. This path integral is

$$\langle 0|\hat{\psi}(k_1)\cdots\hat{\psi}(k_n)\hat{\psi}^\dagger(q_1)\cdots\hat{\psi}^\dagger(q_m)|0\rangle = \frac{\int \mathcal{D}\overline{\psi}\mathcal{D}\psi\,\psi(k_1)\ldots\psi^*(q_m)e^{iS}}{\int \mathcal{D}\overline{\psi}\mathcal{D}\psi\,e^{iS}}$$

$$S = \int d^4x\,\overline{\psi}(i\slashed{\partial} - m)\psi$$

(6.31)

where ψ represents a four-component vector of complex numbers[6]. The eagle-eyed may notice that this equation has a deep flaw, which is that the $\hat{\psi}$ operators anticommute due to the spin-statistics theorem, but the ψ complex numbers in the path integral commute just as normal numbers do, which is inconsistent. This is fixed by defining the ψs in the integrand to be a new type of number — Grassmann numbers — which anticommute just like the $\hat{\psi}$ operators do. The properties of Grassmann numbers are outlined in appendix 20.4, but the only property we will really need is that Grassmann numbers still follow Wick's theorem.

For scalar fields in chapter 5, we used Wick's theorem (5.27) to write an nPCF in terms of the Green's function of the Lagrangian, $G(k)$. Since Wick's theorem holds for fermions too, all we need is this Green's function for the fermion Lagrangian.

Recall that the position-space Green's function $G(z)$ for a differential operator O_x is defined as $O_xG(z) = \delta(z)$. We showed in section 5.1 that, when O_x is symmetric under translations, the Green's function in momentum space is just given by $O(k)G(k) = 1$, where $O(k)$ is the original differential operator with ∂_μ replaced by $-ik_\mu$. For fermions, the

---

[6]ψ has four components, but it is not a four-vector. It is a spinor. Four-vectors and spinors transform differently under boosts and should be regarded as separate, as described in section 6.2

differential operator is $O_x = (i\slashed{\partial} - m)$, so $O(k) = (\slashed{k} - m)$ and

$$G(k) = \frac{i}{\slashed{k} - m - i\epsilon} = \frac{i(\slashed{k} + m)}{(\slashed{k})^2 - m^2 - i\epsilon} = \frac{i(\slashed{k} + m)}{k^2 - m^2 - i\epsilon} \qquad (6.32)$$

JTD: Why the i? where we used the fact that $(\slashed{k})^2 = k_\mu k_\nu \gamma^\mu \gamma^\nu$ and the the Clifford algebra (6.20) to show that $(\slashed{k})^2 = k^2$. The $i\epsilon$ term is inserted to remove the singularity which occurs for $k^2 = m^2$, just as we did for the scalar field case.

JTD: Give an example

Spin and Free Fermion Properties

Now that we've worked out the fermionic propagator, we can use the same diagram technology as for scalar fields to compute scattering amplitudes for fermions. However, using these amplitudes to predict probabilities for a scattering experiment includes one more step than for scalar theory. A scalar field merely holds one complex number at each location, so a scalar plane wave encodes momentum and no other properties. However, we had to introduce a Dirac spinor field $\hat{\psi}(x)$ in order to represent fermions, so fermionic plane waves encode both momentum and the value of this spinor, both of which are observed[7].

To get a scattering amplitude for fermions therefore, we multiply an external fermion leg by the value of the corresponding plane wave's observed spinor. This will always produce a complex number for the scattering amplitude because each fermion propagator has to connect to another fermion at its other end in order to pass on its charge, so that every incoming external fermion leg is connected through a chain of fermion propagators to an outgoing external fermion leg. The value of such a

---

[7]We usually refer to the "spin" (two component) as what is observed, while here we refer to the "spinor" (four component). The difference is subtle and usually not important in this book, but we discuss it in appendix 20.4.

chain is the incoming spinor times gamma matrices from the internal propagators times an outgoing spinor, which is a scalar.

Most of the time, the incoming and outgoing plane waves are unpolarized, meaning that they contain all spins. One could sum the value of the Feynman diagrams for all possible incoming and outgoing spinors, but this is mathematically equivalent to computing the trace of the γ^μ matrices in the internal propagators, as is proved in appendix 20.4. An example of such a trace calculation is given later.

Predictions for the momentum of a fermion are made using the same process as for scalars — draw a Feynman diagram and evaluate it, according to the rules in the previous rules and the supplementary rules specific to fermions in the section below.

Fermionic Quirks for Feynman Diagrams

Due to the presence of γ^μ matrices in the fermionic propagator and the fact that $\hat{\psi}$ operators anticommute, a few new rules must be taken into account when computing diagrams with diagrams with fermions. These changes are summarized below and explained in the following paragraphs, and examples are also given in the next section.

1. Fermionic propagators are directed. We should take the non-reversibility of these propagators into account when computing symmetry factors (since fermions are charged)

2. The propagators should be multiplied in order when traversing backwards along charge arrows (since γ^μ matrices do not commute)

3. If two diagrams are equivalent under odd permutations of the external vertices, they contribute opposite signs[8] (since fermions an-

---

[8] An odd permutation is a permutation which contains an odd number of swaps.

ticommute)

4. To sum over all possible spins, take the trace over products of γ^μ matrices (since the trace is rotation-invariant)

5. We should trace over all propagators inside a loop and multiply by -1 (since fermions anticommute)

Change one is due to the fact that $\psi = \overline{\psi}$, which is clear from the fact that ψ is a column vector and $\overline{\psi}$ is a row vector. They are different quantities and cannot be permuted. This is similar to the case of a complex scalar field, where ψ and ψ^* could not be permuted.

Change two is due to the convention that multiplication is conducted from right to left. We wrote the nPCF with the initial states on the left and the final states on the right, so we must also multiply the propagators in this order. This matters because γ^μ matrices do not commute, whereas for the scalar theory the propagator was a complex number and multiplication order did not matter.

Change three is due to the fact that fermions anticommute. External vertices in diagrams represent the $\hat{\phi}(k_i)$ in nPCFs, so permuting external vertices is equivalent to permuting spinors. Odd permutations result in an overall minus sign, so odd and even permutations should possess different signs. The overall sign doesn't matter because the phase of a scattering amplitude is not observable.

Change four deserves some more explaining. A cross section gives the probability to scatter from one state to another, and for fermions, those states include spin. Mathematically, one should include spin by multiplying each external propagator (which is a 4×4 matrix) by a four-component vector indicating the spin state of the external particle. It is common to represent the incoming particle as u and the outgoing particle as $\overline{u}$ (or v and $\overline{v}$ for antiparticles). Then the cross-section is $uM\overline{u}$ where

M is the product of the propagators between the particles[9]. However, most experiments cannot detect spin, so that all possible final u vectors should be *summed* over. They also cannot produce particles of one spin, so that all possible $\overline{u}$ vectors should be *averaged* over; Since spin vectors are "complete" in that they fully cover the four-dimensional unit sphere on which spin states live, the sum of $uM\overline{u}$ over their values is equal to $\text{tr}(M)$. The average is $\text{tr}(M)/4$ since there are four independent spin states.

The trace in the fifth change follows from the fourth change; a loop in a diagram represents the spontaneous creation and annihilation of a fermion, which may have any spin. We should therefore sum over all spins by taking the trace of the propagators of the loop. The minus sign comes from the fact that an implicit commutation of fermions must occur to calculate the loop. The Green's function which we use as a propagator is a 2PCF where the integrand is $\psi\overline{\psi}e^{iS}$. The order of ψ and $\overline{\psi}$ comes from the time ordering symbol of the nPCFs; the early time operator must come first. But when calculating a larger diagram, we traverse propagators backwards according to change number one. Thus, the integrand is $(\overline{\psi}_1\psi_1)(\overline{\psi}_2\psi_2)\ldots(\overline{\psi}_n\psi_n)e^{iS}$, where the subscript indicates which number propagator the fermions belong to. Wick's theorem requires us to write this integrand as a product of $\psi\overline{\psi}$s so that we can represent the integrand as a product of propagators, but this requires us to compute the last ψ_n to the front — $(\psi_n\overline{\psi}_1)(\psi_1\overline{\psi}_2)(\psi_2\ldots\overline{\psi}_n)e^{iS}$ — which is an odd number of permutations. The loop therefore contains a negative sign.

This gives us all the mathematical machinery needed to compute scattering amplitudes for a fermionic theory. For the rest of the chapter, we

---

[9]This spin state is usually obtained by first writing u in the particle's rest frame and then boosting it via the transformation laws for spinors discussed in section 6.2. In the rest frame, the left-handed and right-handed spin states are the same two-component vector, analogous to the non-relativistic, two-component particle spin. After the boost, the left- and right-handed components differ.

will work with a common interaction mechanism between scalars and fermions known as Yukawa coupling, using this machinery to compute the consequences the theory. The next section introduces Yukawa coupling and computes several simple diagrams showing particle decay to practice with the above changes to the Feynman rules.

6.5   Yukawa Theory: Decay in the Massive Scalar Limit

The Yukawa theory consists of a fermion and a real scalar particle together. The Lagrangian for such a theory is simply the sum of the scalar and fermion Lagrangians. We also add an interaction between the two particles, which goes as $\overline{\psi}\psi\phi$. This kind of coupling is called *Yukawa coupling*. The full Lagrangian is

$$\mathcal{L} = \frac{1}{2}|\partial_\mu\phi|^2 - m_\phi\phi^2 + \overline{\psi}(i\slashed{\partial} - m_\psi)\psi + g\overline{\psi}\psi\phi \qquad (6.33)$$

The Yukawa coupling is actually the only way we could couple a fermion to a scalar. $\phi\psi$ is disallowed because the spinor ψ is not dotted with anything and the Lagrangian cannot be spinor-valued. On the other hand, $\overline{\psi}\psi\phi$ already has dimension 4 so any additional particles would make it an irrelevant operator. This is one reason why Yukawa interactions are so studied. Another is that they appear in the Standard Model, as will be discussed later.

We'll represent the scalar ϕ particle as a dashed line and the fermion ψ as a solid line. The vertex corresponding to this new interaction is
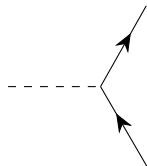
and contributes a factor of $-ig$ to the amplitude. You might remember this vertex from the previous section on scalar fields, but that vertex was a scalar-scalar interaction and this Yukawa interaction is a scalar-fermion interaction.

In this section, we'll assume a massive scalar: $m_\phi \gg m_\psi$. Intuitively, we know that having a massive scalar implies that the decay of one ϕ particle to two (or more) ψ particles is energetically possible. The rest of this section studies this decay.

The leading order diagram for $\phi \to \overline{\psi}\psi$ decay is just the interaction vertex
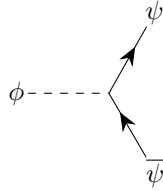
which has no internal legs, so it merely has a scattering amplitude of $-ig$. The next order diagrams are more interesting:
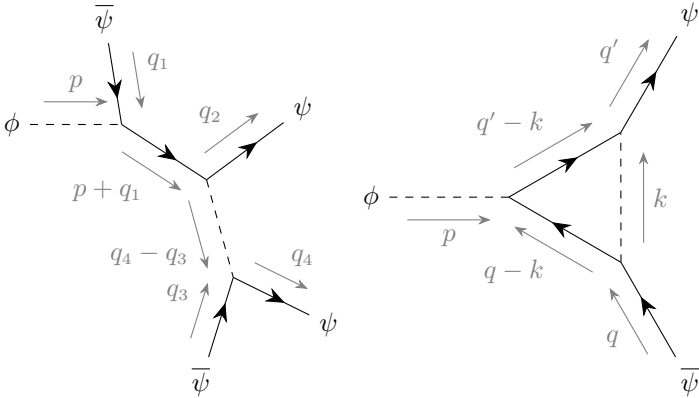
The first represents $\phi \to \bar{\psi}\psi\bar{\psi}\psi$ while the second is another $\phi \to \bar{\psi}\psi$ diagram. Both now have internal propagators, so their matrix elements are more complicated. The first diagram is

$$-iM_{\phi\to\bar{\psi}\psi\bar{\psi}\psi} = \frac{1}{4}(-ig)^3 \frac{i\,\mathrm{tr}(\slashed{p} + \slashed{q}_1 - m_\psi)}{(p+q_1)^2 - m_\psi^2} \frac{1}{(q_4-q_3)^2 - m_\phi^2} \quad (6.34)$$

The first fraction is the internal fermion propagator while the second is the internal scalar propagator. The factor of $1/4$ and the trace appear so that M is the spin-averaged cross section. Unfortunately, this diagram evaluates to zero because, as appendix 20.4 indicates, the trace of an odd number of γ^μ matrices is zero. Thus, $\phi \to \bar{\psi}\psi\bar{\psi}\psi$ is a higher-order process.

The second diagram is not zero, however. It has three internal legs arranged in a loop, meaning that there is one free momentum to integrate over.

$$-iM_{\phi\to\bar{\psi}\psi} = -ig + \frac{(-ig)^3}{4}\int\frac{d^4k}{(2\pi)^4}\,\mathrm{tr}\left[\frac{i(\slashed{q}' - \slashed{k} - m_\psi)}{(q'-k)^2 - m_\psi^2}\frac{i(\slashed{q} - \slashed{k} - m_\psi)}{(q-k)^2 - m_\psi^2}\right]$$
$$\times \frac{1}{k^2 - m_\phi^2}.$$
$$(6.35)$$

There were two fermion propagators this time, which we computed in backwards order. The trace is another spin average. Using the fact that $\mathrm{tr}(\gamma^\mu\gamma^\nu) = 4\eta^{\mu\nu}$,

$$-iM_{\phi\to\bar{\psi}\psi} = -ig - ig^3\int\frac{d^4k}{(2\pi)^4}\frac{(q'-k)\cdot(q-k) + m_\psi^2}{((q'-k)^2 - m_\psi^2)((q-k)^2 - m_\psi^2)}\frac{1}{k^2 - m_\phi^2}.$$
$$(6.36)$$

$$M_{\phi\to\bar{\psi}\psi} = g + g^3\int\frac{d^4k}{(2\pi)^4}\frac{(q'-k)\cdot(q-k) + m_\psi^2}{((q'-k)^2 - m_\psi^2)((q-k)^2 - m_\psi^2)}\frac{1}{k^2 - m_\phi^2}.$$
$$(6.37)$$

Although the integral is complicated, it depends on few variables. Conservation of momentum dictates that $q' + q$ is the initial momentum of the ϕ particle, and if we work in the center of mass frame this momentum is $q' + q = (m_\phi, \mathbf{0})$. Thus, $q = (m_\phi/2, \mathbf{q})$ and $q' = (m_\phi/2, -\mathbf{q})$ where $m_\phi^2/4 - (\mathbf{q})^2 = m_\psi^2$. Thus, $M_{\phi \to \bar{\psi}\psi}$ depends only on m_ψ and m_ϕ.

To leading order in $m_\psi/m_\phi \ll 1$, the cross section is

$$M_{\phi \to \bar{\psi}\psi}^{\text{2D}} \approx g + \frac{g^3}{8\pi}\left(\frac{m_\psi}{m_\phi}\right)^2 \qquad M_{\phi \to \bar{\psi}\psi}^{\text{3D}} \approx g + \frac{g^3}{8\pi}\left(\frac{m_\psi}{m_\phi}\right) \qquad (6.38)$$

in two and three spacetime dimensions respectively. These answers can be confirmed numerically, or computed analytically using the tools of chapter 8. Using the conversion from scattering amplitude to decay rate given in (2.13), we can triumphantly write our first lifetime for a particle:

$$\tau_\phi^{\text{2D}} = \frac{2m_\phi}{g}\left(1 - \frac{g^2}{8\pi}\frac{m_\psi}{m_\phi}\right) \qquad \tau_\phi^{\text{3D}} = \frac{2m_\phi}{g}\left(1 - \frac{g^2}{8\pi}\left(\frac{m_\psi}{m_\phi}\right)^2\right) \quad (6.39)$$

JTD: This has the wrong dimensions Quantum field theory therefore successfully describes particle decay — the first ever consistent theory of physics to do so.

We have given these answers in unphysically low dimensionality because $M_{\phi \to \bar{\psi}\psi}$ diverges as the number of spacetime dimensions approaches four. This divergence also occurred in the case of the scalar particle in chapter 5, and in general occurs for a particle which has a dimensionless coupling constant (λ for scalar theory; g for Yukawa theory). This was a serious problem historically in developing QFT as mentioned in chapter 1, eventually fixed using the techniques we introduce in chapter 8. The divergences also have a physical origin which can be understood already as follows:

At the beginning of this section, we justified the Yukawa Lagrangian by stating that any additional interactions are so high in dimension that their cross sections would be proportional to energy (or a positive power

of energy). When using two or three spacetime dimensions, even the Yukawa coupling becomes too high in dimensions and we see that $M_{\phi \to \overline{\psi}\psi}$ is proportional to $1/m_\phi^{4-d}$, where d is the number of spacetime dimensions. Since $1/m_\phi$ is small, this confirms our intuition that high-dimensionality interactions are irrelevant in this scenario.

But in the four dimensional case, this trend requires that $M_{\phi \to \overline{\psi}\psi}$ is proportional to $1/m_\phi^0$ — that is, it is a pure number depending only on g. However, this cannot be true since m_ϕ cannot decay if $m_\phi < 2m_\psi$ due to energy conservation[10]. Its lifetime therefore must depend on m_ϕ. The divergence of the amplitude is a mathematical indication of this contradiction.

The decay studied in this section occurs in the Standard Model, where the Higgs boson (mass $125\,\text{GeV}$) takes the role of the scalar particle and the bottom quark (mass $4\,\text{GeV}$) takes the role of the fermion (though many other decay paths are possible). The decay is so quick that the Higgs has a lifetime of a few times $10^{-22}\,\text{s}$, so that it would travel only a few proton widths before decaying if moving at the speed of light, making it difficult to observe.

Since we've already seen these diverging amplitudes in two theories, we will soon describe how to resolve them. Before this, however, we have two more problems to discuss. One is to further explore the properties of fermions in the remainder of this chapter. The other is to describe spin-one particles, such as light, in the next chapter.

---

[10]Even the low-dimensional lifetimes of (6.39) predict decay for $m_\phi < 2m_\psi$, which is incorrect by this logic. But they were computed to leading order in m_ϕ/m_ψ, and the higher orders will resolve the conflict by forcing the lifetime to blow up as m_ϕ nears $2m_\psi$. This defense mechanism cannot occur in four spacetime dimensions because m_ϕ and m_ψ do not appear in the lifetime.

6.6   Yukawa Potential & the Meson Model of the Strong Nuclear Force

In this section, we'll continue using the Yukawa Lagrangian (6.33) of the previous chapter, but now we'll consider the opposite limit of the mass ratio, where $m_\psi \gg m_\phi$. We will see that in this limit, the bosonic ϕ particle acts as a carrier of a long-distance, attractive force between the fermions which constituted the first model of the atomic nucleus. The mechanism of producing long range forces with bosons occurs quite often in the Standard Model; of the five known fundamental bosons, four are force carriers. Electromagnetism is the most famous example of a long-range force carried by the massless photon. The exception is the Higgs boson, which is quite massive and therefore decays before it carries forces, as discussed in the previous section. Light bosons have use even outside QFT; other theories of physics that use field theory techniques sometimes introduce a light boson to carry long-range forces.

   The interaction we'll consider is two heavy fermions which interact by passing a boson from one to the other:
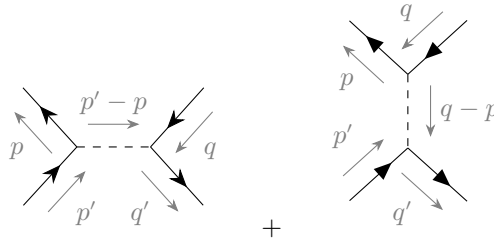
These diagrams' amplitudes are simply

$$-iM = \frac{(-ig)^2}{(p'-p)^2 - m_\phi^2} + \frac{(-ig)^2}{(q-p)^2 - m_\phi^2} \tag{6.40}$$

In section 5.4, we studied similar diagrams and made use of Mandelstam variables (5.51) to simplify the expression. There we had three diagrams,

but conservation of charge forbids one diagram (the u-channel), so we just have

$$M = -ig^2 \left(\frac{1}{s^2 - m_\phi^2} + \frac{1}{t^2 - m_\phi^2} \right). \tag{6.41}$$

The physical consequences of this scattering amplitude are not immediately clear, so let us imagine it in the context of an experiment. Suppose that two beams of distinguishable fermions A and B are fired together and the momenta of the final states are measured. A physicist in the 1920s would have analyzed this situation using non-relativistic quantum mechanics, and determined using the Born approximation that

$$M = -2\pi V(\boldsymbol{p} - \boldsymbol{q})\delta(E_{\boldsymbol{p}} - E_{\boldsymbol{q}}) \tag{6.42}$$

where V is the Fourier transform of the potential between the particles. This equation was derived in 2. **JTD: Derive this equation** Equating this amplitude to the QFT amplitude above **JTD: Why the i?** connects the relativistic notion of a three-particle Yukawa interaction to the non-relativistic (and familiar) notion of a potential.

Since the particles in our experiment are assumed to be distinguishable, the t-channel which mixes particle types should be neglected. Thus, equating (6.41) and (6.42) gives

$$V(\boldsymbol{k}) = -\frac{g^2}{\boldsymbol{k}^2 - m_\phi^2}. \tag{6.43}$$

This potential can be converted back to position-space using the tools of complex analysis (discussed in chapter 19) to get a spherically symmetric potential:

$$V(r) = -\frac{g^2 e^{-m_\phi r}}{4\pi r}. \tag{6.44}$$

This attractive potential is called the *Yukawa potential*. It dies of quickly for $r \gg m_\phi$, but for $r \ll m_\phi$ it behaves like the $V(r) \sim 1/r$ Coulomb potential. Indeed, if $m_\phi = 0$, the Yukawa potential becomes the Coulomb

potential, though this prediction is difficult to achieve in practice because ϕ is unlikely to be massless, as discussed in 5.8[11].

The fact that QFT can produce this long-range potential is at first surprising. A QFT Lagrangian can only include interactions between fields at the same spacetime coordinate, making it difficult for ψ fields at different locations to reliably respond to distant fluctuations. But with the addition of a boson light enough to propagate long distances, we showed that the particles may communicate via the boson. Not only does this interaction reproduce the phenomenon of non-relativistic quantum mechanics; it also forces the interactions to be causal, since the force-carrying boson cannot move faster than the speed of light.

With that said, one could have expected that a correct physical theory should predict long-range forces because electromagnetism is a long-range force. But the Yukawa potential is crucially different from electromagnetism because it does not possess charges. All nucleons attract each other, whereas in electromagnetism particles of the same charge repel. This property is necessary to create a nucleus, which are observed to contain hundreds of particles for millennia without ejecting them. In fact, the Yukawa potential was invented specifically for the purpose of modeling the nucleus by Hideki Yukawa in 1935. He predicted the existence of a hitherto undiscovered scalar boson known as a *meson* to mediate the nucleon-nucleon force termed the *strong force*, or strong nuclear force. Soon after his prediction, the muon was discovered and thought to be the meson in question, though it was later determined to be an unrelated particle[12]. Soon afterwards a true meson — the pion — was discovered and Yukawa won the Nobel prize in 1949.

---

[11]The true Coulomb potential comes from a three-point interaction between fermions and a spin-one particle, rather than this Yukawa interaction between fermions and a spin-zero particle.

[12]In *The Feynman Lectures*, Richard Feynman refers to the muon as the "mu meson" because the fact that the muon does not interact through the strong force.

Somewhat surprisingly, the discoveries continued and many more mesons were discovered until the list became too large to name. This may be surprising because the Yukawa interaction requires only one meson to exist, but now we understand that the Yukawa interaction is approximate. Protons, neutrons, and mesons are really made up of quarks, and the true Feynman diagram of nucleon-nucleon interactions is as follows

JTD: Diagram

which looks like a Yukawa interaction diagram unless the substructure of the proton, neutron, or meson are revealed. The vast family of mesons exist merely because there are many ways to combine quarks into distinct particles.

The accurate theory of quarks and mesons is called Quantum Chromodynamics, or QCD, and we discuss it in chapter 14 after we have developed the mathematics necessary to understand them. That theory was layed out int he 70s **JTD: check** , delayed due to its complexity. We'll therefore focus on the simpler theory of light — and spin-one particles in general — in the next chapter, moving on to the first quantum theory of electromagnetism in the following chapter.

Problems

6.1   Verification of the Clifford algebra. Show that the Clifford algebra (6.20) satisfies the Lorentz algebra (6.18) under the definition of the γ^μ functions in (6.19).

6.2   Relativistic Stern-Gerlach Experiment.

Chapter 7

Spin One

In the last chapter, we saw that a massless scalar boson can carry a long-range force between two fermions with a potential of $V(r) \propto 1/r$. This is similar to the electric potential, but a magnetic field is missing. In order to achieve a theory for the magnetic field, we propose a new type of particle named a vector boson which consists of a vector field $A^\mu(x)$, on analogy with the four-vector of electromagnetism. This particle can be called "spin one" because it transforms as a vector under Lorentz boosts and rotations, and the only SU(2) or SO(3) representation that rotates like a four vector is the spin-one representation. Our only problem is that the Lagrangian for this field is so high in dimension that A^μ should not be observable.

We address this dimensionality problem by imposing a gauge symmetry on A^μ which forbids the offending terms from the Lagrangian. Gauge symmetry is an intrinsic property even of classical electromagnetism, so it is no surprise that it should reappear in QFT. After writing down the possible Lagrangians of A^μ and how it could couple to a fermion while obeying gauge symmetry, we use the combined vector-fermion theory to reproduce the Coulomb potential between fermions, proving that A^μ is a photon. Furthermore, the theory predicts that the

fermion should have a magnetic dipole moment of about 2, which is true of the electron. Finally, we show that a massless vector particle does not suffer from the same problems as a scalar particle; its masslessness is not ruined by including higher order Feynman diagrams. We're therefore safe to propose that this massless vector particle is the photon, the fermion is the electron, and the theory is a valid theory of Quantum Electrodynamics, or QED.

7.1   Gauge Invariance

For a field A^μ, the simplest kinetic term one can put in the Lagrangian is $\partial_\mu A^\mu$, and the simplest potential term is $A^\mu A_\mu$. As with the scalar and fermion fields, we normalize the Lagrangian by giving the kinetic term a coefficient of one while all other terms get their own constant:

$$\mathcal{L} = \partial_\mu A^\mu - m A^\mu A_\mu. \tag{7.1}$$

where m is the particle mass. The first term requires that A^μ have dimension 3, so that m has dimension -2. Therefore, any dynamics involving this second term are not observable at low energies.

This is a serious problem for any physical model with a vector boson, and seemingly the only solution is to replace the kinetic term $\partial_\mu A^\mu$ with a different term, so that the dimensionality of A^μ could change. But why should $\partial_\mu A^\mu$ not exist in the Lagrangian when for the scalar and fermion particles every simple term that could exist in the Lagrangian did so?

We may take inspiration from the classical definition of the vector potential, which is that A^0 is the electric potential and A^i or $\boldsymbol{A}$ is the vector potential, meaning the electric and magnetic fields are $\boldsymbol{E} = -\nabla A^0 - \partial_t \boldsymbol{A}$ and $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. These $\boldsymbol{E}$ and $\boldsymbol{B}$ fields are observable, while $\boldsymbol{A}$ is not.

Suppose we make the following change to A^μ

$$A^0(x) \to A^0(x) - \partial_t \lambda(x) \qquad \boldsymbol{A}(x) \to \boldsymbol{A}(x) + \nabla \lambda(x) \tag{7.2}$$

where $\lambda(x)$ is an arbitrary function. In spacetime notation, this is $A^\mu(x) \rightarrow A^\mu(x) + \partial^\mu \lambda(x)$. Since derivatives commute and the curl of a gradient is zero, the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are unchanged; that is, there is no physical change to the system. We might say that the above transformation is a "symmetry" of the system.

The symmetries we have discussed so far are global symmetries, meaning that they refer to actions taken on the entire field at once. For example, CPT symmetry is a global symmetry because it states that physics is invariant if one flips the charge, parity, and time of everything in the universe, not just the particles in a single region. (7.2) is different because the field $\lambda(x)$ can be different at every position; it may even be zero everywhere except in a small region. We call these local symmetries *gauge symmetries*.

We've referred to symmetries such as the SU(2) symmetry of spin by their group before. The group for the gauge symmetry of electromagnetism must have only one generator because $\lambda(x)$ was the sole generator of (7.2). The only such group is U(1), where the U stands for "Unitary." This group is usually represented as rotating fields by a phase $e^{i\alpha}$, where α is the Euler angle and the generator is one. In (7.2) the representation is more complicated than a phase, but the single generator guarantees that the symmetry group is still U(1).

Enforcing this gauge symmetry of classical electromagnetism as a symmetry of the new quantum field A^μ will destroy the $\partial_\mu A^\mu$ kinetic term. If we perform the transformation of (7.2) on $\partial_\mu A^\mu$, we get

$$\partial_\mu A^\mu \rightarrow \partial_\mu A^\mu + \partial_\mu \partial^\mu \lambda(x) \tag{7.3}$$

which is not equal to $\partial_\mu A^\mu$. The term is therefore not gauge invariant and cannot be included.

A useful tool for writing gauge invariant equations is to define the

Faraday tensor, which is an antisymmetric tensor of fields defined as

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{7.4}$$

which is invariant under Gauge transformations. This allows for a gauge invariant kinetic term $F^{\mu\nu}F_{\mu\nu}$. However, no potential term is possible because $A^\mu A_\mu$ is not gauge invaraint. Thus, our new Lagrangian for a vector boson is

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{7.5}$$

where the $\frac{1}{4}$ factor is a matter of convention. This particle is required to be massless since there is no second term, but in the next section we will show that the addition of an fermion admits another, low-dimensional term, proving that the dynamics of A^μ are observable.

7.2   The Fermion-Vector Boson Lagrangian

Now that we have understood the importance of respecting Gauge freedom, we can more thoroughly explore the possible Lagrangians of a photon field. The above Lagrangian contains the only two low-dimension terms for a vector particle by itself, but if we include a fermion, things get more complicated. Any term must be invariant under a spin rotation, so it should contain γ^μ and contract ψ with $\overline{\psi}$. It should also be Lorentz invariant, so all indices should be contracted. The only low-dimension term satisfying both constraints is

$$ie\overline{\psi}\gamma_\mu A^\mu \psi = e\overline{\psi}\slashed{A}\psi$$

where e is a constant with dimension $\frac{1}{2}$. The last thing to check is whether this term is invariant under the new gauge U(1) symmetry, but we cannot do this until we determine how ψ transforms under the gauge symmetry.

In the previous two chapters on fermions and scalars, we required that the Lagrangian be invariant under $\psi \rightarrow e^{i\alpha}\psi$, where α is a constant.

This is a global U(1) symmetry, global because α does not depend on position. We also associated this phase with charge, because doing so allowed us to use the CPT theorem to justify that ψ and $\overline{\psi}$ (or ϕ ad ϕ^* for complex scalars) are antiparticles. Since we're pursuing a quantum version of electromagnetism, it is very tempting to connect these U(1) global and U(1) gauge symmetries.

Let's promote the U(1) global symmetry of the fermion ψ to a U(1) gauge symmetry. That is, we require that the action be invariant under

$$\psi \rightarrow e^{ie\lambda(x)}\psi \tag{7.6}$$

where $\lambda(x)$ is the same $\lambda(x)$ as in (7.2). We can then explicitly compute how $e\overline{\psi}\slashed{A}\psi$ transforms under gauge transformations:

$$
\begin{aligned}
e\gamma_\mu\overline{\psi}A^\mu\psi &\rightarrow e\gamma_\mu\overline{\psi}\left[e^{-ie\lambda(x)}\left(A^\mu + \partial^\mu\lambda(x)\right)e^{ie\lambda(x)}\right]\psi \\
&= e\gamma_\mu\overline{\psi}A^\mu\psi + e\gamma_\mu\overline{\psi}\psi\partial^\mu\lambda(x)
\end{aligned}
\tag{7.7}
$$

which is not gauge invariant. But this promoting the global U(1) to a gauge U(1) also rendered the fermion part of the Lagrangian non-invariant, so we must check how that transforms too:

$$
\begin{aligned}
\overline{\psi}\left(i\slashed{\partial} - m\right)\psi &\rightarrow i\gamma_\mu\left(\overline{\psi}e^{-ie\lambda(x)}\right)\partial^\mu\left(e^{ie\lambda(x)}\psi\right) - m\overline{\psi}e^{-ie\lambda(x)}e^{ie\lambda(x)}\psi \\
&= i\overline{\psi}\slashed{\partial}\psi + ie\gamma_\mu\overline{\psi}\left(i\partial^\mu\lambda(x)\right)\psi - m\overline{\psi}\psi \\
&= \overline{\psi}\left(i\slashed{\partial} - m\right)\psi - e\gamma_\mu\overline{\psi}\psi\partial^\mu\lambda(x).
\end{aligned}
\tag{7.8}
$$

This term is also not gauge invariant, but adding it with the previous term causes the extra contribution from the gauge transformations to cancel, so that the sum (and therefore the total action) is gauge invariant.

To reprise, the full Lagrangian for a vector boson plus an electron is

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \overline{\psi}\left(i\slashed{\partial} - m\right)\psi + e\overline{\psi}\slashed{A}\psi. \tag{7.9}$$

though it is sometimes written as

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \overline{\psi}\left(i\not{D} - m\right)\psi \tag{7.10}$$

where $D_\mu = \partial_\mu - ieA_\mu$ is known as the *covariant derivative*.

The same logic also applies to a complex scalar boson, which is why we referred to such particles as charged. There, the Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}|D_\mu\phi|^2 - m|\phi|^2 \tag{7.11}$$

with the same covariant derivative D_μ as before. One can verify manually that this Lagrangian is also gauge invariant.

For several chapters, we have claimed without justification that complex conjugation represents charge reversal. This is necessary to understand why ϕ and ϕ^* are antiparticles for a complex scalar field ϕ, and why ψ and $\overline{\psi}$ are antiparticles for a fermion ψ. U(1) gauge symmetry was the missing link. In the above, we demonstrated that the phase of these particles determines how they couple with the vector boson A^μ, and when we show that A^μ is a photon, then the coupling constant e will manifestly be the charge of the particle. You can check that if we conjugate ψ by replacing it with $\overline{\psi}$ in the above argument, the coupling constant e in the covariant derivative becomes $-e$, proving that complex conjugation reverses charge.

7.3   Fermions, Gauge Bosons, and Noether's Theorem

In this chapter we started with a vector boson and introduced a U(1) gauge symmetry, inspired by electromagnetism, in order to make the boson observable. But it is also possible to perform the same task backwards: one can promote any global continuous symmetry to a gauge symmetry and create a vector boson known as a *gauge boson*. We'll review that process in this section.

When we "gauged" the U(1) symmetry of the fermion by promoting its $e^{i\alpha}$ global symmetry of ψ to an $e^{i\lambda(x)}$ gauge symmetry, we made the kinetic term of the fermion no longer gauge invariant. This was because the kinetic term involves a derivative $i\overline{\psi}\slashed{\partial}\psi$ which acted on the $\lambda(x)$ in the gauge transformation and did not act on the constant $e^{i\alpha}$ global transformation. The same derivative will act on any gauge symmetry, even one which isn't U(1).

In the above section, the gauged symmetry required modifying this offending derivative to a covariant derivative D_μ which contains a contribution from the new gauge boson. After adding a kinetic term for the gauge boson, the action was gauge invariant.

The form of D_μ derives directly from the symmetry being gauged. For U(1) we subtracted a new term ieA_μ but for a general symmetry with n generators $\tau_1, \ldots, \tau_n$, we will have to create n new vector bosons and add define the covariant derivative as

$$D_\mu = \partial_\mu - ie\tau_1 A^1_\mu - \cdots - ie\tau_n A^n_\mu. \tag{7.12}$$

Physically, the gauging of this symmetry creates a long-range force between fermions. If the phase of fermions were a global symmetry, then positrons and electrons would be identical except in their parity and time properties and would never interact. But electromagnetism arose when we gauged that symmetry. Every vector boson in the standard model is a consequence of a different symmetry being gauged, and all result in forces between fermions.

Gauging a global symmetry is an excellent tool for simplifying the standard model. One needn't explain why vector bosons are there if a gauge symmetry *requires* them to be there. It was also a crucial historical development, since seemingly bizarre interactions such as the weak force could be explained by proposing a global symmetry and gauging it.

Later in this book, we will explain weak interactions by gauging isospin symmetry, and strong interactions by gauging color symmetry.

This connection between interactions and gauge symmetries was first proposed in 1915, well before the invention of QFT, by Emmy Noether. A mathematician by training and profession, Noether published her incredibly influential Noether's Theorem during a brief six-month break from math, though she was prevented from publishing the result for several years due to institutional sexism against her and her work. The theorem states that any physical theory a gauge symmetry contains a corresponding conserved charge (a charge for which she gives a formula). Our above statement connecting gauge symmetries to vector bosons is same statement in a QFT context, since vector bosons conserve charge.

7.4   Feynman Rules and Conserved Quantities

Before we compute any scattering amplitudes involving vector bosons we need to find the Feynman rules for the new particle. Vector bosons are typically denoted with wavy lines in diagrams, so that the $e\overline{\psi}\slashed{A}\psi$ interaction term is depicted as

$$= -ie\gamma^{\mu}.$$
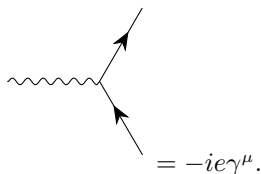
When we introduced the fermion, we used the arrows to distinguish between ψ and $\overline{\psi}$. Now we can also interpret them as arrows pointing in the direction of the flow of charge, and the vector boson interacts with these charge arrows because it carries the charge force. Furthermore, the vector bosons don't have arrows on them but they do have spacetime indices. These indices are contracted with the γ^{μ} of a vertex when they meet fermions.

To find the propagator of the vector boson we first find the differential operator corresponding to the kinetic term: need to compute the Green's function of the differential operator in the Lagrangian. The kinetic term in the Lagrangian can be written as

$$
\begin{aligned}
-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} &= -\frac{1}{4}\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)\left(\partial^\mu A^\nu - \partial^\nu A^\mu\right) \\
&= -\frac{1}{2}\left((\partial_\mu A_\nu)^2 - \partial_\nu A_\mu \partial^\mu A^\nu\right) \\
&= \frac{1}{2}A_\mu\left(\partial^2 \delta^\mu_\nu - \partial_\nu \partial^\mu\right)A^\nu \\
&= \frac{1}{2}A_\mu\left(\partial^2 \delta^\mu_\nu\right)A^\nu
\end{aligned}
\tag{7.13}
$$

where the third line comes from integration by parts, which we may use because this kinetic term is always integrated over in the action, $S = \int d^4x\,\mathcal{L}$. The last line acknowledges that $\partial_\mu \partial^\nu \to -\partial_\mu \partial^\nu$ with a second integration by parts, and therefore this term will vanish in the action. Therefore, the differential operator for a vector boson is just the scalar field operator times δ^μ_ν and the Green's function follows the same pattern. It is convenient to write the propagator with both indices down for a vector boson photon, so

$$
G_{\text{spin 1}}(p^2) = -\frac{g_{\mu\nu}}{p^2 + i\epsilon}
\tag{7.14}
$$

is the propagator of a vector boson.

Polarization

Finally, there is the question of how a vector boson plane wave manifests itself in experiment. The value of A^μ for a plane wave of vector bosons is observable just as the value of a spinor was observable for a plane wave of fermions. For vector bosons, this value ϵ^μ is called the polarization.

However, different ϵ^μ vectors might correspond to the same physical system. For example, the vector's normalization is not observable.

To remove ambiguity, we therefore normalize polarization according to $\epsilon^\mu \epsilon_\mu = -1$. Another ambiguity is caused by gauge invariance, which we could choose to remove by picking a gauge. For example, we already discussed that $\partial_\mu A^\mu$ is not gauge-invariant, so if we set $\partial_\mu A^\mu = 0$ we have broken the invariance and picked a gauge known as the Lorentz gauge. If this derivative is computed for a plane wave, it states that $p_\mu \epsilon^\mu = 0$, meaning that polarization is *transverse*. **JTD: This isn't fully right yet** .

Thus, if a vector photon were traveling along the z-axis, the possible polarization vectors are superpositions of $\epsilon^\mu = (0, 1, 0, 0)$ and $\epsilon^\mu = (0, 0, 1, 0)$. To compute a Feynman diagram, the polarizations of the external legs should be multiplied onto the $g_{\mu\nu}$ terms in the vector propagators, just as spinors are multiplied onto the Dirac γ^μ matrices of fermion propagators. If the internal or external plane waves are unpolarized, we can instead multiply the diagram by $-g^{\mu\nu}$ where μ and ν are the indices corresponding to those external legs.

7.5   Coulomb Potential

Let's address any ambiguity from the previous section by computing a Feynman diagram. Inspired by the near-reproduction of the Coulomb potential in the previous chapter, we'll study the same diagram, but replace the force-carrying scalar with a force-carrying gauge boson:
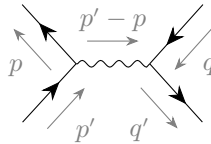
The contribution from this diagram is

$$iM = \frac{1}{4} \operatorname{tr}\left[(-ie\gamma^\mu)\frac{-g_{\mu\nu}}{p^2 + i\epsilon}(-ie\gamma^\nu)\right]. \tag{7.15}$$

As you can see, the indices of the propagator are contracted with the vertex so that M remains a scalar. The trace indicates that we assume the scattering beams do not have a coherent spin. Since the trace of $\gamma^\mu g_{\mu\nu} \gamma^\nu$ is 4,

$$iM = \frac{e^2}{p^2 + i\epsilon}. \tag{7.16}$$

This amplitude is identical to the Coulomb potential amplitude we computed for a scalar force-carrier except with the opposite sign, meaning that in the vector boson case positrons and electrons are attracted rather than repelled, with a Coulomb potential of

$$V(r) = -\frac{e^2}{4\pi r} \tag{7.17}$$

If we thought of the vector boson as a photon and the fermion as the electron, this Coulomb potential is exactly that of electromagnetism! Furthermore, if the fermion beams had a coherent spin then the γ^μ and γ_ν matrices flip the spin of each particle, and the $g_{\mu\nu}$ term ties together the spin flips so that both occur simultaneously and equally across all three spatial axes. Thus, the "photon" can be thought of as transferring spin, inducing a total spin change of 1 in both particles. This is another demonstration of why vector bosons are called spin one.

The fact that the vector boson interacts with spin cements our suspicion that a photon is a vector boson. Interaction with spin is the defining property of magnetic fields as we discuss in the next section, and we've already shown that photons interact with charge similarly to an electric field. Thus, vector bosons are true electromagnetic waves, and 7.11 can truly be thought of as the Lagrangian describing photons and electrons, and the electromagnetic force. In the context of quantum field theory, we call the theory Quantum Electrodynamics, or QED.

7.6   Magnetic Dipole Moment

Another interesting interaction to consider is the scattering of an electron off of an ambient electric or magnetic field. The leading order diagram is simply
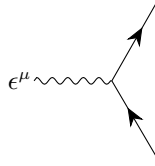
which contributes the amplitudes

$$iM = -ie\epsilon_\mu \overline{v}\gamma^\mu u \tag{7.18}$$

where u and $\overline{v}$ are the spinors of the incoming and outgoing electrons. **JTD: Where's the electric force?**   This is reminiscent of the Hamiltonian for spin-magnetic field interaction in non-relativistic quantum mechanics:

$$H = -g\frac{e}{2m}\boldsymbol{B}\cdot\boldsymbol{\sigma} \tag{7.19}$$

The g-factor is a unitless constant measured by experiment. It has no official name, but the quantity $g\frac{e}{2m}$ is referred to as the *magnetic moment*. where e is the electron charge, m is its mass, $\boldsymbol{\sigma}$ is the spin operator, and $\boldsymbol{B}$ is the magnetic field. Applying the rules of non-relativistic quantum mechanics to compute a scattering amplitude would remove the mass dependence, yielding

$$iM = -ie\frac{g}{2}\boldsymbol{B}\cdot\boldsymbol{\sigma}. \tag{7.20}$$

Though they use different notation, (7.18) and (7.20) are very similar. They have the same sign, are both proportional to e, and both depend on the dot product of the magnetic field (either $\boldsymbol{B}$ or ϵ^μ) with spin (either $\overline{v}\gamma^\mu u$ or $\boldsymbol{\sigma}$). Averaging over either spin or polarization would set both

equal to zero. This leaves only one difference between (7.18) and (7.20), which is the presence of g. In fact, if $g = 2$, then the QFT scattering amplitude would exactly equal the non-relativistic scattering amplitude.

When this $g = 2$ conclusion was originally published by Paul Dirac, it represented the first ever correct theoretical determination of the magnetic moment of the electron. A classical model in which the electron is a spinning solid sphere with uniformly distributed charge would yield (incorrectly) $g = 1$. Thus, the $g = 2$ conclusion is a fundamental result of adding relativity to quantum mechanics.

If we pressed on to higher order Feynman diagrams, however, we would be forced to break modify this conclusion. The next order diagram is
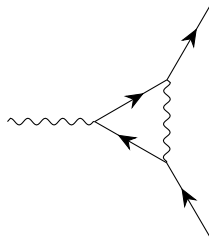
which would both contribute a correction to e and to g in order to maintain the equivalence of (7.18) and (7.20). These corrections both have important effects. The change in e is difficult to measure unless an experiment is conducted in a very small space, where this second order diagram can be suppressed, in which case the uncorrected value of e is observed instead. This effect is responsible for the Lamb shift in the spectrum of hydrogen, where the electron is extremely close to the nucleus.

The correction to g is easier to measure since the uncorrected value $g = 2$ is known already. QED's prediction for the $g - 2$ correction is a great achievement of QED and is confirmed by experiment to many

decimal places. In the next chapter, we develop the mathematical tools necessary to make this prediction.

In the next section we'll discuss a somewhat different but equally important historical achievement of QED: the prediction for how relativistic electrons scatter off matter. This "Compton scattering" cross section is of great importance in astrophysics.

7.7   Compton Scattering

The phenomenon of photons scattering off atoms has been known long before quantum field theory. A classical model might be to imagine an electromagnetic plane wave incident on a charged particle. The oscillating electric field causes the particle to accelerate, emitting dipole radiation due to Maxwell's equations. From afar, this dipole radiation appears to be scattering of the original plane wave.

This classical explanation of *Thompson scattering* has two consequences of note. First, the scattered light should have the same frequency as the particle's oscillation and therefore of the incident plane wave. The light loses no energy by scattering. More quantitatively, the total scattering cross section off a particle equal in charge to the electron is

$$\sigma_T = \frac{2\alpha}{3m} \approx 6.65 \times 10^{-29}\,\mathrm{m}^2 \approx 0.66\,\mathrm{b}. \tag{7.21}$$

This cross section is known as the *Thompson cross section*. Here, $\alpha \approx 1/137$ is the fine structure constant and m is the mass of the charged particle.

However, Arthur Compton discovered in 1923 that both of these consequences are violated. In fact, the higher energy light downshifts in energy after scattering and the cross section decreases from σ_T. This detection makes perfect sense if you imagine light as a particle; a scattering photon changes momentum and therefore must transfer some of it to the charged particle. Since the charged particle gains kinetic energy due to

the additional momentum, the photon must lose energy as well. Further-more, the energy loss should be more detectable for photons with larger momentum—i.e., higher energy.

Compton scattering as it became known is one of the first definitive proofs that Max Planck was right, and light comes in quanta. But since light is also a relativistic object, a relativistic quantum theory is necessary to understand exactly how Compton scattering arises. In this section, we'll use QFT to derive the cross section of Compton scattering, known as the Klein-Nishina equation.

The first-order Feynman diagram for Compton scattering is very simple:
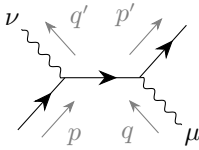
$$iM = -ie^2 \epsilon^\mu u \left(\gamma_\mu \frac{\not{p} - \not{q}' + m}{(p - q')^2 - m^2} \gamma_\nu \right) \epsilon^\nu \overline{v}.$$

But as usual, we can spin-average the cross section. This will eliminate any odd power of γ, as well as the initial and final polarization and spinor states. Thus, the amplitude simplifies to

$$M = -\frac{me^2}{(p - q')^2 - m^2}. \tag{7.22}$$

Working out the denominator requires the use of momentum conserva-tion and special relativity, which reveals that $(p - q')^2 = m^2 \left(1 - \frac{2\delta}{1 + \sin\theta} \right)$ where δ is the energy loss of the photon divided by the initial energy (so $\delta < 1$). Thus,

$$M = -\frac{e^2}{m} \frac{1 + \sin\theta}{1 + \sin\theta - 2\delta}. \tag{7.23}$$

JTD: There's a problem with the sign of $(p - q')^2$ and I think M should decrease with rising δ, not increase.

7.8   The Ward Identity & Massless Photons are Protected

Problems

7.1   **Repulsion Between Electrons**. Use the same techniques as section 7.5 to compute the Coulomb potential between two like-charged particles. Conclude that these particles repel.

Chapter 8

Renormalization

8.1 Magnetic Moment of the Electron
8.2 Photon self-energy?
8.3 Electronic Scattering
8.4 Dimensional Regularization
Problems

8.1 **Mass shift of a scalar boson**. In section 5.8 we began a calculation for the mass shift of a massless scalar boson but could not complete it because we lacked the tools. Use the techniques of section 8.2 to compute the mass shift of the χ particle under the theory used in section 5.8 to first loop order.

Chapter 9

Effects of Ultraviolet Divergences

Chapter 10

Infrared divergences

10.1 Bremsstrahlung Diagrams

10.2 Massive photon renormalization

Part III

Electroweak Theory & the Higgs Mechanism

Chapter 11

Mesons

Chapter 12

Yang-Mills Theory

12.1 Non-Abelian Symmetry Breaking

Chapter 13

Electroweak theory

13.1 Non-Abelian Symmetry Breaking
13.2 Feynman Rules for the Electroweak theory
13.3 Muon lifetime
13.4 Higgs Decay

Part IV

Strong Theory & The Standard Model

Chapter 14

Quantum Chromodynamics

Chapter 15

Standard Model

Part V

Review

Chapter 16

Review of Special Relativity

16.1 Tenets of Special Relativity

16.2 Time Dilation, Length Contraction, and the Lorentz Group

16.3 Particle Momenta in Special Relativity

Chapter 17

Review of Electromagnetism

17.1   Maxwell's Equations

17.2   Covariant Electromagnetism: the Faraday Tensor

17.3   Gauge invariance

Chapter 18

Review of Non-relativistic Quantum Mechanics

18.1 Tenets of Quantum Mechanics

18.2 Schrödinger and Heisenberg Picture

18.3 Non-Relativistic Scattering

Since the predominant use of relativistic QFT is to compute scattering cross sections, it is worth reviewing non-relativistic scattering in quantum mechanics. Consider a particle propelled toward some potential $V(\boldsymbol{x})$ which tails off at large $\boldsymbol{x}$. For example, an electron incident on a proton with Coulomb attraction. When the particle is far from the potential, its wavefunction is a superposition of plane waves. For simplicity, suppose the initial state is just one plane wave of momentum $\boldsymbol{p}$:

$$|\psi(t \to -\infty)\rangle = |\boldsymbol{p}\rangle = \int dx \, e^{i\boldsymbol{p}\cdot x} \, |x\rangle \, . \qquad (18.1)$$

The final state is also far from the potential, so it is a superposition of many plane waves

$$|\psi(t \to +\infty)\rangle = A \, |\boldsymbol{p}\rangle - i \int d^3 \boldsymbol{k} \, |\boldsymbol{k}\rangle \, M(\boldsymbol{k}). \qquad (18.2)$$

To understand what $M(\boldsymbol{p})$ is, use the fact that $\langle \boldsymbol{k}|\boldsymbol{p}\rangle = \delta(\boldsymbol{k}-\boldsymbol{p})$ so that

$$\langle \boldsymbol{k}|t \to +\infty\rangle = A\delta(\boldsymbol{k}-\boldsymbol{p}) - iM(\boldsymbol{k}). \tag{18.3}$$

Thus, $|A|^2$ is the probability for the particle not to be scattered at all, and $|M(\boldsymbol{k})|^2$ is the probability of scattering into momentum $\boldsymbol{k}$.

Our goal is to compute $M(\boldsymbol{k})$ using the Schrödinger equation, which states that (for a static potential)

$$|\psi(t)\rangle = e^{-2iHt}\,|\psi(-t)\rangle\,. \tag{18.4}$$

Since our initial and final states are simple in the momentum basis, we'll write H in the momentum basis too. The kinetic energy part of H is simple:

$$K = \int d^3\boldsymbol{k}\,|\boldsymbol{k}\rangle\,\frac{k^2}{2m}\,\langle \boldsymbol{k}| \tag{18.5}$$

and the potential energy can be written in the momentum basis via its Fourier transform $V(\boldsymbol{k})$ and the fact that $\langle p|x\rangle = e^{i\boldsymbol{k}\cdot x}$.

$$\begin{aligned}
V &= \int d^3\boldsymbol{x}\,|\boldsymbol{x}\rangle\,V(\boldsymbol{x})\,\langle \boldsymbol{x}| \\
&= \int d^3\boldsymbol{x}\,d^3\boldsymbol{k}\,d^3\boldsymbol{k}'\,|\boldsymbol{k}\rangle\,\langle \boldsymbol{k}|\boldsymbol{x}\rangle\,V(\boldsymbol{x})\,\langle \boldsymbol{x}|\boldsymbol{k}'\rangle\,\langle \boldsymbol{k}'| \\
&= \int d^3\boldsymbol{x}\,d^3\boldsymbol{k}\,d^3\boldsymbol{k}'\,|\boldsymbol{k}\rangle\,e^{-i(\boldsymbol{k}'-\boldsymbol{k})\cdot \boldsymbol{x}}V(\boldsymbol{x})\,\langle \boldsymbol{k}'| \\
&= \int d^3\boldsymbol{k}\,d^3\boldsymbol{k}'\,|\boldsymbol{k}\rangle\,V(\boldsymbol{k}'-\boldsymbol{k})\,\langle \boldsymbol{k}'|\,.
\end{aligned} \tag{18.6}$$

We'll consider the case of weak scattering, where V is usually small[1] In this case, the particle will spend only a short time in the potential well, so we can strategically pick the time t such that $|\psi(t)\rangle \approx |\psi(\infty)\rangle$ (and

---

[1] For example, the expectation value of the scattered particle's kinetic energy is much greater than its potential energy.

$|\psi(-t)\rangle \approx |\psi(-\infty)\rangle$), but t is still small. We will therefore drop all terms of order $\mathcal{O}(t^3)$, $\mathcal{O}(t^2 V)$, and $\mathcal{O}(tV^2)$ in our calculation for M:

$$\langle \boldsymbol{k}|\psi(t)\rangle = \langle \boldsymbol{k}| \left[1 - 2iKt - 2K^2t^2 - 2iVt\right] |\psi(-t)\rangle$$
$$= \left(1 - i\frac{k^2}{m} - \frac{k^4}{2m^2}\right) \langle \boldsymbol{k}|\boldsymbol{p}\rangle - 2it \int d^3k'\, V(\boldsymbol{k}' - \boldsymbol{k}) \langle \boldsymbol{k}'|\boldsymbol{p}\rangle. \tag{18.7}$$

Equation (18.3) connects the left hand side to the scattering amplitude M. Since $\langle \boldsymbol{k}|\boldsymbol{p}\rangle$ is the same $\delta(\boldsymbol{k} - \boldsymbol{p})$ as the coefficient of A, we identify A with the first term of the previous equation and iM with the second:

$$A = 1 - it\frac{p^2}{m} - t^2\frac{p^4}{2m^2}$$
$$M(\boldsymbol{k}) = 2tV(\boldsymbol{k} - \boldsymbol{p}). \tag{18.8}$$

This is an interesting result: $|A|^2 = 1$ to order t^2, so that an unscattered particle is offset in phase by $\Delta\phi = -tp^2/m$. **JTD: Somehow this phase is -1** . It follows that our scattering amplitude is

$$M(\boldsymbol{k}) = 2\frac{m}{p^2}V(\boldsymbol{k} - \boldsymbol{p}). \tag{18.9}$$

For an inelastic position (i.e., when the potential contains no internal degrees of freedom which could steal energy form the particle), we expect the initial and final energies to be the same. This implies that $M(\boldsymbol{k}) \propto \delta(k^2 - p^2)$ so that the scattering amplitude lies on a sphere of radius p^2.

JTD: Alternate method

This time, instead of framing the problem as a single particle moving from late times to early times, we'll imagine a stream of particles continuously scattering from a distant source. At far distances, the wavefunction is a superpostion of the incoming plane wave and the scattered wavefunction

$$\psi(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{x}} + \phi(\boldsymbol{r}), \tag{18.10}$$

and this wavefunction satisfies the Schrödinger equation

$$H\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \tag{18.11}$$

Splitting the Hamiltonian into kinetic and potential energy $H = K + V$,

$$(E - K)\psi(\boldsymbol{r}) = V\psi(\boldsymbol{r}). \tag{18.12}$$

Suppose we have some function $G(\boldsymbol{x})$, which satisfies

$$(E - K)G(\boldsymbol{x}) = \delta(\boldsymbol{x}). \tag{18.13}$$

This equation is useful because it suggests that the following fact is true of $\psi(\boldsymbol{x})$

$$\psi(\boldsymbol{r}) - \psi_0(\boldsymbol{r}) = \int d^3\boldsymbol{x}'\, G(\boldsymbol{x} - \boldsymbol{x}')V(\boldsymbol{x}')\psi(\boldsymbol{r}) \tag{18.14}$$

where $\psi_0(\boldsymbol{r})$ is the wavefunction of free space, where $V = 0$. **JTD: Explain the fact that if we use an advanced Green's function, we can use $\psi_0 = $ plane wave.**

The left hand side of (18.14) is therefore the scattered wavefunction $\phi(\boldsymbol{r})$. A simple yet powerful approximation is to set $\phi(\boldsymbol{r})$ in the integrand equal to $\psi_0(\boldsymbol{r})$. This is known as the *Born approximation,* and it readily gives the scattered wavefunction.

Chapter 19

Complex Analysis

Complex analysis is the branch of math that deals with calculus for complex functions. It's status as an upper level undergraduate class for majors—often not even taught to physicists—gives the subject a forboding field. But in fact it is much more intuitive than it's cousin real analysis (calculus for real functions) and is quite similar to the vector calculus most first-year physics majors learned. It's techniques, particularly the Residue theorem, are very important in QFT, so we provide a summary of them.

We restrict ourselves to a subclass of complex functions called *analytic* or sometimes *holomorphic* functions. These functions are continuous and infinitely differentiable functions of a single complex variable z. Uses of the complex conjugate of z are forbidden. Examples of *non*-analytic functions are

$$z^* z, \qquad |z|, \qquad \text{piecewise functions.}$$

We can express this requirement of analyticity in another way: if we write a complex number $z = a + bi$ as a vector $\boldsymbol{z} = (a, -b)^T$, then a complex function $f(z)$ being analytic is equivalent to the vector function $\boldsymbol{f}(\boldsymbol{z})$ being curl-free and divergence-free in the vector calculus sense:

$$\nabla \cdot \boldsymbol{f} = 0 \qquad \nabla \times \boldsymbol{f} = 0. \tag{19.1}$$

Holomorphic functions obey several very useful theorems concerning power series (section 19.1) and integration (section 19.2) discussed next.

19.1   Series, Poles, and Branch Cuts

The constraint for analyticity (19.1) seems strict, but most functions in physics nevertheless satisfy it. Polynomials, rational functions, and exponents are all analytic functions. In fact, any function with a Taylor series is analytic. Likewise, analytic functions can be written as a Taylor series which converges in some disk of radius R centered on z_0:

$$f(z) = \sum_{n=0}^{\infty} \frac{(z - z_0)^n}{n!} \left. \frac{d^n f}{dz^n} \right|_{z=z_0} \quad \text{converges for all } |z - z_0| < R. \quad (19.2)$$

This statement has an immediate consequence, which is that if you know the value of $f(z)$ in a small region—but one big enough to compute all the derivatives of f there—then (19.2) tells you the value of f everywhere in the much larger disk of convergence. In this sense, analytic functions are *unique*.

In some functions (called *entire* functions), that disk is the entire complex plane, but for most functions the disk can only grow so large before the series diverges. The maximum radius of the disk, called the *radius of convergence R*, is always the distance to the nearest singularity of the function.

A *singularity* is a location where the function is undefined. The most common singularity is a *pole*, where f blows up due to dividing by a small number. For example, the function $f(z) = 1/z$ has a pole at $z = 0$. $1/z^2$ also has a pole at zero, but we would call this a pole of multiplicity two due to the square in the denominator.

The relationship between series and poles is very useful, and sheds light on why some Taylor series of real functions diverge. For example,

the function

$$f(x) = \frac{1}{x^2 + 1} \tag{19.3}$$

is defined for all real x, and has a seemingly innocent Taylor series

$$\frac{1}{x^2 + 1} = \sum_{n=0}^{\infty} (-1)^n x^{2n}. \tag{19.4}$$

But try plugging in $x > 1$ or $x < -1$ and the Taylor series diverges. That is, it has a radius of convergence 1. We can see why by returning to the original function (19.3) and promoting x to a complex variable z. Then $f(z)$ has poles at $z = \pm i$, since $1/(i^2 + 1)$ diverges. The distance between the center of the Taylor series (0) and the pole (i) is one, so $R = 1$.

Luckily, a series diverging is often fixable. Consider $f(z) = \ln z$, defined as the function such that $e^{f(z)} = z$. This function has a pole at $z = 0$, since $e^{-\infty} = 0$. For this reason, the Taylor series around $z_0 = 1$ has a radius of convergence of $R = 1$, which you can verify for yourself by looking at the series' terms

$$\ln(z) = (z - 1) - \frac{(z - 1)^2}{2} + \frac{(z - 1)^3}{3} - \frac{(z - 1)^4}{4} + \cdots - (-1)^n \frac{(z - 1)^n}{n}. \tag{19.5}$$

However, we can extend the domain of this function by evaluating the series at a new point, such as $z_1 = (1 + i)/\sqrt{2}$. Since z_1 is well within the radius of convergence of (19.5), we can also find all the derivatives of $\ln(z_1)$ and write out the Taylor series:

$$\ln(z) = \sum_{n=0}^{\infty} a_n (z - z_1)^n \tag{19.6}$$

where a_n are new complex constants. Since the only pole of the logarithm is at $z = 0$, this series still has a radius of convergence of 1. Since $|i - z_1| < 1$, we can use it to define the logarithm of i, which turns out to be $\ln i = i\pi/2$.

The point is that while i is within the radius of convergence of (19.6), it is outside the radius of convergence of the original series because $|i - 1| > 1$. So by introducing the intermediate series centered on z_1, we have managed to beat the radius of convergence and define $\ln i$! If we wished, we could go farther and define $\ln z$ at any point in the complex plane (except $z = 0$) despite the pole, simply by introducing z_2, z_3, etc. and walking the radius of convergence of the series over to z.[1] This process is called *analytic continuation* and can be performed on any analytic function. A diagram is given in Figure 19.1.

Qualitatively speaking, we've shown that the divergence of (19.5) was a red herring except at $z = 0$. The function was perfectly well defined.

Or was it? There is still one property left to check: does the value we got for $\ln i$ depend on the value we chose for z_1? At first it seems like the answer is no because the disk of radius 1 surrounding z_1 directly connects i to 1, and the logarithm must be unique in this disk as we discussed just after (19.2).

But there's a flaw in this argument. What if we didn't connect 1 to i directly? We could instead define $z_1 = (1 - i)/\sqrt{2}$, $z_2 = -i$ and so forth, making a series of disks wrapping clockwise around the origin instead of counterclockwise. Try this, and we would get $\ln i = -3\pi i/2$ instead.[2]

But it cannot be that $\ln i = \pi i/2$ and $-3\pi i/2$ at the same time. This problem illustrates the weakness of analytic continuation. It allows a smooth function to be defined anywhere except at a singularity, but it may induce inconsistencies like constant offsets in the value of the func-

---

[1] Note that we do not have a single Taylor series which works for every point; you must chose the right Taylor series which is centered on a nearby point.

[2] In the case of the logarithm, it is simple to find the 2π offset which analytic continuation produces. It happens that $\ln(re^{i\theta})$, where r and θ are real, is $\ln r + i\theta$. Check this by exponentiating $\ln r + i\theta$ and confirming it is equal to $re^{i\theta}$. Then defining z_i along a circle of radius $r = 1$ keeps increasing the imaginary component of z_i as θ is increased until the circle is complete and the analytic continuation has accumulated a $2\pi i$ offset.
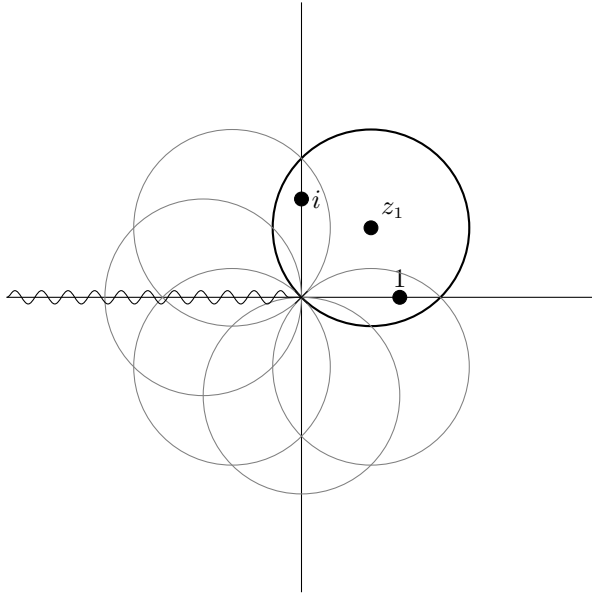
Figure 19.1: Black: The analytic continuation that gives $\ln i = \pi i/2$ via a Taylor series centered on z_1. Gray: the contradicting path giving $\ln i = -3\pi i/2$, and the wavy line is the branch cut preventing it. The cross in the center is the pole.

tion. We fix these contradictions by defining a *branch cut*, which is an arbitrary path extending from the singularity to infinity at which we say $\ln z$ is undefined. That way, the circular path used above to create an inconsistency is interrupted by the branch cut and the argument can no longer be made. This branch cut is represented with a wavy line in Figure 19.1 and chosen to lie along the negative real axis by convention. In general, any singularity in a function generated by analytic continuation should have a branch cut attached to it. For example, $\sqrt{z}$ also has a branch cut extending from $z = 0$, conventionally chosen to lie along the negative

real axis as well.

Awareness of the poles and branch cuts in a complex function is a crucial trick when it comes to Taylor expanding or integrating a function. The connection between integration and poles is worked out in the next chapter.

19.2   Residue Theorem

This section is on the integrals of complex functions such as $f(z)$. Like in vector calculus, when we integrate a function $\int_\gamma dz\, f(z)$ we must specify a curve γ to integrate over and we define the integral in the usual way using Riemann sums. The parametrization of the curve is unimportant, but the location of the curve in complex space is. In vector calculus a two-dimensional vector field $\boldsymbol{f}(\boldsymbol{z})$ obeys Green's theorem for integration around a closed contour γ:

$$\oint_\gamma d\boldsymbol{\ell} \cdot \boldsymbol{f}(\boldsymbol{z}) = \int_D d^2\boldsymbol{z}\, \nabla \times \boldsymbol{f}(\boldsymbol{z}) \tag{19.7}$$

where D is the interior of γ. If $\nabla \times \boldsymbol{f}(\boldsymbol{z}) = 0$ everywhere in D, then the line integral is zero. We would say that $\boldsymbol{f}(\boldsymbol{z})$ is *conservative*.

But remember, analytical functions are curl free. Thus,

$$\oint_\gamma dz\, f(z) = 0 \tag{19.8}$$

for functions f that are analytical everywhere inside γ. An equivalent statement that works for open curves is that if one is integrating $f(z)$ over a curve γ from a to b in the complex plane, γ can be moved without changing the value of the integral as long as f is analytic in the region γ was moved through.[3] But what if f had a few poles at in that region?

---

[3]This statement about an unclosed curve γ is equivalent to (19.8) for the following reason. Consider your starting curve γ_1 from a to b and the curve γ_2 that you're moving to. Reverse γ_2 to make a new curve $-\gamma_2$ from b to a. Merge these two curves and you get a new curve $\gamma_\circ = \gamma_1 - \gamma_2$ which is closed. If f is analytic between the curves then (19.8) guarantees that $\int_{\gamma_1} dz\, f(z) + \int_{-\gamma_2} dz\, f(z) = 0$, or $\int_{\gamma_1} dz\, f(z) = \int_{\gamma_2} dz\, f(z)$.

Then $f(z_j)$ wouldn't be analytic, we couldn't use Green's theorem, and the integral might change if we moved γ across the pole. Let us focus on these poles specifically now.

Suppose we have a function $f(z) = z^{-n}$, with a pole of multiplicity n at the origin. We can find the value of a closed integral around the origin by integrating in a nearly-closed contour in a circle around the pole and taking a limit as the ends approach each other. Note that the radius of this circle doesn't matter because if we shrink or expand the circle to another radius, we do not pass the curve through any poles, so that the value of the integral doesn't change.

Complex functions obey the fundamental theorem of calculus:

$$\int_a^b dz\, f(z) = F(b) - F(a) \qquad \text{where} \qquad \frac{dF}{dz} = f(z). \qquad (19.9)$$

A function $F(z)$ is generally called an *antiderivative* of f. We changed notation here by not specifying γ and instead only specifying its endpoints. This is allowed because, as we just learned, γ may be moved through space where f is defined so its exact path doesn't matter.

An antiderivative of z^{-n} is $F(z) = z^{1-n}/(1-n)$, just like for real z, except when $n = 1$ in which case the antiderivative is $F(z) = \ln z$. For $n > 1$, $F(z)$ is continuous so that $F(b) - F(a) \to 0$ as $b \to a$. It follows that the closed integral around z^{-n} is zero for $n > 1$. However, for $n = 1$ we have to deal with the branch cut in the logarithm, over which $F(z)$ is *dis*continuous. We've already mentioned that $F(z)$ increases by $2\pi i$ when moving counterclockwise over the branch cut, so that the value of $\oint dz\, z^{-1}$ must likewise be $2\pi i$. To summarize,

$$\oint dz\, z^{-n} = \begin{cases} 0 & n \neq 1 \\ 2\pi i & n = 1 \end{cases}. \qquad (19.10)$$

Returning to integrating an arbitrary function $f(z)$ with poles at z_j, by definition $f(z)$ can be expanded around a pole of multiplicity m as

follows:

$$f(z) = \frac{a_{-m}}{(z - z_j)^m} + \frac{a_{1-m}}{(z - z_j)^{m-1}} + \cdots = \sum_{n=-m}^{\infty} a_n (z - z_j)^n. \quad (19.11)$$

The integral of this function is then zero for each term except $n = -1$, where it is $2\pi i a_{-1}$. We generally write the coefficient a_{-1} of this Taylor series centered at z_j as $a_{-1} = \text{Res } f(z_j)$, or the *residue* of f at z_j. Thus

$$\oint_\gamma dz f(z) = 2\pi i \sum_{z_j} \text{Res } f(z_j) \quad (19.12)$$

where z_j are the poles inside γ.

(19.12) is an incredibly powerful theorem. It can be used to integrate even very complicated functions over closed contours. It can even be used to perform integrals over non-closed contour, by closing the contour through some region of space where we know $f(z)$ is negligible. For example, we can easily find that

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \pi \quad (19.13)$$

via the Residue theorem. First we split the integrand into partial fractions to find the residue:

$$\frac{1}{2} \int_{-\infty}^{\infty} dx \left(\frac{i}{x + i} - \frac{i}{x - i} \right). \quad (19.14)$$

The function has poles at $z = \pm i$, with $\text{Res } f(i) = -i$ and $\text{Res } f(-i) = i$. To make a closed curve, we can close the contour in an enormous loop with very large $|z|$ from the $+\infty$ of the real axis to the $-\infty$ side. This can be done over the real line, in which case the curve will run counterclockwise and enclose the $+i$ pole, or it can be done under the real axis in which case it will run clockwise and enclose the $-i$ pole. Both these contours should give the same integral, and they do. The two poles have

opposite residues, but integrating over a clockwise curve flips the sign of the integral so that the result for both is

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = 2\pi i \left(\frac{1}{2}(-i) \right) = \pi. \tag{19.15}$$

19.3   The Gamma Function

We conclude our function on special analysis with the Gamma function, $\Gamma(z)$. $\Gamma(z)$ is one of many special functions, much like the Riemann zeta function $\zeta(z)$, which are discussed in conjunction with complex analysis because their properties are so much easier to understand in the context of complex variables. Nevertheless, when $\Gamma(z)$ appears in physics, a real argument is usually used.

The Gamma function is a crucial function that appears in QFT due to its relation to the surface area of a sphere, discussed in the next section. It is also relevant to statistics and can be helpful in computing miscellaneous integrals as well. It is a generalization of the factorial:

$$\Gamma(n + 1) = n! \tag{19.16}$$

for integer n. A consistent definition that works for all Re $z > 0$ is

$$\Gamma(z) = \int_0^{\infty} dt \, t^{1-z} e^{-t}. \tag{19.17}$$

One can check that this definition is consistent with the factorial by showing that

$$\Gamma(x + 1) = x\Gamma(x) \tag{19.18}$$

and $\Gamma(0) = 1$, which are also properties of factorials. Many other useful properties are true of the Gamma function, but we leave these to Wikipedia.

As defined in (19.17), $\Gamma(z)$ is analytic and smoothly connects all the values of the factorial on the positive real axis. On the negative real axis,
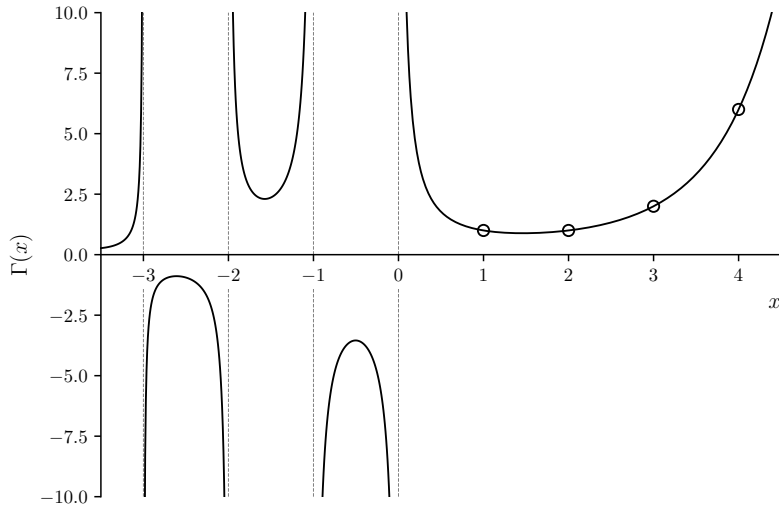
Figure 19.2: The Gamma function, $\Gamma(x)$ on the real line. Open circles represent the value of $(x+1)!$, and dotted vertical lines are the poles at every non-positive integer.

it has poles at every negative integer and at zero (Figure 19.2). At the zero pole, $\Gamma(z)$ has a convenient expansion:

$$\Gamma(\epsilon) = \frac{1}{\epsilon} - \gamma_E + \mathcal{O}(\epsilon) \tag{19.19}$$

which generalizes to

$$\Gamma(-n+\epsilon) = \frac{(-1)^n}{n!}\left(\frac{1}{\epsilon} - \gamma_E + \sum_{j=1}^{n}\frac{1}{j}\right) + \mathcal{O}(\epsilon) \tag{19.20}$$

where $\gamma_E \approx 0.57721$ is the Euler-Mascheroni constant.

Another useful formula is the values of the Gamma function at half-integers:

$$\Gamma\left(z + \frac{1}{2}\right) = \frac{\sqrt{\pi}}{4^{z-\frac{1}{2}}}\frac{\Gamma(2z)}{\Gamma(z)} \tag{19.21}$$

which reduces to the special case of $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. The Gamma function can also be used to aid in evaluating the following integral

$$B(a, b) = \int_0^1 dt\, t^{a-1}(1 - t)^{b-1} \tag{19.22}$$

which is called the *Beta function*. It appears in QFT as well as in probability and various trigonometric functions. It is related to the Gamma function via

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}. \tag{19.23}$$

Below, we use the Beta and Gamma functions to compute the surface area of a d-dimensional sphere.

Surface Area of Spheres

In physics, integrals over quantities which are rotationally symmetrical are very common. Such integrals can be separated into a radial part and an angular part, even in an arbitrary d-dimensional space, by integrating over d-dimensional spheres like so

$$\int d^d x\, f(|x|) = \left(\int dr\, r^{d-1} f(r) \right) A_d \tag{19.24}$$

where A_d is the surface area of the d-dimensional unit sphere. For example, $A_3 = 4\pi$ and $A_2 = 2\pi$.

The task of this section is to find the surface area of this d-dimensional sphere as a function of d, which we will do via recursion. Suppose we know the area A_{d-1}. This knowledge allows us to separate the d dimensions into one $\hat{z}$ axis and $d - 1$ other axes, which together form a hyperplane P. Now we can think two-dimensionally in terms of the plane P and $\hat{z}$. (Figure 19.3)

The d-sphere corresponds to many $d - 1$-spheres stacked along the $\hat{z}$ axis, just as a normal three dimensional sphere is a stack of circles. As the
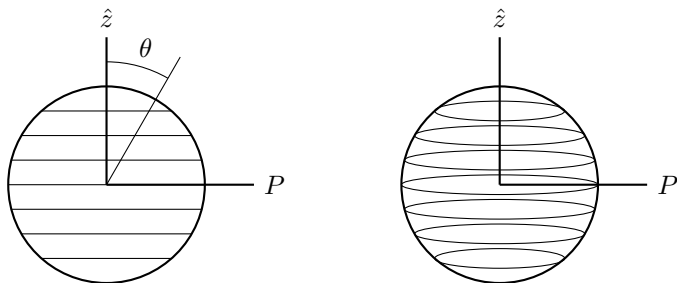
Figure 19.3: Stacking $d-1$-dimensional spheres to form a d-dimensional sphere for $d = 3$.

stack grows, the circles (or $d-1$-spheres) shrink, following $r = \sin\theta$ as θ rises from $\pi/2$ on the equator to zero at the pole. The general scaling law of area with radius implies that the area also shrinks as r^{d-2}. To compute the area in d-space dA, we must also multiply by the thickness of each slice so that $dA = A_{d-1}\sin^{d-2}\theta d\theta$ to the total area per slice. Integrating this equation,

$$
\begin{aligned}
A_d &= A_{d-1} \int_0^\pi d\theta \, \sin^{d-2}\theta \\
&= -A_{d-1} \int_0^\pi d(\cos\theta) \, \sin^{d-3}\theta \\
&= A_{d-1} \int_{-1}^1 dx \, (1-x^2)^{\frac{d-3}{2}} \\
&= A_{d-1} \int_0^1 dt \, t^{-\frac{1}{2}} (1-t)^{\frac{d-3}{2}}
\end{aligned}
\tag{19.25}
$$

where in the second line we substituted variables from $\theta \to \cos\theta$, in the third line we expressed $\sin\theta = \sqrt{1-\cos^2\theta}$, and in the last line we substituted variables $x^2 \to t$. The last integral is the Beta function $B(\frac{1}{2}, \frac{d-1}{2})$ we defined in the previous section, so we may immediately write it in terms

of Γ functions:

$$A_d = A_{d-1}\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{d-1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} = A_{d-1}\sqrt{\pi}\frac{\Gamma\left(\frac{d-1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}. \tag{19.26}$$

Now we must turn the above recursive equation into a formula for A_d, but fortunately this is not difficult due to the Γ function's own recursion relation (19.18). The following formula checks out:

$$A_d = \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)}. \tag{19.27}$$

(19.27) draws a fascinating and unexpected connection between the Gamma function, which descended from factorials, and spheres. While it is possible to evaluate the first line (19.25) without referring to the Gamma function if d is an integer, our result is actually applicable for all d because the Beta function is valid for any complex input. The validity of (19.27) for non-integers seems irrelevant, but it is actually critical in QFT and will be used in chapter 8.

Chapter 20

Review of Group Theory and Lie-Groups

Properties of the γ^μ atrices

Grassmann numbers

Above, we stated the fact that spinor operators anticommute. But in the QFT principle of least action, only the left hand side (the nPTFs) contain spinors. The right hand side is a path integral over the complex values of the field, which in this case is just a vector of four numbers ϕ_i to emulate the four operators in the $\hat{\phi}$ spinor.

However, the anti-commuting nature of the fermion must be taken into account. To do so, we define a new number line of anticommuting numbers, which are called Grassmann numbers. These will represent the possible fermionic field values in the path integral of the QFT principle of least action.

The definition of these numbers is that, for any two Grassmann numbers η and θ,

$$\theta\eta = -\eta\theta. \tag{1}$$

A special case is that $\theta^2 = \eta^2 = 0$. These numbers may seem like odd, but future investigation will show that their properties are very simple do to their square vanishing in this way. For example, the Taylor series of any function $f(\theta)$ is $\alpha + \beta\theta + \gamma\theta^2 + \dots$ for more Grassmann numbers constants α, β, γ, etc. But since $\theta^2 = 0$, this is just

$$f(\theta) = \alpha + \beta\theta. \tag{2}$$

All functions are therefore linear.

Integration is likewise simple for Grassmann numbers. We will only need integrals over the entire space of Grassmann numbers — the equivalent of integrals from $-\infty$ to ∞ on the real line — and therefore the integral of a function $\alpha + \beta\theta$ should depend only on α and β. Furthermore, substitution of variables implies that $\int d\theta\, f(\theta) = \int d\theta\, f(\theta + \eta)$, so that

$$\int d\theta\, \alpha + \beta\theta = \int d\theta\, \alpha + \beta(\theta + \eta) = \int d\theta\, (\alpha + \beta\eta) + \beta\theta. \qquad (3)$$

The intercept α of the function has been shifted by this substitution of variables, so the integrand can only be proportional to β. By convention, we set the integrand equal to β.

$$\int d\theta\, \alpha + \beta\theta = \beta. \qquad (4)$$

For multidimensional integrals, we treat the volume element $d\theta$ as another Grassmann number, so that $d\theta\, d\eta = -d\eta\, d\theta$. By convention, we set

$$\int d\eta\, d\theta\, \theta\eta = 1 \implies \int d\theta\, d\eta\, \theta\eta = -1. \qquad (5)$$

The last detail we need is that of complex Grassmann numbers, which are defined just like normal complex numbers: $\psi = \theta + i\eta$. Their complex conjugate is $\psi^* = \theta - i\eta$.

Wick's theorem for Grassmann numbers

JTD: Do this

Connecting Spin and Spinors

Spin 2 Particles and Gravitons

Dark Matter

Particles of the Standard Model

Fundamental Particles

JTD: Quark, lepton, gauge boson, Higgs diagram

Mesons

JTD: Eightfold way diagram

Baryons

JTD: Eightfold way but baryons

Other

JTD: Tetraquark, pentaquark, WIMP or Axions

Encyclopedia of Quantum Field Theories

Particles

List propagators and conserved quantities for

1. Spin zero: scalar boson

2. Spin one half: fermion

3. Spin one: vector boson

Gauge Groups

List D_μ and $F^{\mu\nu}$ for

1. U(1)

2. SU(2)

3. SU(3)

Interactions

1. ϕ^4

2. Yukawa coupling

3. QED

4. Electroweak

5. Standard Model

6. Axions

7. SU(5)

8. SO(10)

Important Formulas

Index