

# QFT Lectures Notes

**Jared Kaplan**

*Department of Physics and Astronomy, Johns Hopkins University*

## **Abstract**

Various QFT lecture notes to be supplemented by other course materials.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	What are we studying? . . . . .	2
1.2	Useful References and Texts . . . . .	3
<b>2</b>	<b>Fall Semester</b>	<b>4</b>
2.1	Philosophy . . . . .	4
2.2	Review of Lagrangians and Harmonic Oscillators . . . . .	8
2.3	Balls and Springs . . . . .	11
2.4	Special Relativity and Anti-particles . . . . .	17
2.5	Canonical Quantization and Noether's Theorem . . . . .	24
2.6	Free Scalar Quantum Field Theory with Special Relativity . . . . .	27
2.7	Dimensional Analysis, or Which Interactions Are Important? . . . . .	33
2.8	Interactions in Classical Field Theory with a View Towards QFT . . . . .	36
2.9	Overview of Scattering and Perturbation Theory . . . . .	39
2.10	Relating the S-Matrix to Cross Sections and Decay Rates . . . . .	41
2.11	Old Fashioned Perturbation Theory . . . . .	46
2.12	LSZ Reduction Formula – S-Matrix from Correlation Functions . . . . .	51
2.13	Feynman Rules as a Generalization of Classical Physics . . . . .	56
2.14	The Hamiltonian Formalism for Perturbation Theory . . . . .	59
2.15	Feynman Rules from the Hamiltonian Formalism . . . . .	64
2.16	Particles with Spin 1 . . . . .	72
2.17	Covariant Derivatives and Scalar QED . . . . .	78
2.18	Scattering and Ward Identities in Scalar QED . . . . .	83
2.19	Spinors and QED . . . . .	87
2.20	Quantization and Feynman Rules for QED . . . . .	95
2.21	Overview of 'Renormalization' . . . . .	100
2.22	Casimir Effect . . . . .	102
2.23	The Exact 2-pt Correlator – Kallen-Lehmann Representation . . . . .	106
2.24	Basic Loop Effects – the 2-pt Correlator . . . . .	109
2.25	General Loop Effects, Renormalization, and Interpretation . . . . .	116
2.26	Large Logarithms and Renormalization Flows . . . . .	121
2.27	QM Example of Wilsonian Renormalization . . . . .	128
2.28	Wilsonian Renormalization Flows . . . . .	130
2.29	Path Integrals . . . . .	134
2.30	Generating Function(al)s and Feynman Rules . . . . .	140
2.31	Path Integrals and Statistical Physics . . . . .	146
2.32	Soft Limits and Weinberg's Theorem . . . . .	148

<b>3</b>	<b>Spring Semester</b>	<b>151</b>
3.1	Effective Actions, Renormalization, and Symmetry . . . . .	151
3.2	Schwinger-Dyson from the PI and Ward Identities . . . . .	153
3.3	Discrete Symmetries and Spinors . . . . .	157
3.4	More on Spin and Statistics . . . . .	164
3.5	QED Vacuum Polarization and Anomalous Magnetic Moment . . . . .	169
3.6	Renormalization and QED . . . . .	176
3.7	IR Divergences and Long Wavelength Physics . . . . .	182
3.8	Implications of Unitarity . . . . .	191
3.9	Interlude on Lie Groups and Lie Algebras . . . . .	197
3.10	Overview of Lie Algebra Classification and Representations . . . . .	201
3.11	Spontaneously Broken Global Symmetries and Goldstone Bosons . . . . .	207
3.12	Non-Abelian Gauge Theories . . . . .	216
3.13	Quantization of Yang-Mills Theories . . . . .	222
3.14	Renormalization in YM and Asymptotic Freedom . . . . .	225
3.15	Higgs Mechanism . . . . .	227
3.16	Lattice Gauge Theory, QCD, and Confinement . . . . .	234
3.17	Parton Model and Deep Inelastic Scattering . . . . .	238
3.18	DIS, CFTs, and the OPE . . . . .	245
3.19	Anomalies – A Summary . . . . .	249
3.20	Anomalies as Almost Local Effects . . . . .	251
3.21	A Lecture on Cosmological Perturbation Theory . . . . .	254

# 1 Introduction

## 1.1 What are we studying?

Almost everything.

- The Standard Model of particle physics, which accounts for all observed phenomena on length scales larger than  $\sim 10^{-18}$  meters, is a quantum field theory. In traditional QFT courses one mostly learns how to compute the probabilities for various scattering processes in the standard model.
- General Relativity can also be viewed as a quantum field theory, or an ‘effective quantum field theory’ describing the long-distance behavior of a massless spin 2 particle, the graviton, as it interacts with the particles from the rest of the standard model. (The division between GR and the SM is linguistic/historical, and not very sensible conceptually.)
- The perturbations in the energy density in the early universe, which seeded structure formation, seem to have arisen from quantum fluctuations of a quantum field that we refer to as ‘the inflaton’.

- The behavior of systems near phase transitions, especially the universality of such systems. For example, systems of spins behave in the same way as water near its critical point – both are well-described by a simple quantum field theory, which we will study.
- The universal long-distance theory of metals, as Fermi liquids of weakly interacting electrons. This accounts for the universal properties of metals as shiny materials that are good at conducting electricity and heat, and that superconduct at low temperatures.
- Fluids and superfluids. Strangely enough, the latter are easier to understand than the former.
- Sound waves in materials, such as metals and crystals, as we will discuss.
- The quantum hall effect can be described using a very simple ‘topological’ quantum field theory, which is like electrodynamics, but much simpler.

What aren’t we studying? Basically, all that’s left are some systems from condensed matter physics that have a finite size, or a detailed lattice structure that’s crucially important for the physics. Or systems full of dirt (e.g. systems exhibiting localization). Systems with very little symmetry usually aren’t well described by QFT.

## 1.2 Useful References and Texts

Here’s a list of useful references with a brief summary of what they are good for.

- **Official Course Textbook: Quantum Field Theory and the Standard Model**, by **Matt Schwartz**
- Weinberg’s *Quantum Theory of Fields* – A profound, instant classic, which you should eventually absorb as completely as possible. It’s not used as a primary textbook for a first course in QFT because of its sophistication and its various idiosyncrasies.
- Zee’s *QFT in a Nutshell* – This book provides a wealth of the conceptual information about QFT and its applications; it can’t be used as a text because it’s not very systematic, and it doesn’t teach you how to calculate anything.
- Howard Georgi’s *Lie Algebras and Particle Physics* – One major stumbling block for students learning QFT is that an understanding of Lie Groups and their Representation Theory is absolutely essential, and yet this subject is almost never taught. This book is the most useful way to learn Lie Theory... I’ve read mathematics books on the subject, but they are far less useful because they rarely compute anything.
- Banks *Modern Quantum Field Theory* – This is an interesting alternative text, which basically forces readers to teach themselves QFT by working through a number of exercises. If you’re (re-)learning QFT on your own, or want a different, more self-directed perspective, this would probably be great... doing all the exercises would insure that you really understand the subject.

- Srednicki's *Quantum Field Theory* – This seems to be a well-liked standard text based on the path integral. If you want to see everything developed from that perspective, this is probably the reference for you.
- Peskin & Schroeder's *Introduction to QFT* – This has been a standard text for teaching for a long time, because it immediately involves students in relevant particle physics calculations... it's a great reference for computing Feynman diagrams and for some Standard Model subjects.
- Weinberg's *Quantum Mechanics* – this is a good reference for background on Lagrangian and Hamiltonian mechanics, as applied to Quantum Mechanics and Canonical Quantization (see chapter 9). It also has a lot of background on scattering.

Here are a few more advanced texts that are a natural place to go after this course, or for independent reading this year:

- Sidney Coleman's *Aspects of Symmetry* – This is a classic text with many important and fascinating advanced topics in QFT.
- Polyakov's *Gauge Fields and Strings* – This is basically Polyakov's notebook... deep and fascinating, touching on condensed matter and particle physics.
- Wess and Bagger *Supersymmetry* – Classic, bare-bones text on supersymmetry.
- Slava Rychkov's CFT Lectures notes and my lectures notes on AdS/CFT from fall 2013.
- Donoghue, Golowich, & Holstein's *Dynamics of the Standard Model* – Nice book about all sorts of standard model phenomenology and processes, especially the description of low-energy QCD phenomena.
- Whatever you find interesting and engaging from Inspire searches, hearsay, etc.

## 2 Fall Semester

### 2.1 Philosophy

Why is Quantum Field Theory the way it is? Does it follow inevitably from a small set of more fundamental principles? These questions have been answered by two very different, profoundly compatible philosophies, which I will refer to as *Wilsonian* and *Weinbergian*. Together with the *Historical* approach to QFT, we have three philosophies that motivate the introduction and quantization of spacetime fields. Now a bit about each philosophy.

The Historical philosophy arose from the discovery of quantum mechanics, and the subsequent realization that theories should be 'quantized'. This means that for any given classical theory of physics, we should look for a quantum mechanical theory that reduces to the classical theory in a classical limit. Since Maxwell's theory of the electromagnetic field was known long before the

discovery of QM, physicists in the early twentieth century had the immediate task of ‘quantizing’ it. There is a general procedure for quantizing a classical theory, called ‘Canonical Quantization’ (Canonical just means standard; you can read about these ideas in chapters 2 and 6 of Shankar’s Quantum Mechanics textbook). So one can motivate QFT by attempting to quantize classical field theories, such as the theory of the electromagnetic field. We will start with a simpler theory in this course, but the motivation remains.

The Wilsonian philosophy is based on the idea of zooming out. Two different physical systems that look quite different at short distances may behave the same way at long distances, because most of the short distance details become irrelevant. In particular, we can think of our theories as an expansion in  $\ell_{short}/L$ , where  $\ell_{short}$  is some microscopic distance scale and  $L$  is the length scale relevant to our experiment. We study the space of renormalizable quantum field theories because this is roughly equivalent to the space of universality classes of physical systems that one can obtain by ‘zooming out’ to long distances. Here are some famous examples:

- The Ising Model is a model of spins on a lattice with nearest-neighbor interactions. We can zoom out by ‘integrating out’ half of the spins on the lattice, leaving a new effective theory for the remainder. However, at long distances the model is described by the QFT with action

$$S = \int d^d x \left( \frac{1}{2} (\partial\phi)^2 - \lambda\phi^4 \right) \quad (2.1.1)$$

The details of the lattice structure become ‘irrelevant’ at long distances.

- Metals are composed of some lattice of various nuclei along with relatively free-floating electrons, but they have a universal phase given by a Fermi liquid of their electrons. Note that the Fermi temperature, which sets the lattice spacing for the atoms, is around 10,000 K whereas we are most interested in metals at  $\sim 300\text{K}$  and below. At these energies metals are very well described by the effective QFT for the Fermi liquid theory. See [1] for a beautiful discussion of this theory and the Wilsonian philosophy. Research continues to understand the effective QFT that describes so-called strange metals associated with high temperature superconductivity.
- Quantum Hall fluids seem to be describable in terms of a single Chern-Simons gauge field; one can show that this is basically an inevitable consequence of the symmetries of theory (including broken parity), the presence of a conserved current, and the absence of massless particles.
- The Standard Model and Gravity. There are enormous hierarchies in nature, in particular from the Planck scale to the weak scale.
- Within the Standard Model, we also have more limited (and often more useful!) effective descriptions of QED, beta decay, the pions and nucleons, and heavy quarks. Actually, general relativity plus ‘matter’ is another example of an effective description, where the details of the massive matter are unimportant at macroscopic distances (e.g. when we study the motion of the planets, it’s irrelevant what they are made of).

So if there is a large hierarchy between short and long distances, then the long-distance physics will often be described by a relatively simple and universal QFT.

Some consequences of this viewpoint include:

- There may be a true physical cutoff on short distances (large energies and momenta), and it should be taken seriously. The UV theory may not be a QFT. Effective Field Theories with a finite cutoff make sense and may or may not have a short-distance = UV completion.
- UV and IR physics may be extremely different, and in particular a vast number of distinct UV theories may look the same in the IR (for example all metals are described by the same theory at long distances). This means that a knowledge of long-distance physics does not tell us all that much about short-distance physics – TeV scale physics may tell us very little about the universe’s fundamental constituents.
- Symmetries can have crucially important and useful consequences, depending on whether they are preserved, broken, emergent, or anomalous. The spacetime symmetry structure is essential when determining what the theory can be – high-energy physics is largely distinguished from condensed matter because of Poincaré symmetry.
- QFT is a good approach for describing both particle physics and statistical physics systems, because in both cases we are interested in (relatively) long-distance or macroscopic properties.

For a classic review of the Wilsonian picture of QFT see Polchinski [1].

The Weinbergian philosophy [3] finds Quantum Field Theory to be the only way to obtain a Lorentz Invariant, Quantum Mechanical (Unitary), and Local (satisfying Cluster Decomposition) theory for the scattering of particles. Formally, a “theory for scattering” is encapsulated by the S-Matrix

$$S_{\alpha\beta} = \langle \alpha_{in} | S | \beta_{out} \rangle \quad (2.1.2)$$

which gives the amplitude for any “in-state” of asymptotically well-separated particles in the distant past to evolve into any “out-state” of similarly well-separated particles in the future. Some aspects of this viewpoint:

- Particles, the atomic states of the theory, are defined as irreducible representations of the Poincaré group. By definition, an electron is still an electron even if it’s moving, or if I rotate around it! This sets up the Hilbert space of incoming and outgoing multi-particle states. The fact that energies and momenta of distant particles *add* suggests that we can use harmonic oscillators  $a_p$  to describe each momentum  $p$ , because the harmonic oscillator has evenly spaced energy levels.
- The introduction of creation and annihilation operators for particles is further motivated by the Cluster Decomposition Principle<sup>1</sup>. This principle says that very distant processes don’t

---

<sup>1</sup>In AdS/CFT we can *prove* that this principle holds in AdS directly from CFT axioms.

affect each other; it is the weakest form of locality, and seems necessary to talk sensibly about well-separated particles. Cluster decomposition will be satisfied if and only if the Hamiltonian can be written as

$$H = \sum_{m,n} \int d^3 q_i d^3 k_i \delta \left( \sum_i q_i \right) h_{mn}(q_i, k_i) a^\dagger(q_1) \cdots a^\dagger(q_m) a(k_1) \cdots a(k_n) \quad (2.1.3)$$

where the function  $h_{mn}$  must be a non-singular function of the momenta.

- We want to obtain a Poincaré covariant S-Matrix. The  $S$  operator defining the S-Matrix can be written as

$$S = T \exp \left( -i \int_{-\infty}^{\infty} dt V(t) \right) \quad (2.1.4)$$

Note that this involves some choice of  $t$ , which isn't very covariant-looking. However, if the interaction  $V(t)$  is constructed from a local Hamiltonian density  $\mathcal{H}(x)$  as

$$V(t) = \int d^3 \vec{x} \mathcal{H}(t, \vec{x}) \quad (2.1.5)$$

where the Lorentz-scalar  $\mathcal{H}(x)$  satisfies a causality condition

$$[\mathcal{H}(x), \mathcal{H}(y)] = 0 \quad \text{for} \quad (x - y)^2 \text{ spacelike} \quad (2.1.6)$$

then we will obtain a Lorentz-invariant S-Matrix. How does this come about? The point is that the interactions change the definition of the Poincaré symmetries, so these symmetries do not act on interacting particles the same way they act on free particles. To preserve the full Poincaré algebra with interactions, we need this causality condition.

- Constructing such an  $\mathcal{H}(x)$  satisfying the causality condition essentially requires the assembly of local fields  $\phi(x)$  with nice Lorentz transformation properties, because the creation and annihilation operators themselves do not have nice Lorentz transformation properties. The  $\phi(x)$  are constructed from the creation and annihilation operators for each species of particle, and then  $\mathcal{H}(x)$  is taken to be a polynomial in these fields.
- Symmetries constrain the asymptotic states and the S-Matrix. Gauge redundancies must be introduced to describe massless particles with a manifestly local and Poincaré invariant theory.
- One can prove (see chapter 13 of [3]) that only massless particles of spin  $\leq 2$  can couple in a way that produces long range interactions, and that massless spin 1 particles must couple to conserved currents  $J_\mu$ , while massless spin 2 particles must couple to  $T_{\mu\nu}$ . This obviously goes a long way towards explaining the spectrum of particles and forces that we have encountered.

*In theories that include gravity, the Weinbergian philosophy accords perfectly with the idea of Holography:* that we should view dynamical spacetime as an approximate description of a more fundamental theory in fewer dimensions, which ‘lives at infinity’. Holography was apparently not a motivation for Weinberg himself, and his construction can proceed with or without gravity. But the philosophy makes the most sense when we include gravity, in which case the S-Matrix is the only well-defined observable in flat spacetime.



## 2.2 Review of Lagrangians and Harmonic Oscillators

The harmonic oscillator is probably the most important system in physics. This isn't just bluster or experimental observation – there is a simple, universal reason.

Consider some classical system whose state can be specified using a set of coordinates. If the system is (meta)stable, then if we apply some small, momentary perturbing force, its state will not change drastically – this is what we mean by metastability. But in order for that to be true, there must be a restoring force. The restoring force will be a smooth function of the coordinates which vanishes when the system is in equilibrium. This means that unless there is a tuning, the force can be expanded in the displaced coordinate as

$$F = -kx \quad (2.2.1)$$

But this just means that for small  $x$ , the coordinate  $x$  is the position of a harmonic oscillator. Since most systems we encounter are metastable (otherwise they'd be undergoing rapid change right before our eyes) most of the degrees of freedom in most systems behave like harmonic oscillators under small perturbations.

We will discuss a harmonic oscillator using the Lagrangian  $\rightarrow$  Hamiltonian  $\rightarrow$  Quantum Mechanical description. Why?

1. There is a standard procedure called *Canonical Quantization* for finding a quantum mechanical version of a classical theory, once we have a Lagrangian description of the classical theory. We will review this as its one useful way to get to QFT.
2. Lagrangians are the most natural formalism for relativistic QFT, because they make it easy to formulate a theory that's manifestly Poincaré (Lorentz + Translation) invariant. You might ask, why not just start with the Hamiltonian, but to derive a Hamiltonian one must choose a specific time coordinate in order to define the canonical momenta, whereas the Lagrangian is agnostic about our choice of time. More generally, the Lagrangian formalism makes it easy to study other symmetries, and this will be crucial in QFT.

Let's review the harmonic oscillator from the point of view of canonical quantization and the Lagrangian and Hamiltonian formalisms. The action is

$$S = \int dt \left( \frac{1}{2} \dot{x}^2 - \frac{1}{2} \omega^2 x^2 \right) \quad (2.2.2)$$

The Euler-Lagrange equations are derived by imposing that the action is invariant under variations  $\delta x(t)$  from the classical trajectory  $x_{cl}$ :

$$\begin{aligned} 0 &= \delta S = \int dt \left( \dot{x}_{cl} \delta \dot{x} - \omega^2 x_{cl} \delta x \right) \\ &= \int dt \left( -\ddot{x}_{cl} - \omega^2 x_{cl} \right) \delta x \end{aligned} \quad (2.2.3)$$

Since this must vanish for any  $\delta x(t)$ , we find that

$$\ddot{x}_{cl} = -\omega^2 x_{cl} \quad (2.2.4)$$

which is the usual EoM for a spring.

Now let us discuss *Canonical Quantization*, in case it is unfamiliar. If you'd like to read about all the general details, I recommend Weinberg's Quantum Mechanics, chapter 9. We have derived the classical equations of motion, but how do we find a quantum mechanical description that reduces to the classical oscillator in the  $\hbar \rightarrow 0$  limit?

We define the momentum canonically conjugate to the  $x$  operator via

$$p \equiv \frac{\partial L}{\partial \dot{x}} = \dot{x} \quad (2.2.5)$$

and we impose the *canonical commutation relation*

$$[\hat{x}, \hat{p}] = i\hbar \quad (2.2.6)$$

When canonically quantizing, we always impose this commutation relation between the coordinate variables and their canonical momenta defined by equation (2.2.5).

There are several reasons for this. One explanation is that when  $\hat{x}$  actually measures the position of an object in space, then  $\hat{p}$  acts as the *generator of translations*, ie it acts to move  $\langle \hat{x} \rangle$  infinitesimally, so that

$$\langle e^{i\hat{p}\epsilon/\hbar} \hat{x} e^{-i\hat{p}\epsilon/\hbar} \rangle \approx \left\langle \hat{x} + \frac{i\epsilon}{\hbar} [\hat{p}, \hat{x}] \right\rangle = \langle \hat{x} \rangle + \epsilon \quad (2.2.7)$$

So at least we know that this canonical commutation relation makes sense when  $\hat{x}$  really is a position operator. A more general explanation is that insuring canonical commutation relations means that we recover Poisson brackets from commutators in the classical limit, so we recover the classical description of the system. In particular, these canonical commutation relations insure that *symmetries* work correctly – that is, that conserved currents associated with symmetries generate those symmetries. But we're getting ahead of ourselves, so let's leave it at the fact that we get the correct classical limit.

Now we can construct the (classical or quantum) Hamiltonian, which is defined as

$$H \equiv p\dot{x} - L = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\omega^2 x^2 = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 x^2 \quad (2.2.8)$$

and we just send  $x \rightarrow \hat{x}$  and  $p \rightarrow \hat{p}$  for the quantum version. The expression is the same whether  $p$  and  $x$  are functions (classical) or operators on a Hilbert space (quantum). In the latter case this Hamiltonian defines the time evolution of our system. Note that on the equations of motion, we have

$$\frac{d}{dt}H = \dot{x}\ddot{x} + \omega^2 x\dot{x} = \dot{x}(\ddot{x} + \omega^2 x) = 0 \quad (2.2.9)$$

so the Hamiltonian, or energy, is conserved. This follows whenever the Lagrangian has no explicit dependence on time; it is a consequence of a time translation symmetry, as we will discuss in more detail and generality later on.

Furthermore, you might remember that the Hamiltonian operator is supposed to generate time translations. So notice that

$$[\hat{H}, \hat{x}] = \frac{1}{2}[\hat{p}^2, \hat{x}] = i\hbar p = i\hbar \dot{\hat{x}} \quad (2.2.10)$$

and

$$[\hat{H}, \hat{p}] = \frac{1}{2}[\omega^2 \hat{x}^2, \hat{p}] = -i\hbar \omega^2 x = i\hbar \ddot{x} = i\hbar \dot{\hat{p}} \quad (2.2.11)$$

where in the second case we have used the equations of motion. This means that  $[H, \hat{O}]$  generates the time translations of an operator  $\hat{O}$  made out of  $\hat{x}$  and  $\hat{p}$ .

Note that the canonical commutation relation  $[\hat{x}, \hat{p}] = i\hbar$  is satisfied if we choose the *representation*  $\hat{p} = -i\hbar \partial_x$ . Note that this is an example of an explicit choice for  $\hat{p}$  that satisfies the canonical commutation relation, but it's not necessarily the only choice – there can be many representations for a given algebra of commutators. Now we can write the Schrodinger equation (switching now from the Heisberg to the Schrodinger picture, that is from a picture where operators depend on time, but states don't, to a picture where operators do not depend on time, and states do)

$$-i\hbar \partial_t \Psi = \hat{H} \Psi(t, x) = \left( -\frac{\hbar^2}{2} \partial_x^2 + \frac{1}{2} \omega^2 x^2 \right) \Psi(t, x) \quad (2.2.12)$$

and get the usual result for a non-relativistic harmonic oscillator. If instead we had a general  $V(x)$  potential, we would have recovered that correctly too. This is another, more familiar way of saying that the Hamiltonian generates time translations – it literally tells us the time derivative of the Schrodinger picture state.

To solve the system we must diagonalize the Hamiltonian. As you know, if we define annihilation and creation operators  $a$  and  $a^\dagger$ , with

$$[a, a^\dagger] = 1 \quad (2.2.13)$$

then we can write

$$x = \sqrt{\frac{\hbar}{2\omega}}(a + a^\dagger) \quad \text{and} \quad p = i\sqrt{\frac{\hbar\omega}{2}}(a^\dagger - a) \quad (2.2.14)$$

so that the canonical commutation relation is satisfied. But the real reason to do this is that now the Hamiltonian is just

$$H = \hbar\omega \left( a^\dagger a + \frac{1}{2} \right) \quad (2.2.15)$$

The overall constant of  $\omega/2$  doesn't have any meaning in the absence of gravity. The Hilbert space consists of the vacuum with

$$a|0\rangle = 0 \quad (2.2.16)$$

and the states built from it as  $(a^\dagger)^n|0\rangle$ . These have energy  $\omega(n + 1/2)$ .

We explained above why harmonic oscillators are universal. But what do they have to do with QFT and particle physics?

The point is to think deeply about what it means to have a *free* particle. If I have several different free particles with various energies, the energy of the total state is just *a sum of the energies of each particle*. But this is a feature of the harmonic oscillator system – the energy of the  $(n + m)$ th level is just the sum of the energies of the  $n$ th and  $m$ th level! This is why, as we will see in the next few lectures, in QFT, the oscillator creation operators and annihilation operators will become *particle creation and annihilation operators*. And the free particle Hamiltonian will just be an (infinite) sum of oscillator Hamiltonians, where the oscillators are labeled by the momenta and energies of the free particles.

As a final comment, what if we consider adding a small correction to the Hamiltonian

$$H_I = gx^{2n} = g \left( \frac{\hbar}{\omega} \right)^n (a + a^\dagger)^{2n} \quad (2.2.17)$$

If we treat this correction in perturbation theory, it looks like an interaction between  $2n$  of the oscillator modes (for example it could create  $2n$ , destroy  $2n$ , etc). This does not seem very interesting in the case of the oscillator, but in QFT these sort of interactions will be the perturbations that create and destroy particles, and lead to non-trivial scattering processes.

## 2.3 Balls and Springs

(See John McGreevy's notes <http://physics.ucsd.edu/mcgreevy/s14/239a-lectures.pdf> for some more material on getting a QFT from a lattice.)

Consider a model for a solid, or more specifically, a crystal. There's a regular pattern of atoms in specific positions. We will put the atoms in a position

$$x_i = x_{0i} + \phi_i \quad (2.3.1)$$

so  $\phi_i$  represents the displacement from the resting position. With our definitions, this means that  $\phi_i$  has *units of length*.

We are about to study sound waves in the solid. When these are quantized they are called *phonons*. We will learn much later in the course why their properties are universal (namely, because they are Goldstone bosons).

So back to basics – what happens if we try to move the  $i$ th atom? It cannot be held in place by anything other than its neighbors, since both the laws of physics and an empty universe are translation invariant. This means the potential must be a function

$$V(\phi_i - \phi_j) \quad (2.3.2)$$

where  $i$  and  $j$  are different atoms and  $\phi_i$  and  $\phi_j$  are their positions relative to their resting states. Let us define  $\phi_i = 0$  for all  $i$  when the atoms are at rest.

The neighbors don't exert a force on the atom when it's at rest, because otherwise it wouldn't be able to stay stably at rest. So the potential for the solid must satisfy

$$\left. \frac{\partial}{\partial \phi_i} V(\phi_a - \phi_b) \right|_{\text{all } \phi_a=0} = 0 \quad (2.3.3)$$

for any  $\phi_i$ . Furthermore, the *locality* of the interactions means that atoms that are far apart do not exert significant forces on each other.

This means that the Lagrangian should take the form (we'll discuss a 1-d solid for simplicity)

$$L = \sum_i \left[ \frac{1}{2} m \dot{\phi}_i^2 - \frac{1}{2} k \left( \frac{\phi_i - \phi_{i+1}}{a} \right)^2 - g_3 \left( \frac{\phi_i - \phi_{i+1}}{a} \right)^3 - g_4 \left( \frac{\phi_i - \phi_{i+1}}{a} \right)^4 - \dots \right] \quad (2.3.4)$$

There's much to unpack here. We have assumed that all interactions are nearest neighbor. The mass  $m$  of the atoms appears as a coefficient of the kinetic term (we have assumed they all have the same mass). The parameter  $a$  we have inserted could be absorbed away, but we have included it because it is the physical distance between atoms, or the 'lattice spacing'. So it is *the natural short-distance length scale* in our problem. Realizing that such short-distance 'cutoffs' exist will be absolutely crucial for understanding what's going on throughout the QFT course. You may have heard that there are 'infinities' in various QFT calculations; most of these come from imagining that cutoffs like  $a$  are simply zero.

Now let's consider dimensional analysis. Clearly  $k$ ,  $g_3$ ,  $g_4$ , etc have the same units, namely units of energy (the same units that the Lagrangian has). It's reasonable to guess that since these parameters can be compared, they are numerically comparable, i.e. no one of them is much bigger or smaller than the others. But this means that since (so that the solid is near its rest state, and is not anywhere close to breaking, or melting)

$$\left( \frac{\phi_i - \phi_{i+1}}{a} \right) \ll 1 \quad (2.3.5)$$

that the terms proportional to  $g_3$  and  $g_4$  are unimportant compared to  $k$ ! So now let's just study the simplified Lagrangian

$$L = \sum_i \left[ \frac{m}{2} \dot{\phi}_i^2 - \frac{m}{2} c^2 \left( \frac{\phi_i - \phi_{i+1}}{a} \right)^2 \right] \quad (2.3.6)$$

where I have re-labeled  $k \rightarrow mc^2$ . I will also assume that our lattice is periodic, so site  $N + 1$  is the same thing as site 1.

We interpret the first term in the Lagrangian as the kinetic energy of each atom, while the second term is  $-V$ , the negative of the potential energy stored in the bonds between atoms.

This is just a theory describing coupled harmonic oscillators. If we can diagonalize their interaction then we can solve the theory. Let's look at the equations of motion:

$$\ddot{\phi}_n = -c^2 \left( \frac{2\phi_n - \phi_{n+1} - \phi_{n-1}}{a^2} \right) \quad (2.3.7)$$

Now note that if we write  $\phi_n$  in terms of a Fourier series in  $n$ , then we have the transform and inverse transform

$$\phi_k(t) \equiv \frac{1}{\sqrt{N}} \sum_n e^{ikan} \phi_n(t) \quad \text{and} \quad \phi_n(t) = \frac{1}{\sqrt{N}} \sum_{k>0}^{2\pi/a} e^{-ikan} \phi_k(t) \quad (2.3.8)$$

where

$$k = \frac{2\pi}{Na} s \quad (2.3.9)$$

for integers  $s$  and  $N$  is the total number of atoms in the 1d solid. The equation of motion becomes

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{k>0}^{2\pi/a} e^{-ikan} \ddot{\phi}_k &= -\frac{c^2}{a^2} \frac{1}{\sqrt{N}} \left[ 2 \sum_{k>0}^{2\pi/a} e^{-ikan} \phi_k - \sum_{k>0}^{2\pi/a} e^{-ika(n+1)} \phi_k - \sum_{k>0}^{2\pi/a} e^{-ika(n-1)} \phi_k \right] \\ &= -\frac{c^2}{a^2} \frac{1}{\sqrt{N}} \sum_{k>0}^{2\pi/a} e^{-ikan} (2\phi_k - e^{-ika} \phi_k - e^{ika} \phi_k) \\ &= -\frac{c^2}{a^2} \frac{1}{\sqrt{N}} \sum_{k>0}^{2\pi/a} e^{-ikan} \phi_k (2 - 2\cos(ka)) \end{aligned} \quad (2.3.10)$$

If we take the inverse Fourier series, then we must have

$$\ddot{\phi}_k = -\frac{c^2}{a^2} (2 - 2\cos(ka)) \phi_k \quad (2.3.11)$$

so by Fourier representing, we have diagonalized the equations of motion! In other words, we have  $N$  independent harmonic oscillators with frequencies

$$\omega(k) = 2\frac{c}{a} \sin\left(\frac{ka}{2}\right) \quad (2.3.12)$$

where  $k = \frac{2\pi}{Na}, \frac{4\pi}{Na}, \dots, \frac{2\pi}{a} \frac{N-1}{N}$ . Note that there are only  $N-1$  modes because the lagrangian only depended on the differences  $\phi_i - \phi_{i+1}$ .

Plot value of  $\omega$  as a function of  $k$ . Also now look at what a single wave with definite  $k$  looks like on the lattice. It's a sound wave! Clearly  $k$  just sets the wavelength. We can use  $k = -\frac{\pi}{a}, \frac{\pi(N-2)}{aN}, \frac{\pi(N-4)}{aN}, \dots, \frac{\pi}{a}$  instead for a more symmetric set of  $k$ . This is called the Brillouin zone. These oscillator modes are called phonons, or more specifically, acoustic phonons.

### 2.3.1 Quantization of Phonons and Fock Space

We are studying a theory of atoms with nearest neighbor interactions, or ‘balls and springs’. How do we quantize this theory? Diagonalizing the equations of motion also diagonalizes the Lagrangian and Hamiltonian, for example the kinetic term is

$$\sum_n \dot{\phi}_i^2 = \sum_n \frac{1}{N} \sum_{s,r} e^{-i\frac{2\pi}{N}(sn+rn)} \dot{\phi}_s \dot{\phi}_r \quad (2.3.13)$$

$$= \frac{1}{N} \sum_{s,r} N \delta_{s,-r} \dot{\phi}_s \dot{\phi}_r \quad (2.3.14)$$

$$= \sum_{s=-N/2}^{N/2} \dot{\phi}_s \dot{\phi}_{-s} \quad (2.3.15)$$

where  $k = \frac{2\pi s}{Na}$  as usual. You might be surprised to see  $\phi_s \phi_{-s}$ , instead of  $\phi_s^2$ , but note that

$$\phi_k^\dagger \equiv \left( \frac{1}{\sqrt{N}} \sum_n e^{ikan} \phi_n \right)^\dagger = \frac{1}{\sqrt{N}} \sum_n e^{-ikan} \phi_n = \phi_{-k} \quad (2.3.16)$$

because  $\phi_n^\dagger = \phi_n$ , so in effect we have  $N$  real degrees of freedom packaged into  $N/2$  complex modes. This can also be understood as conservation of momentum – the sum of the momenta has to vanish. We can re-write the Lagrangian as

$$L = \frac{1}{2} \sum_{k=-\frac{\pi}{Na} \frac{N}{2}}^{\frac{\pi}{Na} \frac{N}{2}} \left[ \dot{\phi}_k \dot{\phi}_{-k} - \frac{c^2}{a^2} \sin^2 \left( \frac{ka}{2} \right) \phi_k \phi_{-k} \right] \quad (2.3.17)$$

where I set  $m = 1$  for simplicity. Now we just have  $N - 1$  independent harmonic oscillators. We recalled how to quantize oscillators in a past lecture.

Thus we can write

$$\phi_k = \sqrt{\frac{a}{4c \sin \left( \frac{|k|a}{2} \right)}} \left( a_k + a_{-k}^\dagger \right) \quad \text{and} \quad \pi_k = i \sqrt{\frac{c \sin \left( \frac{|k|a}{2} \right)}{a}} \left( a_{-k} - a_k^\dagger \right) \quad (2.3.18)$$

for the canonical variables  $\phi_k$  and their canonical momenta. Don’t confuse the lattice spacing  $a$  with the annihilation operators  $a_k$ . Remember that  $k$  is a label here, which can take  $N - 1$  different values. The crucial point is that the Hilbert space is the space generated by states like

$$(a_{k_1}^\dagger)^{n_1} (a_{k_2}^\dagger)^{n_2} \dots (a_{k_l}^\dagger)^{n_l} |0\rangle \quad (2.3.19)$$

where all  $a_k$  annihilate the vacuum  $|0\rangle$ . This state is a superposition of  $n_i$  waves with momentum  $k_i$ . The space of states is a *Fock Space*. This is simple, but it’s important to understand what this

means. Note that the fact that energies add is related to a very familiar property of waves (e.g. water waves) – that they pass right through each other without scattering! This is also what free particles do.

What is the Hamiltonian? It is just a sum of Hamiltonians for each oscillator, so it is

$$H = \sum_{k=-\frac{\pi}{a}}^{\frac{\pi}{a}} \hbar\omega(k) \left( a_k^\dagger a_k + \frac{1}{2} \right) \quad (2.3.20)$$

The total energy is just the sum of energies in each  $k$  mode. Note that the zero mode's  $1/2$  now gets summed over all of these modes, so the sum may be quite large!

### 2.3.2 Continuum Limit

Now let us consider the continuum limit of the theory. This means we send the lattice spacing  $a \rightarrow 0$  keeping other parameters fixed. However, since  $a$  is dimensionful, its numerical value is meaningless; what we really mean by the continuum limit is that  $a/L$  goes to zero, where  $L$  is some length scale at which we are conducting an experiment. Since we have modes with inverse wavelength  $k$ , this means  $ka \rightarrow 0$ .

First, let us consider what happens to  $k$  itself. The dimensionless combination  $ka = \frac{2\pi n}{N}$  for integer  $n$ , so the fact that this is small means that we should take  $N \rightarrow \infty$ . In this case  $(ka)$  becomes a continuum, so the inverse wavelength just becomes a real parameter. What happens to the energies (frequencies)? We find

$$\omega(k) = 2\frac{c}{a} \left| \sin \left( \frac{ka}{2} \right) \right| \rightarrow ck \quad (2.3.21)$$

This is the *relativistic dispersion relation for a massless particle*. For example, this is the dispersion relation for photons, if  $c$  is the speed of light. In our case we have *phonons*, or quantized sound waves, and  $c$  is the speed of sound.

What happens to the Lagrangian and Hamiltonian? The sums over  $k$  must become integrals, so

$$L = \frac{1}{2} \sum_{k=-\frac{2\pi}{Na}}^{\frac{2\pi}{a} \frac{N-1}{N}} \left[ \dot{\phi}_k \dot{\phi}_{-k} - \frac{c^2}{a^2} \sin^2 \left( \frac{ka}{2} \right) \phi_k \phi_{-k} \right] \rightarrow \frac{1}{2} \int dk \left[ \dot{\phi}_k \dot{\phi}_{-k} - c^2 k^2 \phi_k \phi_{-k} \right] \quad (2.3.22)$$

This is the momentum space version of a simple, position space Lagrangian

$$L = \int dx \frac{1}{2} [(\partial_t \phi)^2 - c^2 (\partial_x \phi)^2] \quad (2.3.23)$$

This is actually obvious from our original form of the Lagrangian – the finite differences

$$\frac{\phi_i - \phi_{i+1}}{a} \rightarrow \partial_x \phi \quad (2.3.24)$$



in the limit that  $a \rightarrow 0$  and  $N \rightarrow \infty$ , the continuum limit. We're just using the definition of the derivative.

The quantized fields become

$$\phi_k = \sqrt{\frac{1}{\omega(k)}} (a_k + a_k^\dagger) \quad \text{and} \quad \pi_k = i\sqrt{\omega(k)} (a_k - a_k^\dagger) \quad (2.3.25)$$

where now  $a_k$  annihilates a continuum wave with momentum  $k$ . So we have a Fock space of states

$$(a_{k_1}^\dagger)^{n_1} (a_{k_2}^\dagger)^{n_2} \dots (a_{k_l}^\dagger)^{n_l} |0\rangle \quad (2.3.26)$$

except that the labels  $k_i$  are continuous. Taking the mature point of view that  $\hbar = 1$ , so that frequencies and energies are the same thing, the Hamiltonian becomes

$$H = \int_{-\infty}^{\infty} dk \omega(k) \left( a_k^\dagger a_k + \frac{1}{2} \right) = \int_{-\infty}^{\infty} dk c|k| \left( a_k^\dagger a_k + \frac{1}{2} \right) \quad (2.3.27)$$

You might notice that the  $\frac{1}{2}$  part of the integral is infinity! This is our first QFT infinity. It's a contribution to the energy of the vacuum, the lowest energy state, and it's meaningless here, since it cannot be measured. It would clearly have been finite if we still had our shortest distance  $a$ , since then  $k < 1/a$  would have been mandatory. So the short distance cutoff is also a high-energy cutoff.

### 2.3.3 Other Interaction Terms in Perturbation Theory, and Scaling

Let us look at what sort of effect we would have from the interaction terms we ignored. For example, we discussed a possible term

$$g_4 \left( \frac{\phi_i - \phi_{i+1}}{a} \right)^4 \rightarrow g_4 (\partial_x \phi)^4 \quad (2.3.28)$$

in the continuum limit. Let us consider a situation where the effect of this term is extremely small, so it can be treated in perturbation theory in  $g_4$ . In such a case it is just a small perturbative correction to the Hamiltonian. Schematically, since  $\partial_x \sim k$ , this correction looks like

$$H_I \sim g_4 \sum_{k_i} \delta(k_1 + k_2 + k_3 + k_4) k_1 k_2 k_3 k_4 (a_{-k_1} + a_{k_1}^\dagger) (a_{-k_2} + a_{k_2}^\dagger) (a_{-k_3} + a_{k_3}^\dagger) (a_{-k_4} + a_{k_4}^\dagger) \quad (2.3.29)$$

The integral over  $x$  gave a momentum conserving delta function, the  $\partial_x$  factors became  $k_i$ , but most importantly, this is *an interaction among 4 sound waves*. If we included it, they wouldn't just pass through each other anymore, but could scatter. Note that although it appears that it could create or destroy four sound waves out of nothing, that actually can't happen, due to momentum and energy conservation.

Notice that this term is of order (momentum)<sup>4</sup>. Another example of such a term would have been

$$(\partial_x^2 \phi)^2 \sim k^4 \phi_k^2 \quad (2.3.30)$$

Since we are studying the long wavelength limit,  $\lambda \rightarrow \infty$ , we want  $k \rightarrow 0$  (or more precisely  $ka \rightarrow 0$ ), and so higher powers of the momenta are smaller. So by ignoring such terms we are not really leaving anything out – *we are just making a long-distance approximation*.

## 2.4 Special Relativity and Anti-particles

For our purposes, special relativity is tantamount to the statement that space and time have certain symmetries.

### 2.4.1 Translations

The most obvious part of those symmetries is translation invariance – the fact that the laws of physics are the same everywhere, and at all times. Let us consider what this means in an extremely simple context, namely just the space of functions of various spatial variables  $x_i$  labeling a bunch of points scattered in space.

If a function  $f(x_1, x_2, \dots, x_n)$  is translation invariant, then

$$f = f(x_i - x_j) \quad (2.4.1)$$

Translation invariance is the reason that momentum space (and the Fourier Transform) is a useful idea. We have that

$$f(x + y) = f(x) + y\partial f(x) + \dots = e^{y\partial_x} f(x) \quad (2.4.2)$$

so we translate a function using spatial derivatives. But in Fourier space

$$f(x_i) = \int_{-\infty}^{\infty} \left( \prod_{i=1}^n \frac{dp_i}{2\pi} \right) \tilde{f}(p_i) e^{-ip_i x_i} \quad (2.4.3)$$

which means that

$$\frac{\partial}{\partial x_i} f(x_i) = ip_i \tilde{f}(p_i) \quad (2.4.4)$$

so momentum space *linearizes the action of translations*. Instead of having to compute a (spatial) derivative, in momentum space we can differentiate by simply multiplying by  $ip_i$ . This exponentiates, so moving  $f(x) \rightarrow f(x + y)$  just requires multiplication by the phase  $e^{ipy}$ .

Translation invariance implies that

$$f(x_i - x_j) = \int \left( \prod_{i=1}^n \frac{dp_i}{2\pi} \right) \tilde{f}(p_i) e^{-ip_i x_i} \delta(p_1 + p_2 + \dots + p_n) \quad (2.4.5)$$

In other words, translation invariance implies momentum conservation, and vice versa. Momentum and energy are conserved because the laws of physics are invariant under translations in space and time, respectively. We will explain this again in a more sophisticated way very soon. In general, the fact that *symmetries imply conservation laws* is called Noether's Theorem.

### 2.4.2 Lorentz Transformations and Invariants

Next we want to talk about the part of special relativity that seems non-trivial – Lorentz transformations. But first, rotations, because Lorentz transformations are essentially just a mixture of rotations and rotations with a sign flip. If we just have an  $(x, y)$  plane this is just

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (2.4.6)$$

As a bit of notation, we can write this as

$$x_i \rightarrow R_{ij}x_j \equiv R_{i1}x_1 + R_{i2}x_2 \quad (2.4.7)$$

this is the Einstein summation convention, where repeated indices are summed. Sometimes the summed indices are referred to as having been contracted. Note that  $R^T = R^{-1}$ , so

$$R^T R = (R^T)_{ij} R_{jk} = \delta_{ik} = \mathbb{1}_{ik} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.4.8)$$

This algebraic definition  $R^T R = \mathbb{1}$  is the most useful and generalizable; it tells us that rotations form the group  $SO(2)$ , the special orthogonal group acting on 2-vectors. Note that the norm

$$x^i x_i = (x \ y) \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2 \quad (2.4.9)$$

is invariant under rotations, so

$$x^i x_i = (x^i R_{ij}^T)(R_{jk} x_k) = x^i \delta_{ik} x_k = x^i x_i \quad (2.4.10)$$

In fact, it's probably best to *define*  $SO(2)$  as the set of transformations preserving this inner product.

Now we can consider the larger group  $SO(4)$ , which preserve an inner product  $\tau^2 + x^2 + y^2 + z^2$  on the vector  $(\tau, x, y, z)$ . This is very close to the Lorentz group... the only difference is in the way it treats time. The Lorentz group, or  $SO(1, 3)$ , is simply the group of linear transformations that preserves

$$t^2 - x^2 - y^2 - z^2 = x_\mu x^\mu = \eta_{\mu\nu} x^\mu x^\nu = \eta^{\mu\nu} x_\mu x_\nu \quad (2.4.11)$$

So in other words, Lorentz transformation matrices must satisfy

$$\Lambda^T \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \eta_{\mu\nu} \quad (2.4.12)$$

This includes the  $SO(3)$  group of rotations in space, plus Boosts, which are just ‘rotations’ between time and space. However, to satisfy this defining equation, we need to switch from cosines and sines

to cosh and sinh, e.g.

$$\Lambda = \begin{pmatrix} \cosh \beta_x & \sinh \beta_x & 0 & 0 \\ \sinh \beta_x & \cosh \beta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.4.13)$$

where the  $\beta$  are hyperbolic angles sometimes called rapidities. We can relate this to a velocity (a boost just means that we've shifted to a moving frame) by noting that

$$x \rightarrow \frac{x + vt}{\sqrt{1 - v^2}}, \quad t \rightarrow \frac{t + vx}{\sqrt{1 - v^2}} \quad (2.4.14)$$

leave  $t^2 - x^2$  invariant and reduce to the Galilean transformation  $x \rightarrow x + vt$  at small  $v$ . Thus we see that

$$\sinh \beta = \frac{v}{\sqrt{1 - v^2}} \quad (2.4.15)$$

and  $\beta \approx v$  when these quantities are small.

When it comes to symmetries such as those of special relativity, the major question is how do various quantities of interest transform under the symmetry? In mathematical language, this is the question of the *representation of the symmetry group*.

The book doesn't really talk about Lorentz transformations in a fully consistent way. Surprisingly, these issues can get really confusing, and there are even two distinct ways of discussing them, often called 'active' and 'passive', depending on whether we think of the fields as transforming, or the coordinates.

So first let's mention what **the book says, implicitly using 'passive' transformations**: The simplest thing that can happen is that a particular quantity doesn't transform at all. Such a quantity is called a *scalar*. If the quantity takes values throughout spacetime, it is called a *scalar field*, and we say

$$\phi(x) \rightarrow \phi(x) \quad (2.4.16)$$

under a Lorentz transformation  $\Lambda_{\mu\nu}$ . Scalars are completely fixed under Lorentz transformations. You might get confused and think that  $\phi(x) \rightarrow \phi(\Lambda x)$ , but this is just a relabeling of points – the field is not changing as long as we evaluate it at the same spacetime point, which has a fixed physical meaning (and in particular, the fact that we can relabel points doesn't tell us anything about the symmetries of the theory).

OK, so that was the passive point of view. It's useful because it sounds very simple, but it brushes a lot under the rug. The simplest example of the 'active' point of view, where the fields transform, is the case of translations. Under an infinitesimal translation  $x \rightarrow x + \epsilon$  we have

$$\phi(x) \rightarrow \phi(x) + \epsilon^\nu \partial_\nu \phi(x) \quad (2.4.17)$$

This defines how translations act on scalar fields. In the example where we discussed sound waves, we had an action that compared  $\phi_n$  and  $\phi_{n+1}$ , ie it compared  $\phi(x)$  to its translation  $\phi(x+a)$ . In the continuum limit this (the difference between the two) became the derivative, as above.

From the ‘active’ point of view, all fields also transform under Lorentz transformations. If we consider an infinitesimal transformation  $\Lambda_{\mu\nu} = 1 + L_{\mu\nu}$  (where  $L_{\mu\nu}$  is an element of the ‘Lie algebra’ with  $L^T \eta = \eta L$  as a matrix), then

$$\phi(x) \rightarrow \phi(x) + L^{\mu\nu} x_\mu \partial_\nu \phi(x) \quad (2.4.18)$$

Note that this is just a generalizations of how rotations act on functions of position.

Perhaps an even more familiar example is a *vector* or vector field  $V^\mu(x)$ . In the passive language of the book, it transforms as

$$V^\mu \rightarrow \Lambda^\mu_\nu V^\nu \quad (2.4.19)$$

where as usual, repeated indices are summed. As concrete examples, one can consider the more familiar case of rotations, and the electric field  $\vec{E}$  and its scalar potential  $\phi$ , where  $\vec{E} = -\vec{\nabla}\phi$ .

Some important examples of vectors (but not fields)

$$x^\mu = (t, x, y, z) \quad (2.4.20)$$

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = (\partial_t, \partial_x, \partial_y, \partial_z) \quad (2.4.21)$$

$$p^\mu = (E, p_x, p_y, p_z) \quad (2.4.22)$$

We build Lorentz invariants by contracting together indices, e.g.

$$V^\mu W_\mu = V^\mu W^\nu \eta_{\mu\nu} \quad (2.4.23)$$

This contraction is Lorentz invariant because it transforms like a scalar field. Tensors of rank  $k$  transform linearly, e.g.

$$T^{\mu\nu} \rightarrow \Lambda^\mu_\alpha \Lambda^\nu_\beta T^{\alpha\beta} \quad (2.4.24)$$

We tend to forget which indices are upper and lower because we can raise and lower with  $\eta_{\mu\nu}$ , which is (almost) the identity matrix, and this also works for  $\Lambda$  and  $\Lambda^T$ .

Lorentz invariants are things like

$$V^2 = V_\mu V^\mu, \quad \phi, \quad 1, \quad \partial_\mu V^\mu \quad (2.4.25)$$

while Lorentz covariant objects are e.g.

$$V_\mu, \quad F_{\mu\nu}, \quad \partial_\mu, \quad x_\mu \quad (2.4.26)$$

Note that some quantities are neither, for example the energy density is the 00 component of the energy-momentum tensor  $T_{\mu\nu}$ .

The book wisely informs you that almost all you really need to know about special relativity is how to identify if a given quantity is invariant, covariant, or none of the above. For this purpose the passive point of view is easier, but you should watch out, as it can be tricky to use once you scratch the surface.

### 2.4.3 Discrete Transformations and Invariants

There are two discrete transformations that may or may not be symmetries, parity and time reversal:

$$P : (t, x, y, z) \rightarrow (t, -x, -y, -z) \quad (2.4.27)$$

$$T : (t, x, y, z) \rightarrow (-t, x, y, z) \quad (2.4.28)$$

These can easily be written as diagonal matrices acting on general vectors and tensors via  $\Lambda_T$  and  $\Lambda_P$  transformations. Parity and time reversal cannot be written as products of rotations and boosts, and usually I won't refer to them as 'Lorentz transformations'. Note that  $P$  and  $T$  leave inner products like  $V_\mu V^\mu$  invariant, because  $(-1)^2 = 1$ .

Note that no Lorentz transformation changes the sign of invariants such as

$$V_\mu V^\mu \quad (2.4.29)$$

so we say that when this is positive,  $V_\mu$  is *timelike*, when it vanishes it is *null* or *lightlike*, and when it's negative it's *spacelike*. Obviously  $(t, 0, 0, 0)$  is timelike and  $(0, x, 0, 0)$  is spacelike. All one-particle momenta are timelike (if the particle is massive) or lightlike (if the particle is massless). So photon momenta are always lightlike.

An important conceptual point is that *causality is ambiguous for spacelike separations*. Concretely, if we have some spacelike separation between two spacetime points  $A$  and  $B$ , say at

$$A = (0, 0, 0, 0) \quad \text{and} \quad B = (0, x, 0, 0) \quad (2.4.30)$$

then we can perform a Lorentz transformation so that  $A_0 > B_0$  or  $A_0 < B_0$ , so there time coordinates can have any order. This is impossible for time-like separated points, whose time-ordering is unambiguous in all reference frames.

This means that for time-like vectors, the sign of the zero component is a Lorentz invariant. Concretely, if we have a timelike vector  $V_\mu = (1, 0, 0, 0)$ , no Lorentz transformation can make  $V_0$  negative, and so  $\text{sign}(V_0)$  is Lorentz invariant. This is easy to see since  $V_\mu V^\mu = 1$  is invariant, so  $t^2 > 1 + x^2$  and  $t$  cannot change sign.

### 2.4.4 Need for Anti-particles and QFT due to Relativity

When we discussed atoms and phonons, or 'balls and springs', we saw that an  $\omega = ck$  dispersion relation and the creation and destruction of 'particles' can *emerge* naturally from the long-distance limit of a fairly conventional system. This provides one motivation for QFTs, without any requirement for special relativity.

However, it turns out that once we demand both quantum mechanics and special relativity, quantum field theory becomes nearly obligatory. In particular, one cannot simply pursue a 'relativistic version' of the problems you studied in non-relativistic quantum mechanics. Let us see why via a beautiful and classic thought experiment. Apologies if it seems a bit abstract at first; it'll be clearer why certain quantities appear later in the semester once you know more QFT.

Consider a classical apparatus that has a probability amplitude  $J_E(x)$  of producing a particle at position  $x = (x^0, x^i)$  in space-time, and an absorption apparatus with amplitude  $J_A(y)$  to absorb a

particle at  $y$ . Let us assume that the particle propagates freely between the emission and absorption apparatus, with a mass  $m$ . Quantum mechanics tells us that the total amplitude for this process is

$$A = \int d^4x d^4y \langle x | e^{-iH(x^0-y^0)} | y \rangle J_A(x) J_E(y) \quad (2.4.31)$$

Let's unpack this formula. The formula  $\langle x | e^{-iH(x^0-y^0)} | y \rangle$ , often called a propagator, is the amplitude for a particle to propagate from  $x$  to  $y$  in the Schrodinger picture. We have integrated over all  $x$  and  $y$  because we are summing over all the places and times where the emission could have occurred, followed by an absorption.

Now we ask – is this expression Lorentz invariant? Note that the energy of this free particle is just  $\sqrt{p^2 + m^2}$ , so we can put that into the Hamiltonian if we use momentum space states. This gives

$$A = \int d^4x d^4y J_A(x) J_E(y) \int d^3p |\langle (z=0) | p \rangle|^2 e^{-ip \cdot (x-y)} \quad (2.4.32)$$

Note that the propagator only depends on  $x - y$ , as one would expect (by invariance under translations), and that the product in the exponent is the manifestly Lorentz-invariant 4-vector product. The matrix element  $\langle (z=0) | p \rangle$  is the overlap of a particle at  $z=0$  with the momentum state  $|p\rangle$ . It is the only part of the expression that may not transform nicely under Lorentz transformations.

However, we can guess its Lorentz invariant form by writing

$$\int d^3p |\langle 0 | p \rangle|^2 e^{-ip \cdot (x-y)} \propto \int d^4p \delta(p^2 - m^2) \theta(p^0) e^{ip \cdot (x-y)} \quad (2.4.33)$$

Doing the integral over  $p_0 = \omega_p$  using the delta function means

$$\langle 0 | p \rangle \propto \frac{1}{\sqrt{\omega_p}} \quad (2.4.34)$$

Note that in the non-relativistic limit this is just  $1/\sqrt{m}$ , so the particle is spread out over its compton wavelength.

Now with this relativistic wavefunction, the overall amplitude is Lorentz invariant (when the  $J$  transform as scalar fields), but there seems to be a paradox: *the amplitude is non-zero even when the separation between points is spacelike*. The causal order of spacetime points is not Lorentz invariant when they are spacelike separated, so this is a problem. To see this, note that the propagator is

$$\int d^3p |\langle 0 | p \rangle|^2 e^{-ip \cdot (x-y)} = \int_0^\infty \frac{dp}{(2\pi)^2} \frac{p^2}{\sqrt{m^2 + p^2}} \int_{-1}^1 d\cos\theta e^{iRp\cos\theta} \quad (2.4.35)$$

$$= \int_0^\infty \frac{dp}{(2\pi)^2} \frac{p}{\sqrt{m^2 + p^2}} \frac{\sin(Rp)}{R} \quad (2.4.36)$$

We can evaluate this when the mass is zero to find

$$\int d^3p |\langle 0 | p \rangle|^2 e^{-ip \cdot (x-y)} \propto \frac{1}{R^2} \quad (2.4.37)$$

so it's non-vanishing outside the lightcone. This means that *we cannot interpret this as a process with definite causal ordering*, because spacelike separated events can have either order in time, depending on the reference frame.

We cannot fix the problem by making the propagator vanish outside the lightcone, because we already wrote down the most general Lorentz invariant expression for this function, given that the particle has positive energy and  $p^2 = m^2$ . This follows because the only Lorentz invariant is  $p^2 = m^2$ , and additional  $m^2$  dependence would not change the dependence on  $x - y$ .

Thus we have learned that *every emission source must be an absorption source, and vice versa*. So we can take  $J = J_E + J_A$  and write

$$\begin{aligned} A &= \int d^4x d^4y J(x) J(y) \int \frac{d^3p}{2\omega_p} [\theta(x^0 - y^0) e^{-ip \cdot (x-y)} + \theta(y^0 - x^0) e^{ip \cdot (x-y)}] \\ &= \int d^4x d^4y J(x) J(y) D_F(x - y) \end{aligned} \quad (2.4.38)$$

The step functions  $\theta$  impose that one term can be interpreted as emission at  $y$  and absorption at  $x$ , and the other has the opposite interpretation. The point is that *different observers will disagree on which events were emission and absorption when  $x - y$  is spacelike, but they will agree on the amplitude*.

Now let's imagine that the particle carries some conserved and Lorentz invariant charge, such as electric charge. In such a case one could associate emission and absorption with charge flow. But this would be inconsistent with the fact that for spacelike  $x - y$ , there is no preferred causal order. Thus we must accept that *for every particle of with some given mass and charge, there must be an otherwise identical anti-particle with opposite charge*. Note that in the absence of charge, particles can be their own anti-particle (note, though, that we can think of a  $+$  helicity photon as the anti-particle of a  $-$  helicity photon).

Next let us consider a more complicated scenario, where the particle scatters off of a potential located at a space-time point  $z$  after it is emitted and before it is absorbed.

$$A_{scatter}(z) = \int d^4x d^4y J(x) J(y) S(z) D_F(x - z) D_F(z - y) \quad (2.4.39)$$

where  $S(z)$  is some amplitude for scattering there. We can draw a simple diagram for this. The point is that the amplitude will be non-zero even when  $x - z$ ,  $y - z$ , and  $x - y$  are all spacelike separated. This means that there will be a Lorentz frame where the scattering at  $z$  occurs before the emission and absorption at  $x$  and  $y$ . This means that an observer in this frame *sees the creation of a pair of particles by the scatterer at  $z$ !* Thus scattering processes necessarily entail the creation and destruction of particles.

Thus quantum mechanics, special relativity, and causality imply that there must exist processes where particles are created and destroyed. These arguments also show that it's difficult to localize a particle in a region smaller than its Compton wavelength – to do so we would need to probe the particle with much larger momenta, but this would lead to the production of more particles, not to the localization of a single particle in a smaller region.



Finally... if these arguments seemed too quick for you, consider returning to them after we study perturbation theory, especially the so-called ‘Old Fashioned Perturbation Theory’ a few lectures on. In that case one can study the annihilation of two charged particles into a pair of neutral particles through a  $t$ -channel exchange of a charged particle, making the argument for anti-particles extremely explicit and physically accurate.

## 2.5 Canonical Quantization and Noether’s Theorem

We’ve seen a simple example of a (not necessarily relativistic) QFT, and we’ve reviewed special relativity and its implications when combined with Quantum Mechanics. From now on we’ll mostly be studying relativistic theories.

### 2.5.1 General Lagrangians and Noether

To study a relativistic theory, we introduce a Lagrangian, and then canonically quantize it to get a quantum Hamiltonian. Why bother with a Lagrangian at all? The reason is that *Lagrangian densities are Lorentz scalars, while Hamiltonians, which require inputting an explicit notion of time, are not*. The Hamiltonian is just the  $T_{00}$  component of the energy-momentum tensor, so it is neither invariant nor covariant. Thus it is easy to obtain a Lorentz invariant theory by writing down Lorentz invariant Lagrangians, and then deriving a Hamiltonian; if we write down a random Hamiltonian, it’s hard to know if it will give rise to a Lorentz invariant theory. This is why all relativistic QFT begins with Lagrangians. In condensed matter physics, you can just start with a Hamiltonian if you like, since Lorentz invariance is usually not relevant.

You’re already familiar with Lagrangians  $L(q_i, \dot{q}_i, t)$ , where the  $q_i$  are the (many, possibly infinitely many) coordinates, e.g. the  $\phi_i$ , the displacements of the atoms in a crystal. We demand that the action

$$S = \int dt L(q_i, \dot{q}_i, t) \quad (2.5.1)$$

is stationary with respect to variations  $\delta q_i(t)$  in the trajectories, and this gives

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \quad (2.5.2)$$

the Euler-Lagrange equations of motion.

Now we will discuss one of the most important theorems in physics, *Noether’s theorem, which says that symmetries  $\implies$  conservation laws*. This theorem is extremely useful in QFT, although it’s true in classical mechanics, non-relativistic quantum mechanics, and classical field theory as well.

Consider some infinitesimal transformation of the coordinate variables  $q_i(t)$

$$q_i \rightarrow q_i + \epsilon F_i(q) \quad (2.5.3)$$

where  $\epsilon$  is an infinitesimal constant, and the  $F_i$  are functions of the  $q_i$  that depend on the symmetry. Our transformation is a symmetry of the Lagrangian if

$$0 = \delta L = \sum_i \left[ \frac{\partial L}{\partial q_i} F_i + \frac{\partial L}{\partial \dot{q}_i} \dot{F}_i \right] \quad (2.5.4)$$

Note that this does not use the equations of motion. However, we can now use the Euler-Lagrange equations to write this as

$$0 = \sum_i \left[ \left( \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) F_i + \frac{\partial L}{\partial q_i} \dot{F}_i \right] = \frac{d}{dt} F \quad (2.5.5)$$

where

$$F \equiv \sum_i \frac{\partial L}{\partial \dot{q}_i} F_i(q) \quad (2.5.6)$$

is a conserved quantity. So we have derived a conservation law from the symmetry!

For example, if we have a theory of two harmonic oscillators with the same frequency

$$L = \frac{1}{2} (\dot{x}^2 + \dot{y}^2) - \frac{1}{2} \omega^2 (x^2 + y^2) \quad (2.5.7)$$

then we could have  $F_x = y$  and  $F_y = -x$ . Then

$$\delta L = \frac{1}{2} (2\dot{x}\dot{y} + 2\dot{y}(-\dot{x})) - \frac{1}{2} \omega^2 (2xy + 2y(-x)) = 0 \quad (2.5.8)$$

so we have discovered a symmetry under rotations in the  $x$ - $y$  plane. This implies that

$$\ell \equiv \sum_i \frac{\partial L}{\partial \dot{q}_i} F_i(q) = \dot{x}y - \dot{y}x \quad (2.5.9)$$

is a conserved quantity. Of course it is just the usual angular momentum in the  $x$ - $y$  plane.

## 2.5.2 Hamiltonians and the Hamiltonian Formalism

There's another, special conserved quantity associated with *time translation symmetry*, namely the Hamiltonian or energy. We define the canonical momenta as

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i} \quad (2.5.10)$$

and then obtain the Hamiltonian

$$H = \sum_i \dot{q}_i p_i - L \quad (2.5.11)$$

Its rate of change is (using the E-L equations)

$$\frac{d}{dt} H = \sum_i \left[ \ddot{q}_i \frac{\partial L}{\partial \dot{q}_i} + \dot{q}_i \frac{\partial L}{\partial q_i} \right] - \frac{d}{dt} L \quad (2.5.12)$$

$$= -\frac{\partial}{\partial t} L \quad (2.5.13)$$

because each of the terms in the first line is also a term in the rate of change of the Lagrangian. So if the Lagrangian has no explicit time dependence (so it's time translation invariant), the Hamiltonian is conserved.

The purpose of the Hamiltonian formalism is to trade 2nd order differential equations in  $q_i$  for first order differential equations in the phase space variables  $(q_i, p_i)$ . So if we eliminate  $\dot{q}_i$  for  $p_i$  everywhere, then we obtain

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad \text{and} \quad \dot{q}_i = \frac{\partial H}{\partial p_i} \quad (2.5.14)$$

although I haven't shown you the details.

### 2.5.3 Canonically Quantizing

The point of canonical quantization is to transform  $q_i, p_i, H$ , and  $F$  (for any of the system's symmetries) so that  $H$  generates time translations, and  $F$  generates its associated symmetry transformation, in the sense that

$$[F, q_i] = -i\hbar F_i(q) \quad (2.5.15)$$

and in the case of the Hamiltonian

$$[H, q_i] = -i\hbar \dot{q}_i \quad \text{and} \quad [H, p_i] = i\hbar \dot{p}_i \quad (2.5.16)$$

Note that this is in the Heisenberg picture, where operators depend on time but the states do not.

Both of these relations are satisfied if we make the canonical choice

$$[q_i, p_j] = i\hbar \delta_{ij} \quad (2.5.17)$$

and all other commutators between the  $q_i$  and  $p_j$  vanish. Let's check this for  $F$ , which we defined as

$$F \equiv \sum_i \frac{\partial L}{\partial \dot{q}_i} F_i(q) = \sum_i p_i F_i(q) \quad (2.5.18)$$

Now we see immediately that

$$[F, q_n] = \sum_i [p_i F_i(q), q_n] = -i\hbar F_n(q) \quad (2.5.19)$$

because the  $i = n$  term is picked out by the commutator, and  $F_n(q)$  commutes with  $q_n$ . So canonical quantization guarantees that symmetries are generated by commutation relations.

## 2.6 Free Scalar Quantum Field Theory with Special Relativity

When we studied a system of atoms in a crystal, ie sound waves in materials, or ‘Balls and Springs’, we already found a motivation for the study of a system with the action

$$S = \int dt d^3x \left( \frac{1}{2} \dot{\phi}^2 - \frac{1}{2} c^2 (\vec{\partial}_x \phi)^2 \right) \quad (2.6.1)$$

where here we have upgraded to  $3 + 1$  spacetime dimensions. A useful point of view is to think of  $\vec{x}$  as a *label*, as though the field  $\phi(t, x) = \phi_x(t)$  is just an infinite collection of coordinates. This was exactly how we obtained  $\phi(t, x)$  in the case of balls and springs.

When we studied balls and springs, we were even able to understand why this Lagrangian is the leading term at long distances – namely because other possible potentials/interactions that we might add to the Lagrangian are small when the spacing between atoms is small compared to the momenta of interest.

Now that we’ve reviewed special relativity, we can view  $c$  as the speed of light, instead of the speed of sound, and set  $c = 1$ , to find the same Lagrangian

$$S = \int dt d^3x \left( \frac{1}{2} \partial_\mu \phi \partial^\mu \phi \right) \quad (2.6.2)$$

Thus we have a new, independent motivation for studying this Lagrangian – it is the unique Poincaré invariant action for a scalar field  $\phi$  that’s quadratic in  $\phi$  and only has 2 derivatives.

When we discussed balls and springs, we viewed  $\frac{m}{2} \dot{\phi}^2$  as the kinetic energy of the atoms, and  $\frac{1}{2} c^2 (\partial_x \phi)^2$  as the potential energy stored in the springs (bonds between atoms). In relativistic QFT, we call the entire  $(\partial_\mu \phi)^2$  the ‘Kinetic Term’. The term

$$\frac{1}{4} F_{\mu\nu}^2 = \frac{1}{2} (\vec{E}^2 + \vec{B}^2) \quad (2.6.3)$$

in electrodynamics is the kinetic term for the electromagnetic field  $A_\mu$ . Roughly speaking, all quadratic terms in the fields are kinetic terms, including the mass term  $\frac{1}{2} m^2 \phi^2$ , which we haven’t talked about yet.

The book makes a big deal about kinetic terms, mostly to distinguish them from interaction terms, which have more (than two) powers of the fields. For example, we’ll talk extensively about

$$g\phi^3, \quad \lambda\phi^4 \quad (2.6.4)$$

as interaction terms in scalar field theories. In electrodynamics the interactions come about, roughly speaking, from a coupling of the photon field  $A_\mu$  to electromagnetic currents  $J_\mu$  made from other fields.

The kinetic terms are motivated by the fact that they describe freely propagating particles, which is definitely the right start for both particle physics and for the physics of ‘quasi-particles’ like phonons. The point is that we will view the kinetic term, as the part of the theory that we can solve exactly, while the interaction terms are treated in perturbation theory.

You might be wondering what we should choose for the interactions, in general, and why something as simple as  $g\phi^3$  would naturally appear, and not e.g.  $\log \cos(\phi^2)$  or some such. We’ll be able answer such questions soon enough.

### 2.6.1 Canonical Quantization of the Free Scalar

Let us finally apply what we learned about Canonical quantization in order to study a free scalar field in  $3 + 1$  spacetime dimensions, which has an action

$$S = \int dt d^3x \left( \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 \right) \quad (2.6.5)$$

This is just the continuum limit of our ‘Balls and Springs’ action, except that we have taken  $c = 1 =$  the speed of light, we are in  $3 + 1$  dimensions, and we have added a potential term,  $V(\phi) = \frac{1}{2} m^2 \phi^2$ .

The Euler-Lagrange equation is

$$\square \phi = m^2 \phi \quad (2.6.6)$$

where  $\square \equiv \partial_\mu \partial^\mu$ , which has solutions

$$f_p(t, \vec{x}) = e^{ip_\mu x^\mu} \quad \text{where} \quad p^2 = \omega_p^2 - \vec{p}^2 = m^2 \quad (2.6.7)$$

So the addition of the  $\frac{1}{2} m^2 \phi^2$  potential led us to the dispersion relation for a *massive relativistic particle*, with the mass set by  $m$ . Again we have an infinite number of harmonic oscillators labeled by spatial momenta  $\vec{p}$ .

The canonical ‘ $q$ ’ variable is just  $\phi(t, x)$  where we can think of  $x$  as a continuous label. Its canonical conjugate is

$$\pi(t, x) = \frac{\delta L}{\delta \dot{\phi}(t, x)} = \dot{\phi}(t, x) \quad (2.6.8)$$

The canonical commutation relations

$$[\phi(t, \vec{x}), \pi(t, \vec{y})] = i \delta^3(\vec{x} - \vec{y}) \quad (2.6.9)$$

are the continuum limit of what we had in the balls and springs case, where a Kronecker delta on atom labels has become a delta function on spatial positions. We must also demand that  $\phi(t, \vec{x})$  satisfies its equations of motion, as above. This means that  $\phi$  must be a linear combination of the functions  $f_p(t, \vec{x})$ .

We can find such a linear combination of  $f_p$  satisfying the canonical commutation relations by introducing

$$[a_k, a_p^\dagger] = (2\pi)^3 \delta^3(\vec{p} - \vec{k}) \quad (2.6.10)$$

These operators create and destroy particles, and describe a Fock space of free particle states. We usually normalize

$$a_p^\dagger = \frac{1}{\sqrt{2\omega_p}} |\vec{p}\rangle \quad (2.6.11)$$

where the  $|\vec{p}\rangle$  is a single-particle state.

The field  $\phi_0$  at a fixed time  $t_0$  is

$$\phi(t_0, \vec{x}) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (a_p e^{i\vec{p}\cdot\vec{x}} + a_p^\dagger e^{-i\vec{p}\cdot\vec{x}}) \quad (2.6.12)$$

In QFT we usually work in the *Heisenberg Picture*, where operators depend explicitly on time and the state is time independent. So the operator  $\phi(t, x)$  should depend on time. In fact the full result is just

$$\phi(t, \vec{x}) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (a_p e^{-ip\cdot x} + a_p^\dagger e^{ip\cdot x}) \quad (2.6.13)$$

where we promoted  $\vec{p} \cdot \vec{x} \rightarrow p_\mu x^\mu$ . Note that this field obeys the equations of motion, which just say that  $p^2 = m^2$ . We also obtain

$$\pi(t, \vec{x}) = -i \int \frac{d^3p}{(2\pi)^3} \sqrt{\frac{\omega_p}{2}} (a_p e^{-ip\cdot x} - a_p^\dagger e^{ip\cdot x}) \quad (2.6.14)$$

from the time derivative. Note that the integrands are just what we'd expect for a harmonic oscillator for the  $p$ -mode.

Let's check the commutation relations. First, note that

$$\begin{aligned} [\phi(t, \vec{x}), \phi(t, \vec{y})] &= \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{4\omega_p\omega_q}} [(a_p e^{i\vec{p}\cdot\vec{x}} + a_p^\dagger e^{-i\vec{p}\cdot\vec{x}}), (a_q e^{i\vec{q}\cdot\vec{y}} + a_q^\dagger e^{-i\vec{q}\cdot\vec{y}})] \\ &= \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{4\omega_p\omega_q}} (e^{i\vec{p}\cdot\vec{x} - i\vec{q}\cdot\vec{y}} [a_p, a_q^\dagger] + e^{-i\vec{p}\cdot\vec{x} + i\vec{q}\cdot\vec{y}} [a_p^\dagger, a_q]) \end{aligned} \quad (2.6.15)$$

But the commutator of the  $a_k$  gives  $(2\pi)^3 \delta^3(\vec{p} - \vec{k})$ , so we have

$$[\phi(t, \vec{x}), \phi(t, \vec{y})] = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} (e^{i\vec{p}(\vec{x}-\vec{y})} - e^{-i\vec{p}(\vec{x}-\vec{y})}) = 0 \quad (2.6.16)$$

Now note that the region of integration, the measure, and  $\omega_p = \sqrt{m^2 + \vec{p}^2}$  are symmetric under  $\vec{p} \rightarrow -\vec{p}$ , so this commutator vanishes, as claimed.

Next let's compute

$$\begin{aligned} [\phi(t, \vec{x}), \pi(t, \vec{y})] &= -i \int \frac{d^3q}{(2\pi)^3} \int \frac{d^3p}{(2\pi)^3} \frac{1}{2} \sqrt{\frac{\omega_p}{\omega_q}} (e^{-i\vec{q}\cdot\vec{x} + i\vec{p}\cdot\vec{y}} [a_q^\dagger, a_p] - e^{i\vec{q}\cdot\vec{x} - i\vec{p}\cdot\vec{y}} [a_q, a_p^\dagger]) \\ &= \frac{i}{2} \int \frac{d^3p}{(2\pi)^3} (e^{i\vec{p}(\vec{x}-\vec{y})} + e^{-i\vec{p}(\vec{x}-\vec{y})}) \\ &= i\delta^3(\vec{x} - \vec{y}) \end{aligned} \quad (2.6.17)$$

So we have verified that the canonical commutation relation holds.

You should check on your own that

$$H = \int \frac{d^3p}{(2\pi)^3} \omega_p \left( a_p^\dagger a_p + \frac{1}{2} \right) \quad (2.6.18)$$

and

$$[H, \phi(t, \vec{x})] = -i\partial_t \phi(t, \vec{x}) \quad (2.6.19)$$

by using the commutation relations.

Finally, let's re-derive the familiar Schrodinger equation for our free particles. We get back 1-particle QM by restricting to 1-particle states. A position space basis for these is given by

$$\langle x| = \langle 0|\phi(t, \vec{x}) \quad (2.6.20)$$

which follows because  $\phi$  creates or destroys one single particle. The Schrodinger picture wavefunction is

$$\psi(t, x) = \langle x|\psi \rangle = \langle 0|\phi(t, \vec{x})|\psi \rangle \quad (2.6.21)$$

for some generic 1-particle state  $|\psi\rangle$ . Since  $\phi(t, \vec{x})$  satisfies

$$-\partial_t^2 \phi = \left( m^2 - \vec{\nabla}^2 \right) \phi \quad (2.6.22)$$

we have that

$$\begin{aligned} i\partial_t \psi(t, x) &= \sqrt{m^2 - \vec{\nabla}^2} \psi(t, x) \\ &\approx m\psi(t, x) - \frac{\vec{\nabla}^2}{2m} \psi(t, x) \end{aligned} \quad (2.6.23)$$

which is the non-relativistic Schrodinger equation, once we remove the constant energy  $mc^2$ .

### 2.6.2 Noether's Theorem in Field Theory

Everything we said about Noether's theorem applies when we study field theory, where the  $q_i \rightarrow \phi(x)$  are continuous and 'labeled' by the space(-time) position  $x$ . But there's more – not only do we obtain conserved charges, but in field theories we also have *currents*  $J_\mu(x)$  that are conserved at every point in space and time.

Let's look at an example, with Lagrangian

$$L = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi \quad (2.6.24)$$

where in this case  $\phi$  is a complex field. We could, of course, decompose it as  $\phi(x) = \frac{a(x) + ib(x)}{\sqrt{2}}$  in terms of real fields, to give a Lagrangian

$$L = \frac{1}{2}(\partial_\mu a)^2 + \frac{1}{2}(\partial_\mu b)^2 - \frac{m^2}{2}(a^2 + b^2) \quad (2.6.25)$$

In either case the Lagrangian has a symmetry

$$\phi \rightarrow e^{-i\alpha} \phi \quad (2.6.26)$$

which we could write in infinitesimal form as

$$\phi \rightarrow \phi - i\alpha\phi, \quad \phi^\dagger \rightarrow \phi^\dagger + i\alpha\phi^\dagger \quad (2.6.27)$$

or alternatively

$$a \rightarrow a + \alpha b, \quad b \rightarrow b - \alpha a \quad (2.6.28)$$

in components. This is a  $U(1) = SO(2)$  symmetry; it is the symmetry associated with electric charge. It is also a kind of continuum version of the x-y plane symmetry we found in the ordinary mechanical example.

Our derivation of Noether's theorem would lead us to the conserved quantity (global charge)

$$Q = \sum_n \left[ \frac{\partial L}{\partial \dot{\phi}_n} (-i\phi_n) + \frac{\partial L}{\partial \dot{\phi}_n^\dagger} (i\phi_n^\dagger) \right] \quad (2.6.29)$$

$$= -i \int d^3x \frac{\partial L}{\partial \dot{\phi}(x)} \phi(x) + \frac{\partial L}{\partial \dot{\phi}^\dagger(x)} (i\phi^\dagger(x)) \quad (2.6.30)$$

$$= i \int d^3x \left( \dot{\phi}(x) \phi^\dagger(x) - \dot{\phi}^\dagger(x) \phi(x) \right) \quad (2.6.31)$$

this a single conserved charge that cannot change throughout the whole universe.

However, in field theory we can do better than this – we can derive a *conserved current*  $J_\mu(x)$ . This is important because... who cares if there's one single number that's conserved in the universe! How could one ever derive any useful, local effects from such an object, if a loss of charge on earth can be compensated by more charge in another galaxy? Current conservation is the physically useful idea that replaces global charge conservation in field theory.

A cute way to derive this is to imagine that the parameter  $\alpha \rightarrow \alpha(x)$ . There are two steps. The idea is that when  $\alpha$  is constant, the action is invariant under the symmetry (even when the fields do not satisfy their EoM!), so if we perform the transformation with  $\alpha(x)$  a function of  $x$ , we find a shift in the action

$$\delta S = \int d^4x \partial^\mu \alpha(x) i \left( \partial_\mu \phi(x) \phi^\dagger(x) - \partial_\mu \phi^\dagger(x) \phi(x) \right) \quad (2.6.32)$$

The point is that the shift of the action is proportional to derivatives of  $\alpha$ , because the variation must vanish when  $\alpha$  is a constant. Here we have shown our example above, although the idea is completely general. Now if we integrate by parts (remembering that fields vanish rapidly at infinity), we find

$$\delta S = - \int d^4x \alpha(x) \partial^\mu J_\mu \quad (2.6.33)$$



Now we derived this equation by varying the Lagrangian, without imposing the EoM. But when the fields do satisfy their equation of motion, the action must be invariant under *any variation*, even when  $\alpha(x)$  is a general function of  $x$ .

Thus we find the current

$$J_\mu(x) = i (\partial_\mu \phi(x) \phi^\dagger(x) - \partial_\mu \phi^\dagger(x) \phi(x)) \quad (2.6.34)$$

must be conserved. In fact we can check directly that *when the EoM are satisfied* this current is conserved, because

$$\partial^\mu J_\mu = i (\square \phi(x) \phi^\dagger(x) + \partial_\mu \phi(x) \partial^\mu \phi^\dagger(x) - \partial_\mu \phi^\dagger(x) \partial^\mu \phi(x) - \square \phi^\dagger(x) \phi(x)) \quad (2.6.35)$$

$$= im^2(\phi^\dagger \phi - \phi^\dagger \phi) = 0 \quad (2.6.36)$$

using the equations of motion for  $\phi$ . The current is not conserved unless we use the EoM, or to say it more physically, charge conservation is a property of physical states and their time evolution. Fields that don't obey the EoM are just arbitrary, physically meaningless functions.

The conserved current from a more general symmetry transformation

$$\phi \rightarrow \phi + \alpha F(\phi) \quad (2.6.37)$$

will be

$$J_\mu = \frac{\delta L}{\delta(\partial_\mu \phi)} F(\phi) \quad (2.6.38)$$

and we can check directly that this will be conserved when the Lagrangian is invariant under the symmetry transformation and the fields satisfy their equations of motion.

The charge  $Q$  that we found before was just  $\int d^3x J_0$ . Note that for any conserved current

$$\partial_t Q = \int d^3x \partial_t J_0 = \int d^3x \vec{\nabla} \cdot \vec{J} = 0 \quad (2.6.39)$$

and this is how we see global charge conservation.

There's a particularly important set of four conserved currents that exist in any Poincaré invariant field theory; together they form the *energy-momentum tensor*

$$T_{\mu\nu} = \frac{\partial L}{\partial(\partial_\mu \phi)} \partial_\nu \phi - \eta_{\mu\nu} L \quad (2.6.40)$$

Note that  $T_{00}$  is just the Hamiltonian of the theory, since

$$T_{00} = \frac{\partial L}{\partial(\dot{\phi})} \dot{\phi} - \eta_{00} L = \pi \dot{\phi} - L = H \quad (2.6.41)$$

As promised, we have seen that the Hamiltonian is just a component of a tensor, so it is neither invariant nor covariant under Lorentz transformations. This is why we start by working with

Lagrangians. The  $T_{\mu\nu}$  are the (four) currents that correspond to the symmetry under *spacetime translations*, where

$$\phi \rightarrow \phi + \epsilon^\nu \partial_\nu \phi \quad (2.6.42)$$

This is why we have 4 currents that transform together as a tensor – because the infinitesimal transformation parameter  $\epsilon^\nu$  is itself a 4-vector, meaning that Lorentz transformations connect translations in different directions.

### 2.6.3 Aside: What's a Current?

Currents  $J_\mu$  can be Noether currents associated with symmetries, as we saw above, physical background currents whose dynamics we don't analyze (e.g. the sort of object that emitted or absorbed particles in our discussion of anti-particles and special relativity), or formal (mathematical) placeholders that we use to streamline discussions. Basically, currents are like extra fields that do not have kinetic terms. The terminology is used very frequently, and in all of these ways.

## 2.7 Dimensional Analysis, or Which Interactions Are Important?

Before studying field theory in detail, let's consider something of direct and far-reaching importance – dimensional analysis. By studying dimensional analysis and thinking about the short-distance origins of long-distance/low-energy *Effective Field Theories*, we can begin to explain the simplicity and universality of QFT.

The first question we want to ask is what are the units of the Action and Lagrangian Density? Recall we are setting  $\hbar = c = 1$ , so

$$[time] = [length] = [energy]^{-1} = [momentum]^{-1} \quad (2.7.1)$$

where we use  $[X]$  to denote the units of  $X$ . Now we have that

$$S = \int d^d x L \quad (2.7.2)$$

where in our case  $d = 3 + 1$  spacetime dimensions. It turns out that the action  $S$  is dimensionless (in particular, it has units of  $\hbar$ ). There are various ways of seeing this, but probably the simplest and most universal is to note that the Hamiltonian is

$$H = \int d^{d-1} x \left( \frac{\delta L}{\delta \dot{\phi}} \dot{\phi} - L \right) \quad (2.7.3)$$

and of course the Hamiltonian has units of energy. This means that

$$[energy] = \left[ \int d^{d-1} x L \right] = [length]^{d-1} [L] \quad (2.7.4)$$

which implies that the Lagrangian density has units

$$[L] = [energy]^d = [length]^{-d} \quad (2.7.5)$$

Note that the implication follows from the fact that we can only sensibly add quantities with identical units. This will be our starting point.

Next let's consider the dimensions of our quantum fields. The free action for a scalar field is

$$S = \frac{1}{2} \int d^d x ((\partial_\mu \phi)^2 - m^2 \phi^2) \quad (2.7.6)$$

We saw this already in the continuum limit of our description of phonons. The action must be dimensionless, so let's first look at the derivative term. It is forced to have units

$$[(\partial_\mu \phi)^2] = \frac{1}{[d^d x]} = [energy]^d \quad (2.7.7)$$

which implies that

$$[\phi] = [energy]^{\frac{d-2}{2}} \quad (2.7.8)$$

because  $[\partial] = [energy]$ . We also must have

$$[m] = [energy] \quad (2.7.9)$$

so that the mass term has the same units as the kinetic term. This is of course correct – since  $c = 1$ , the mass must have units of energy!

Now let's consider what happens if we add an interaction term to the Lagrangian. We already know the dimensions that  $[\phi]$  must have, so we can deduce that since the term in the action

$$\int d^d x \lambda_n \phi^n \quad (2.7.10)$$

must be dimensionless, we must have

$$[length]^d \times [\lambda_n] \times [energy]^{n \frac{d-2}{2}} = 1 \quad (2.7.11)$$

so that

$$[\lambda_n] = [energy]^{d+n(1-\frac{d}{2})} \quad (2.7.12)$$

Let's plug in some numbers; for  $d = 4$  we have  $[\lambda_n] = [energy]^{n-4}$ , and for  $d = 6$  we have  $[\lambda_n] = [energy]^{2n-6}$ . Notice that for  $d = 4$ , a  $\lambda_4 \phi^4$  interaction is dimensionless, while for  $d = 6$  the  $\lambda_3 \phi^3$  interaction is dimensionless. All higher powers of  $\phi$  have a dimension which goes as an inverse power of energy.

Finally, let's look at interactions that involve derivatives. For example, what about

$$\int d^d x g (\partial \phi)^4 \quad (2.7.13)$$

This gives us

$$[g] = [energy]^{-d} \quad (2.7.14)$$

immediately, since this term is just the square the kinetic term. So the coupling constants associated with derivative interactions like this always have dimensions which are a negative power of energy.

So why is this exciting and important?

Let's imagine that our QFT is the long-distance, low-energy description of some system with a short distance cutoff  $a$ , and an equivalent high-energy cutoff

$$\Lambda \quad (2.7.15)$$

An example, of course, is our phonon or balls and springs system from earlier – there the length  $a$  was the lattice spacing, and the energy  $\Lambda \approx \frac{\hbar c}{a}$  was  $\hbar \omega_{max}$ . The point is that up to some  $\mathcal{O}(1)$  numbers, and barring accidents, miraculous cancellations, or symmetries, *all energies appearing in the QFT Lagrangian will be  $\Lambda$ , simply by dimensional analysis.*

To be specific, this means that in four dimensions, we will have an action

$$S = \int d^4 x \left( \frac{1}{2} (\partial_\mu \phi)^2 - \frac{1}{2} c_m \Lambda^2 \phi^2 - c_3 \Lambda \phi^3 - c_4 \phi^4 - \frac{c_5}{\Lambda} \phi^5 - \dots - \frac{g_3}{\Lambda} \phi (\partial_\mu \phi) - \frac{g_4}{\Lambda^2} \phi^2 (\partial_\mu \phi) - \dots \right)$$

Finally, let us consider conducting an experiment on this system at energies  $E \ll \Lambda$ , corresponding to distance  $R \gg a$ . Experimental results must ultimately be phrased in terms of dimensionless quantities. Note that all terms beyond the first four are proportional to powers of  $1/\Lambda$ . This means that the energy scale of the experiment must make up for these negative powers of  $\Lambda$ , or in other words, *all terms beyond the first four contribute to experimental observables in a way that is suppressed by a power of*

$$\left( \frac{E_{exp}}{\Lambda} \right) \ll 1 \quad (2.7.16)$$

This means that at long distances and low-energies, to a good approximation we can ignore all of the terms with negative powers of  $\Lambda$ ! Thus when we study a single scalar quantum field in a Lorentz invariant universe, there is a universal theory with action

$$S_{low-energy} \approx \int d^4 x \left( \frac{1}{2} (\partial_\mu \phi)^2 - \frac{1}{2} c_m \Lambda^2 \phi^2 - c_3 \Lambda \phi^3 - c_4 \phi^4 \right)$$

There are only four possible terms, a kinetic term, a mass term, and two interactions. Those are all of the possibilities! Although the Higgs field interacts with other fields in the Standard Model, when we isolate it, this is its full Lagrangian (in fact it's even simpler, basically because  $\hbar \rightarrow -\hbar$  is

a symmetry, so  $c_3 = 0$ ). The *hierarchy problem* is the fact that if  $\Lambda$  is really big (say at the Planck scale,  $10^{19} \text{ GeV}$ ) then we must have  $c_m \lll 1$  so that the Higgs has a 126 GeV mass.

We will refine some of these statements when we learn about renormalization and renormalization flows, which incorporate effects from quantum mechanics, and you'll learn to appreciate them more when we do a larger variety of (quantum) computations, but the basic ideas here are robustly true and extremely useful for understanding why QFTs are so simple and universal when they are regarded as *Effective Field Theories*, or long-distance, low-energy descriptions of physics.

## 2.8 Interactions in Classical Field Theory with a View Towards QFT

### 2.8.1 Coulomb's Law in Our Formalism

We will mostly be discussing scalar fields, but to make it clear where we'll get eventually, it's nice to see how electrodynamics works as a field theory. We have a Lagrangian density

$$L = -\frac{1}{4}F_{\mu\nu}^2 - A_\mu J^\mu \quad (2.8.1)$$

where  $J^\mu$  is the electromagnetic current. For static charges  $J_0(x) = \rho(x)$  is the charge density, while  $\vec{J} = 0$ . We can write this as

$$L = -\frac{1}{2}(\partial_\mu A_\nu)^2 + \frac{1}{2}(\partial_\mu A^\nu)^2 - A_\mu J^\mu \quad (2.8.2)$$

Now we need to vary we respect to  $A_\mu$ . The book does this formally via the Euler-Lagrange equation, more informally note that

$$0 = \delta S = \int d^4x [-(\partial_\mu A_\nu)(\partial^\mu \delta A^\nu) + (\partial_\mu A^\mu)(\partial_\nu \delta A^\nu) - J^\nu \delta A_\nu] \quad (2.8.3)$$

$$= \int d^4x [(\partial^\mu \partial_\mu A_\nu) \delta A^\nu - (\partial_\nu \partial_\mu A^\mu) \delta A^\nu - J^\nu \delta A_\nu] \quad (2.8.4)$$

so we find the equation of motion

$$\partial^\mu F_{\mu\nu} = \partial^\mu \partial_\mu A_\nu - \partial_\nu \partial^\mu A_\mu = J_\nu \quad (2.8.5)$$

which are Maxwell's equations in gauge invariant, manifestly relativistic form. This already looks like (where  $\square \equiv \partial_\mu \partial^\mu$ )

$$\square A_\nu - \partial_\nu \partial^\mu A_\mu = J_\nu \quad (2.8.6)$$

so if we choose Lorenz gauge,  $\partial_\mu A^\mu = 0$ , then we get the simple equation

$$\square A_\nu = J_\nu \quad (2.8.7)$$

The necessity of choosing a gauge is half of what makes electrodynamics more complicated than scalar theories (the other half is the existence of spin and polarization); it's not a big deal in the classical theory, but it becomes more of a problem in QFT.

Note that if  $J_\nu = 0$ , this is identical to the equation of motion for a free massless scalar (a phonon),  $\square\phi = 0$ , except it has 4 components. This is the Klein-Gordon equation. It has a formal solution

$$A_\nu = \frac{1}{\square} J_\nu \quad (2.8.8)$$

in the presence of a source, which we can define via a Fourier transform. We refer to  $\frac{1}{\square}$  as a ‘Propagator’; we’ll be using propagators extensively throughout QFT.

Note that for Fourier transforms

$$\square^n f(x) = \int \frac{d^4k}{(2\pi)^4} \square^n \tilde{f}(k) e^{ik_\mu x^\mu} = \int \frac{d^4k}{(2\pi)^4} (-k^2)^n \tilde{f}(k) e^{ik_\mu x^\mu} \quad (2.8.9)$$

and so in general, we identify

$$\square \rightarrow (-k^2) \quad (2.8.10)$$

when we take Fourier transforms. Thus  $\frac{1}{\square} \sim \frac{1}{k^2}$ ; really the only issue is what to do when  $k^2 = 0$ .

Now we can use all of this to re-derive the  $A_\mu$  around a static charge, where

$$J_0 = e\delta^3(\vec{x}) \quad (2.8.11)$$

and  $\vec{J} = 0$ . In this case we have

$$A_0 = \frac{e}{\square} \delta^3(\vec{x}) \quad (2.8.12)$$

$$= \int \frac{d^3k}{(2\pi)^3} \frac{e}{k^2} e^{i\vec{k}\cdot\vec{x}} \quad (2.8.13)$$

$$= \frac{e}{(2\pi)^3} \int_0^\infty k^2 dk \int_{-1}^1 d\cos\theta \int_0^{2\pi} d\phi \frac{1}{k^2} e^{ikr\cos\theta} \quad (2.8.14)$$

$$= \frac{e}{8\pi^2} \frac{1}{ir} \int_{-\infty}^\infty dk \frac{e^{ikr} - e^{-ikr}}{k} \quad (2.8.15)$$

There’s no divergence at  $k = 0$  because the exponential factors cancel. Thus we can perform the integral as

$$\int_{-\infty}^\infty dk \frac{e^{ikr} - e^{-ikr}}{k + i\epsilon} \quad (2.8.16)$$

and use contour integration. For the first term we must close the contour in the upper half plane to get zero, while for the second term we get

$$\int dk \frac{-e^{-ikr}}{k + i\epsilon} = -(2\pi i)(-e^{ikr}) = 2\pi i \quad (2.8.17)$$

as  $\epsilon \rightarrow 0$ . So we find

$$A_0 = \frac{e}{4\pi} \frac{1}{r} \quad (2.8.18)$$

as expected for the Coulomb potential.

### 2.8.2 Green's Functions

Let us generalize our analysis above to study more complicated classical fields.

A toy model for gravity has the Lagrangian density

$$L = \frac{1}{2}(\partial h)^2 + \frac{1}{3}\lambda h^3 + Jh \quad (2.8.19)$$

In reality  $h$  is a tensor field, but we are pretending that it's a scalar. The parameter  $\lambda \sim \sqrt{G_N}$  plays the role of the Newton constant – note that the gravitational field is self-interacting, unlike the electromagnetic field. Of course the kinetic term is just what we had for a free scalar field, or for the phonon field we discussed before.

The equation of motion is

$$\square h = J + \lambda h^2 \quad (2.8.20)$$

The idea is to solve while incorporating the effects of both  $J$  and the interaction term  $\lambda h^2$  in the EoM. We will assume that  $\lambda$  is small (we'll discuss more what this means later on), so we can incorporate its effects in perturbation theory. Thus we assume that

$$h = h_0 + h_1 + h_2 + \dots \quad (2.8.21)$$

where the  $n$ th term includes  $n$ th order effects from  $\lambda$ . To zeroth order this means

$$\square h_0 = J \quad (2.8.22)$$

while to first order it implies

$$\square(h_0 + h_1) = J + \lambda(h_0 + h_1)^2 \quad (2.8.23)$$

$$\implies \square h_1 = \lambda h_0^2 \quad (2.8.24)$$

Thus  $h_0^2$  serves as a new source or 'current' for  $h_1$ . This has a solution

$$h_0 = \frac{1}{\square} J \quad (2.8.25)$$

$$h_1 = \frac{\lambda}{\square} h_0^2 = \frac{\lambda}{\square} \left[ \left( \frac{1}{\square} J \right) \left( \frac{1}{\square} J \right) \right] \quad (2.8.26)$$

and it can be continued to get a series solution for  $h$ .

It is worth noting that the propagator, and therefore the solution, depends only on the propagator  $\square$ , which comes from the kinetic term. This is, of course, because we are perturbing about a solution to the free theory.

We can be more precise by defining  $1/\square$  as

$$\square_x \Pi(x, y) = -\delta^4(x - y) \quad (2.8.27)$$

so that

$$\Pi(x, y) = \int \frac{d^4 k}{(2\pi)^4} e^{ik(x-y)} \frac{1}{k^2} \quad (2.8.28)$$

Note that really  $\Pi(x, y) = \Pi(y, x) = \Pi(x - y)$ .

Now  $\Pi$  allows us to write

$$h_0(x) = - \int d^4 y \Pi(x, y) J(y) \quad (2.8.29)$$

as we can check immediately that this satisfies the equation of motion. Thus our more general solution is

$$\begin{aligned} h(x) = & - \int d^4 y \Pi(x, y) J(y) \\ & + \lambda \int d^4 w \Pi(x, w) \int d^4 y \Pi(w, y) J(y) \int d^4 z \Pi(w, z) J(z) + \dots \end{aligned} \quad (2.8.30)$$

which is what we meant earlier when we wrote a formal solution in terms of  $1/\square$ .

This can be represented by pictures where the propagator  $\Pi(x, y)$  looks like a line, while the currents  $J$  look like insertions that come together at points such as  $x$  and  $w$ . These are our first example of *Feynman diagrams*. The rules for associating mathematical expressions with the pictures are called *Feynman rules*. These diagrams give some physical intuition, and they also allow us to generate all the mathematical expressions allowed to any order in  $\lambda$ , via the rules

1. Draw a point  $x$  and a line from  $x$  to a new point  $x_i$ .
2. Either truncate a line at a source  $J$  or let the line branch into two lines, adding a new point a factor of  $\lambda$ .
3. Repeat the previous until all lines truncate at sources.
4. The final value for  $h(x)$  is given by summing up graphs with lines associated with propagators, internal points integrated over, and all points external points except  $x$  associated with  $J$ .

One can solve the EoM for a classical field by drawing these pictures. When we move to QFT, the main difference will be that lines can loop back on themselves.

## 2.9 Overview of Scattering and Perturbation Theory

We learned about symmetries, canonical quantization, and the quantization of a free quantum scalar field, which describes a Fock space of free relativistic particles. We saw examples in classical field theory where we obtained a Feynman diagram perturbation series that solves the classical field equations, and we studied dimensional analysis to see what interactions are important at long distances.



Now we would like to develop the perturbative description of QFT. What sort of processes should we study?

For a variety of reasons we will study *Scattering* in this course. If this were primarily a condensed matter physics course we might study other questions... because the best motivation for studying scattering is that if we are going to do ‘particle physics’ or ‘high-energy physics’, the whole point is to pursue reductionism to the extreme. We would like to ‘see what stuff is made of’. To do that we need a microscope. But due to the uncertainty principle

$$\delta x \cdot \delta p \geq \frac{\hbar}{2} \quad (2.9.1)$$

we cannot look at matter at very short distances, or ‘take it apart’, without very large momenta, and thus very large energies (hence the name high-energy physics). But if we study matter using very high energy probes, we cannot help but destroy it, so to be as precise as possible, it’s easiest to just study what happens when scatter a few particles in isolation.

Obviously this has a precedent, e.g. Rutherford trying to determine the size of an atomic nucleus by looking at  $\alpha$  particles scattering off of gold nuclei in gold foil... all the way to the LHC, today, at 14 TeV... we would like to be able to perform the theoretical computations necessary to understand these results.

Thus we will be primarily interested in computing an observable called the *S-Matrix*. We will usually think of QFT in the Heisenberg picture, where states are time-independent, but operators depend on time. In this language the S-Matrix is

$$\langle f|S|i\rangle_{Heis} = \langle f, t = \infty|i, t = -\infty\rangle_{Schr} \quad (2.9.2)$$

Formally, the  $S$  operator evolves states from the infinite past to the infinite future via the Hamiltonian. However, it is very useful to think of the interactions as being turned off in the distant past and the distant future.

How can we actually perform these computations?

Conceptually, the easiest way to think about the computation of the S-Matrix is to imagine that in the distant past and future, the interactions have been turned off, so that we just have a free QFT. The interactions slowly turn on and off in between, perturbing our QFT and allowing particles to interact. In fact, this happens automatically when we scatter particles, since they become well-separated in the past and future.

Note that since

$$|\psi_{1-part}\rangle = \phi(t, x)|0\rangle \quad (2.9.3)$$

one might imagine that the S-Matrix could be computed from something very roughly like

$$\langle \phi(\infty, \vec{x}_1)\phi(\infty, \vec{x}_2)\phi(-\infty, \vec{x}_3)\phi(-\infty, \vec{x}_4)\rangle \quad (2.9.4)$$

for the example of 2-to-2 scattering, since two of the  $\phi$ s create particles in the past, and two of the  $\phi$ s destroy or measure particles in the future. In fact, there is a formula like this, called the *LSZ Reduction Formula*, which relates QFT *correlation functions* like

$$\langle \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle \quad (2.9.5)$$

to the S-Matrix. This is one major reason why in QFT, these correlation functions are the observables we most often study – for the rest of the course they will be a primary focus.

The precise statement of LSZ for scalar particles is that the S-Matrix

$$\begin{aligned} \langle p_{k+1} \cdots p_n | S | p_1 \cdots p_k \rangle &= \left[ i \int d^4 x_1 e^{-ip_1 x_1} (\square + m_1^2) \right] \cdots \left[ i \int d^4 x_n e^{ip_n x_n} (\square + m_n^2) \right] \\ &\times \langle 0 | T \{ \phi_1(x_1) \phi_2(x_2) \cdots \phi_n(x_n) \} | 0 \rangle \end{aligned} \quad (2.9.6)$$

where  $T$  is a time-ordering symbol, which just says that we act with the fields  $\phi_i$  in an order determined by their time coordinates  $t_i$ , as we'll discuss soon. In words, this formula says that if we Fourier transform a correlation function and pick off the coefficient of certain singularities of the form  $\frac{1}{p^2 - m^2}$ , we get the S-Matrix. Note that the signs of the  $p_i$  flip for initial vs final states.

So how do we compute these correlation functions? We will work in perturbation theory about a free QFT, and there are basically two different methods:

1. Old-Fashioned Perturbation theory, which we'll discuss briefly, is the QFT generalization of the perturbation theory you learned in QM. In this version of perturbation theory, all states are physical, momentum is conserved, but energy is not conserved. Since this methodology treats energy and momentum differently, it cannot be manifestly Lorentz invariant. This is bad because it makes computations more complicated, so that it's easy to make mistakes, and the physical symmetries of special relativity are obscured.
2. Feynman Diagrams, or standard QFT perturbation theory, is the method that's actually been in use since the 1950s. In this formalism energy and momentum are always conserved at every step, and the associated 'Feynman rules' for calculations are manifestly Lorentz invariant. The price to be paid is that we use 'virtual particles', or 'off-shell' states, which do not satisfy  $p^2 = m^2$ , as we will see. We will give two or three different derivations of this method, the first of which connects very closely to our discussion of classical perturbation theory in field theories (using Green's functions).

In what follows we will relate the S-Matrix to experimental observables, derive the LSZ Reduction formula, and discuss these two forms of perturbation theory, focusing mostly on Feynman Diagrams.

## 2.10 Relating the S-Matrix to Cross Sections and Decay Rates

Let us see how we can relate the formal S-Matrix to physical observables, namely cross sections, decay rates, etc.

Classically, we can talk about scattering in terms of cross sectional area – when we shoot one particle at another, what area is occluded? Quantum mechanically, we can have the same discussion, except phrased in terms of probabilities.

To understand what a cross section will mean, it's helpful to think of putting our experiment in a large box of length  $L$ , and in a 'time box', where we turn interactions on for a time of length  $T$ . Then we want to think in terms of a *flux* of particles, which we define as

$$\Phi = \text{number density} \times \text{beam velocity} \quad (2.10.1)$$

It helps to draw a picture of a little box with some number of particles, all moving with some velocity. Then one can see that the flux measures how many particles are crossing a particular cross sectional area per unit time.

It may be most instructive to think in terms of a cross sectional area  $\sigma$  that occludes the particles. Then we obtain a total probability for scattering

$$P = T\Phi\sigma \quad (2.10.2)$$

In other words, if we run the experiment for a length of time  $T$ , with a flux  $\Phi$ , and the scattering process occludes a surface area  $\sigma$ , then  $P$  will be the total probability that a scattering event occurs. We can now invert this equation to define a differential cross section as

$$d\sigma = \frac{1}{T} \frac{1}{\Phi} dP \quad (2.10.3)$$

When we discuss ‘ $dP$ ’ this is a differential probability for scattering. It is a function of angles, quantum numbers, etc. Let us see how we can compute  $dP$  in quantum mechanics and QFT.

### 2.10.1 Transitions Between States in Quantum Mechanics

In quantum mechanics, one of the most important observables is the probability for transitioning from one state to another:

$$P_{fi} = |\langle\psi_f(t_f)|\psi_i(t_i)\rangle|^2 \quad (2.10.4)$$

For example, we might have a hydrogen atom whose electron is in a  $n = 2$  p-orbital at a time  $t_i$ , and we want to know the probability that at a time  $t_f$  it has emitted a photon and transitioned to the  $n = 1$  s-orbital state.

In QFT this is the S-Matrix, for example

$$\langle p_3, p_4 | S | p_1, p_2 \rangle \quad (2.10.5)$$

Note that energy and momentum must be conserved, so this must be proportional to  $\delta^4(p_3 + p_4 - p_1 - p_2)$ . But there’s another issue, which is that if we view the square of this amplitude as a probability, then it’s zero!

Of course it’s zero for a stupid reason, namely that  $p_3$  and  $p_4$  are continuous variables, so the probability of getting *exactly* these momenta vanishes. Instead, we must view this as a differential probability

$$\frac{dP}{d\Pi} \propto |\langle p_3, p_4 | S | p_1, p_2 \rangle|^2 \quad (2.10.6)$$

where the differential *phase space* is roughly

$$d\Pi \sim d^3p_3 d^3p_4 \quad (2.10.7)$$

The reason is that if we integrate this object over the phase space, or the space of final states, parameterized by  $p_3$  and  $p_4$ , then we get a genuine probability.

How can we make this more precise? Let's imagine that the entire universe is in a box with side lengths  $L$ , which are very, very, very large. (This is always an OK thing to do if you get confused about QFT... you can perform computations in a large box and then send  $L \rightarrow \infty$  at the end; you should get sensible answers if you've asked sensible questions.) Then the momenta will be quantized with

$$\vec{p} = \frac{2\pi}{L} \vec{n} \quad (2.10.8)$$

where  $\vec{n} = (n_x, n_y, n_z)$  is a multiplet of integer labels. This is just the standard particle in a box with Dirichlet boundary conditions, which you studied in quantum mechanics.

Now we need a way to sum over particle states in a normalized way. With particles in a box, we have a discretum of state with  $\vec{n}$  for each particle. But we would like to work in a continuum with general momenta  $\vec{p}$ . This means that we need to convert

$$\sum_{\vec{n}} \rightarrow N \int d^3p \quad (2.10.9)$$

This can be done by assigning a continuum momentum  $\vec{p}$  to a discrete momentum 'bin'. We know that each  $\vec{n}$  sits inside a phase space region with a momentum-space volume

$$\left(\frac{2\pi}{L}\right) \left(\frac{2\pi}{L}\right) \left(\frac{2\pi}{L}\right) = \frac{(2\pi)^3}{V} \quad (2.10.10)$$

where  $V = L^3$ , so we know that

$$\sum_{\vec{n}} = V \int \frac{d^3p}{(2\pi)^3} \quad (2.10.11)$$

and this is the relation we needed. We can use this to define

$$d\Pi_{3,4} \propto \left(V \frac{d^3p_3}{(2\pi)^3}\right) \left(V \frac{d^3p_4}{(2\pi)^3}\right) \quad (2.10.12)$$

which is what we will use in what follows.

### 2.10.2 General Cross Section Formula

Now we will derive a general formula for the differential cross section, combining ideas of flux with the quantum mechanical probability discussed above. Let's focus on processes

$$p_1 + p_2 \rightarrow p_3, \dots, p_k \quad (2.10.13)$$

We will work in a fixed finite volume  $V$ , because that makes it much easier to understand various infinities that will drop out at large  $V$ . Since there's only one particle colliding with another, the flux in the center of mass frame is simply

$$\Phi = \frac{|\vec{v}_1 - \vec{v}_2|}{V} \quad (2.10.14)$$

Using our defining formula for the cross section, we have

$$d\sigma = \frac{V}{T} \frac{1}{|\vec{v}_1 - \vec{v}_2|} dP \quad (2.10.15)$$

Quantum mechanically, the probabilities are

$$dP = \frac{|\langle f|S|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle} d\Pi \quad (2.10.16)$$

where we insert the explicit normalizations because (due to the demands of Lorentz invariance) our states are not normalized to unity. Recall that

$$\langle \vec{p}|\vec{p}\rangle = (2\pi)^3 (2E_p) \delta^3(0) = 2E_p V \quad (2.10.17)$$

where we define  $\delta^3(0)$  via the inverse Fourier transform from momentum space. Note also that

$$\delta^4(0) = \frac{TV}{(2\pi)^4} \quad (2.10.18)$$

where  $T$  was the total time.

The final state phase space is

$$d\Pi = \prod_j \frac{V}{(2\pi)^3} d^3 p_j \quad (2.10.19)$$

for the final state particles. This is defined by taking the continuum limit of particles in a box, which have a discretum of momenta with  $p_i = \frac{2\pi}{L} n_i$  for integers  $\vec{n}$ . We can combine this with the  $\langle f|f\rangle\langle i|i\rangle$  to obtain

$$dP = \frac{|\langle f|S|i\rangle|^2}{4E_1 E_2 V^2} \prod_j \frac{d^3 p_j}{2E_p (2\pi)^3} \quad (2.10.20)$$

Finally, we can write the  $S$  operator as

$$S = \mathbb{1} + i\mathcal{T} \quad (2.10.21)$$

where

$$\mathcal{T} = (2\pi)^4 \delta^4 \left( \sum_i p_i \right) \mathcal{M} \quad (2.10.22)$$

This extracts a 4-momentum conserving delta function, and allows us to bypass considering the ‘trivial’ part of the S-Matrix, where no scatterings occur, to focus on  $\mathcal{M}$ . This means that

$$\begin{aligned} |\langle f|S|i\rangle|^2 &= \delta^4\left(\sum_i p_i\right) (2\pi)^8 \delta^4(0) |\mathcal{M}|^2 \\ &= (2\pi)^4 \delta^4\left(\sum_i p_i\right) TV |\mathcal{M}|^2 \end{aligned} \quad (2.10.23)$$

so we can finally write

$$dP = \frac{T}{V} \frac{1}{2E_1 2E_2} |\mathcal{M}|^2 d\Pi_{LIPS} \quad (2.10.24)$$

where

$$d\Pi_{LIPS} \equiv (2\pi)^4 \delta^4\left(\sum_i p_i\right) \prod_{\text{final } j} \frac{1}{2E_j} \frac{d^3 p_j}{(2\pi)^3} \quad (2.10.25)$$

is Lorentz invariant phase space (you’re supposed to verify that it’s Lorentz invariant on the problem set). So the differential cross section is

$$d\sigma = \frac{|\mathcal{M}|^2}{(2E_1)(2E_2)|\vec{v}_1 - \vec{v}_2|} d\Pi_{LIPS} \quad (2.10.26)$$

Note that the two factors of energy in the denominator lead to units of area; the other factors must be dimensionless after integration.

### 2.10.3 Decay Rate Formula

The generalization to decay rates is immediate, since we just replace the initial 2 particle state with a 1 particle state, leading to

$$d\Gamma = \frac{|\mathcal{M}|^2}{2E_1} d\Pi_{LIPS} \quad (2.10.27)$$

In this case  $|\mathcal{M}|^2$  and  $d\Pi_{LIPS}$  must carry units so that after integration we get a rate.

### 2.10.4 Special Case of 2-to-2 Scattering

By far the most common example of scattering is 2-to-2, which in the center of mass frame has

$$\vec{p}_1 = -\vec{p}_2, \quad \vec{p}_3 = -\vec{p}_4 \quad (2.10.28)$$

and  $E_1 + E_2 = E_3 + E_4 = E_{CM}$ . The phase space is

$$d\Pi_{LIPS} \equiv (2\pi)^4 \delta^4\left(\sum_i p_i\right) \frac{1}{2E_3} \frac{d^3 p_3}{(2\pi)^3} \frac{1}{2E_4} \frac{d^3 p_4}{(2\pi)^3} \quad (2.10.29)$$

We can immediately integrate over and eliminate  $\vec{p}_4$  using the delta function, giving

$$d\Pi_{LIPS} \equiv \frac{1}{16\pi^2} d\Omega \int dp_f \frac{p_f^2}{E_3 E_4} \delta(E_3 + E_4 - E_{CM}) \quad (2.10.30)$$

where  $E_3 = \sqrt{p_f^2 + m_3^2}$  and  $E_4 = \sqrt{p_f^2 + m_4^2}$ . Now we can change variables from  $p_f$  to  $x = E_3 + E_4 - E_{CM}$  with Jacobian

$$\frac{dx}{dp_f} = \frac{p_f}{E_3} + \frac{p_f}{E_4} = \frac{E_{CM}}{E_3 E_4} p_f \quad (2.10.31)$$

so we obtain

$$d\Pi_{LIPS} \equiv \frac{1}{16\pi^2} \frac{p_f}{E_{CM}} d\Omega \quad (2.10.32)$$

We can now use the relation

$$|\vec{v}_1 - \vec{v}_2| = \frac{|\vec{p}_1|}{E_1} + \frac{|\vec{p}_2|}{E_2} = \frac{|\vec{p}_{init}|}{E_1 E_2} E_{CM} \quad (2.10.33)$$

which holds by momentum conservation to derive the final result

$$\left( \frac{d\sigma}{d\Omega} \right)_{CM} = \frac{1}{64\pi^2 E_{CM}^2} \frac{|\vec{p}_f|}{|\vec{p}_i|} |\mathcal{M}|^2 \quad (2.10.34)$$

for the center of mass differential cross section with respect to the solid angle of scattering,  $d\Omega$ . If all the masses are equal the ratio of momenta drops out. The book gives an analysis and comparison with the non-relativistic limit.

## 2.11 Old Fashioned Perturbation Theory

We have discussed free quantum field theory, and interacting classical field theory. Now we want to move on to study interacting QFTs. We will ‘warm up’ by discussing old fashioned perturbation theory, which is essentially the perturbation theory you learned in QM applied to QFT. As we will discuss, this is a bad way to organize things, because it’s not Lorentz invariant, but it’s worth understanding because at the end of the day, it’s entirely equivalent to the methods we’ll use later on, and probably more familiar. There are also some situations where it’s a useful way to think about the physics of QFT, and it provides a more precise illustration of our earlier ‘derivation’ of the existence of anti-particles.

### 2.11.1 OFPT Formalism

In quantum mechanics you studied situations where we have a Hamiltonian

$$H = H_0 + V \quad (2.11.1)$$

where  $V$  is an ‘interaction’ term that can be treated in perturbation theory, with the idea that it makes a small correction to the spectrum of  $H_0$ , which we can solve for exactly. The only difference in QFT will be that the spectrum of  $H_0$  is continuous, since particles can have arbitrary momentum and energy.

From our discussion of scattering, it’s clear that we’d like to be able to solve for what happens at late times in terms of an initial state at an early time. So say we start off in a state

$$H_0|\psi\rangle = E|\psi\rangle \quad (2.11.2)$$

Since the spectrum is continuous, we can find a similar state in the full theory satisfying

$$H|\Psi\rangle = E|\Psi\rangle \quad (2.11.3)$$

for the full  $H = H_0 + V$ . This means we can write (formally)

$$|\Psi\rangle = |\psi\rangle + \frac{1}{E - H_0}V|\Psi\rangle \quad (2.11.4)$$

which can be verified by multiplying by  $E - H_0$ . This simple formula is called the *Lippman-Schwinger equation*. You can think of  $\frac{1}{E - H_0}$  as a Green’s function, or the ‘Lippman-Schwinger kernel’.

We want to use this equation to describe scattering, where the states  $|\psi\rangle$  are free at early and late times, and  $V$  induces transitions (scattering) between these free states.

For this purpose its useful to define an operator  $T$ , the transfer matrix, by

$$T|\psi\rangle = V|\Psi\rangle \quad (2.11.5)$$

Plugging this into the Lippman-Schwinger equation gives

$$|\Psi\rangle = |\psi\rangle + \frac{1}{E - H_0}T|\psi\rangle \quad (2.11.6)$$

which formally defines the full interacting state in terms of the free states.

To make this equation useful, let’s take the inner product of both sides with  $\langle\psi_\alpha|V$ , where  $\langle\psi_\alpha|$  is any free state. We find

$$\begin{aligned} \langle\psi_\alpha|V|\psi\rangle + \langle\psi_\alpha|V\frac{1}{E - H_0}T|\psi\rangle &= \langle\psi_\alpha|V|\Psi\rangle \\ &= \langle\psi_\alpha|T|\psi\rangle \end{aligned} \quad (2.11.7)$$

If this is true for all  $\psi_\alpha$  and  $\psi$ , then we must have the operator equation

$$T = V + V\frac{1}{E - H_0}T \quad (2.11.8)$$

This provides a sort of recursion relation for  $T$ , which we can solve perturbatively in  $V$  by inserting it back into itself, giving

$$T = V + V\frac{1}{E - H_0}V + V\frac{1}{E - H_0}V\frac{1}{E - H_0}V + \dots \quad (2.11.9)$$



To make this more concrete, we can insert a complete set of free states, which have definite  $H_0$  eigenvalues (energies), so that we obtain

$$T_{ij} = V_{ij} + \sum_k V_{ik} \frac{1}{E - E_k} V_{kj} + \dots \quad (2.11.10)$$

Note that in a certain sense, ‘energy is not conserved’ in this perturbation series, since  $V_{ik}$  connects states with different  $H_0$  energies.

### 2.11.2 Coulomb’s Law from OFPT

Let’s study what happens when electrons (or, more accurately, a toy model thereof with a scalar electron and a scalar photon) scatter off each other using OFPT. An advantage for us is that we’ll see a very concrete manifestation of ‘particle emission’ and ‘particle absorption’, realizing the abstract framework we discussed when we argued for the existence of anti-particles.

Our initial and final states both have two scalar electrons

$$|i\rangle = |\psi_e^1 \psi_e^2\rangle, \quad \langle f| = \langle \psi_e^3 \psi_e^4| \quad (2.11.11)$$

In QFT, electron interactions are generated by photon exchange. The interaction Hamiltonian is

$$V = \frac{1}{2} e \int d^3x \psi_e(x) \psi_e(x) A(x) \quad (2.11.12)$$

where we treat both the electrons  $\psi_e(x)$  and the photon  $A(x)$  as scalar fields (so this is a simplified model... neither are actually scalars). Note that since this is an interaction Hamiltonian it is only integrated over space.

The transfer matrix is

$$T_{fi} = V_{fi} + \sum_n V_{fn} \frac{1}{E_i - E_n} V_{ni} + \dots \quad (2.11.13)$$

However, the first term vanishes, because the interaction simply turns an electron into an electron and photon, or vice versa. In other words, we need two interactions, so that we can both absorb and emit a photon. So there are two possible terms – one where the photon is emitted by electron 1, and one where the photon is emitted by electron 2. This is just as we saw above, except here the abstract current  $J(x)$  has been replaced by the physical electromagnetic current that can emit or absorb photons. (Note that we are ignoring terms where both electrons are destroyed and turn into a photon, because that cannot happen by charge conjugation, although that’s not obvious in this simplified theory.)

For the first case, we find that

$$V_{ni} = e \langle \psi_e^3 \psi_e^2 A | V | \psi_e^1 \psi_e^2 \rangle = e \langle \psi_e^3 A | V | \psi_e^1 \rangle \langle \psi_e^2 | \psi_e^2 \rangle = e \langle \psi_e^3 A | V | \psi_e^1 \rangle \quad (2.11.14)$$

If we invariantly view the first electron as the source, then in this case we use a ‘retarded propagator’, since the photon moves from the source to the sink. In the opposite case we use an ‘advanced propagator’, because the photon moves from the sink to the source. We must add up both terms.

Let's evaluate the matrix elements. Note that

$$V_{ni} = e \langle \psi_e^3 A | V | \psi_e^1 \rangle = \frac{e}{2} \int d^3x \langle \psi_e^3 A | \psi_e(x) \psi_e(x) A(x) | \psi_e^1 \rangle \quad (2.11.15)$$

To evaluate this we must use the quantized form of the fields

$$\phi(t=0, \vec{x}) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^\dagger e^{i\vec{p}\cdot\vec{x}} + a_p e^{-i\vec{p}\cdot\vec{x}}) \quad (2.11.16)$$

so we find, for example, that

$$\langle A | A(x) | 0 \rangle = e^{-i\vec{p}_A \cdot \vec{x}} \quad (2.11.17)$$

where we have chosen the time as  $t_0 = 0$ , and we recall that the one-particle states are defined so that

$$|\vec{p}\rangle = \left( \sqrt{2E_p} \right) a_p^\dagger | 0 \rangle \quad (2.11.18)$$

so the factors of  $\sqrt{2E_p}$  cancel. Actually, this is a bad thing, since when we sum over a complete set of states we should be using states that are unit normalized, or in other words, we should use

$$\frac{|A\rangle\langle A|}{\langle A|A\rangle} \quad (2.11.19)$$

But for now we will just use the convention above. At the end of the computation we'll remember to divide by  $\langle A|A\rangle = 2E_A$ .

Thus we obtain (note there's a 2 from the 2 electron field contraction possibilities)

$$V_{ni} = e \int d^3x e^{i(\vec{p}_1 - \vec{p}_3 - \vec{p}_A) \cdot \vec{x}} = e(2\pi)^3 \delta^3(\vec{p}_1 - \vec{p}_3 - \vec{p}_A) \quad (2.11.20)$$

The matrix element on the other side is nearly identical, and so we obtain

$$T_{fi} = \int d^3\vec{p}_A (2\pi)^6 \delta^3(\vec{p}_1 - \vec{p}_3 - \vec{p}_A) \delta^3(\vec{p}_2 - \vec{p}_4 + \vec{p}_A) \frac{e^2}{E_i - E_n} \quad (2.11.21)$$

The delta functions enforce

$$\vec{p}_1 + \vec{p}_2 = \vec{p}_3 + \vec{p}_4 \quad (2.11.22)$$

as well as

$$\vec{p}_A = \vec{p}_1 - \vec{p}_3 \quad (2.11.23)$$

Thus we see that

$$E_A(\vec{p}_A) = |\vec{p}_1 - \vec{p}_3| \quad (2.11.24)$$

since the photon is massless.

Now we have two cases to consider, the ‘retarded’ and ‘advanced’. In the retarded case the 2nd electron hasn’t changed, but the first has emitted the photon, so it’s in its final state, with energy  $E_3$ . Thus

$$E_n^R = E_2 + E_3 + E_A \quad (2.11.25)$$

In contrast, in the advanced case we have

$$E_n^A = E_1 + E_4 + E_A \quad (2.11.26)$$

This means that we find (stripping off an overall momentum conserving delta function)

$$T_{fi}^R = \frac{e^2}{(E_1 + E_2) - (E_2 + E_3 + E_A)} = \frac{e^2}{E_1 - E_3 - E_A} \quad (2.11.27)$$

and

$$T_{fi}^A = \frac{e^2}{(E_1 + E_2) - (E_1 + E_4 + E_A)} = \frac{e^2}{E_2 - E_4 - E_A} \quad (2.11.28)$$

Overall energy conservation says  $E_1 + E_2 = E_3 + E_4$  so that  $E_1 - E_3 = E_4 - E_2 = \Delta E$ , so we have

$$T_{fi}^R + T_{fi}^A = \frac{2e^2 E_A}{(\Delta E)^2 - (E_A)^2} \quad (2.11.29)$$

That’s the answer we’ve obtained. However, recall that we were supposed to normalize the photon, which means dividing by  $2E_A$ , yielding

$$T_{fi}^R + T_{fi}^A = \frac{e^2}{(\Delta E)^2 - (E_A)^2} \quad (2.11.30)$$

Soon we will define the sum of the retarded and advanced propagators that have appeared here – this is the Feynman propagator.

Now note that if we define the Lorentz 4-vector

$$k^\mu \equiv p_3^\mu - p_1^\mu = (\Delta E, \vec{p}_A) \quad (2.11.31)$$

then the full answer can be written as

$$\frac{e^2}{k^2} \quad (2.11.32)$$

Thus we see the momentum space Green’s function of our scalar photon  $A$  appearing in the answer! If we like we can plug this into the cross section formula to derive the differential cross section for Coulomb scattering.

As a final note, we can use OFPT to unravel some of the mysteries associated with our argument that the existence of anti-particles follows from Lorentz invariance and quantum mechanics. The two different terms we found in OFPT correspond precisely with ‘a positively charged particle propagating forward in time’ and ‘a negatively charged particle propagating back in time’, respectively. In both cases the particle has positive energy  $E = |\vec{p}_1 - \vec{p}_3|$ , but the individual processes in OFPT do not themselves conserve energy! Note that the lack of energy conservation at intermediate stages follows because we have picked special intermediate times for the calculation, effectively breaking time translation invariance. But this time translation breaking is only associated with intermediate steps in the calculation, and overall energy is still conserved.

## 2.12 LSZ Reduction Formula – S-Matrix from Correlation Functions

We saw in the section before last how to compute scattering cross sections from the S-Matrix. Now let’s see how to compute the S-Matrix from QFT correlation functions.

The LSZ reduction formula tells us how to derive the S-Matrix from QFT correlation functions

$$\begin{aligned} \langle p_{k+1} \cdots p_n | S | p_1 \cdots p_k \rangle &= \left[ i \int d^4 x_1 e^{-ip_1 x_1} (\square + m_1^2) \right] \cdots \left[ i \int d^4 x_n e^{ip_n x_n} (\square + m_n^2) \right] \\ &\times \langle 0 | T \{ \phi_1(x_1) \phi_2(x_2) \cdots \phi_n(x_n) \} | 0 \rangle \end{aligned} \quad (2.12.1)$$

This says that if we write time-ordered correlators in momentum space, the simultaneous coefficient of the poles  $\frac{1}{p_i^2 - m_i^2}$  is the S-Matrix.

### 2.12.1 Deriving LSZ

In slightly more detail, we can write the initial and final states as

$$|i\rangle = \sqrt{2E_1} \sqrt{2E_2} a_{p_1}^\dagger(-\infty) a_{p_2}^\dagger(-\infty) |\Omega\rangle \quad (2.12.2)$$

and

$$|f\rangle = \sqrt{2E_3} \cdots \sqrt{2E_k} a_{p_3}^\dagger(\infty) \cdots a_{p_k}^\dagger(\infty) |\Omega\rangle \quad (2.12.3)$$

so we have

$$\langle f | S | i \rangle = 2^{\frac{k}{2}} \sqrt{E_1 \cdots E_k} \langle \Omega | a_{p_k}(\infty) \cdots a_{p_3}(\infty) a_{p_1}^\dagger(-\infty) a_{p_2}^\dagger(-\infty) | \Omega \rangle \quad (2.12.4)$$

where the important point is that the  $a_p(\pm\infty)$  operators create and destroy particles in the future and past.

We quantized a free scalar field

$$\phi(t, \vec{x}) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (a_p e^{-ip \cdot x} + a_p^\dagger e^{ip \cdot x}) \quad (2.12.5)$$

What about the interacting case? Formally, since

$$[\phi(t, \vec{x}), \pi(t, \vec{y})] = i\delta^3(\vec{x} - \vec{y}) \quad (2.12.6)$$

with  $\pi = \frac{\partial L}{\partial \dot{\phi}}$  we can always write  $\phi$  as above, except with a time dependent Heisenberg picture

$$a_p(t) = e^{i(H-E_p)(t-t_0)} a_p(t_0) e^{-i(H-E_p)(t-t_0)} \quad (2.12.7)$$

so that

$$\phi(t, \vec{x}) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} \left( a_{\vec{p}}(t) e^{-ip \cdot x} + a_{\vec{p}}^\dagger(t) e^{ip \cdot x} \right) \quad (2.12.8)$$

The point is that the  $a_p(t)$  at any fixed time  $t$  satisfy the same commutation relations as before, except that they depend non-trivially on time. If the interactions are only turned on for a finite time, then  $a_p(-\infty)$  will create particles in the past, and  $a_p(\infty)$  in the future.

Note that we could take  $a_p(-\infty) = a_p(\text{free})$ , but then  $a_p(+\infty) \neq a_p(\text{free})$  for the same theory – the relation between particle creation in the far future and far past will be very non-trivial. For instance, since

$$\square \phi = -\frac{\partial V}{\partial \phi} \quad (2.12.9)$$

when  $V$  has non-trivial interaction terms we must have a very non-trivial time evolution for  $a_p(t)$ .

Now that we have defined  $a_p(t)$  as a formal object, let us prove LSZ.<sup>2</sup> A keep point is the relation

$$i \int d^4x e^{ip \cdot x} (\square + m^2) \phi(x) = \sqrt{2E_p} [a_p(\infty) - a_p(-\infty)] \quad (2.12.10)$$

To see this, first note

$$\begin{aligned} i \int d^4x e^{ip \cdot x} (\square + m^2) \phi(x) &= i \int d^4x e^{ip \cdot x} (\partial_t^2 - \vec{\nabla}^2 + m^2) \phi(x) \\ &= i \int d^4x e^{ip \cdot x} (\partial_t^2 + \vec{p}^2 + m^2) \phi(x) \\ &= i \int d^4x e^{ip \cdot x} (\partial_t^2 + E_p^2) \phi(x) \end{aligned} \quad (2.12.11)$$

We also have that

$$\begin{aligned} \partial_t [e^{ipx} (i\partial_t + E_p) \phi(x)] &= [iE_p e^{ipx} (i\partial_t + E_p) + e^{ipx} (i\partial_t^2 + E_p \partial_t)] \phi(x) \\ &= i e^{ipx} (\partial_t^2 + E_p^2) \phi(x) \end{aligned} \quad (2.12.12)$$

---

<sup>2</sup>You can see chapter 10 of Weinberg for a very different, and much more sophisticated proof.

Plugging this relation into the equation above gives

$$i \int d^4x e^{ip \cdot x} (\square + m^2) \phi(x) = \int dt \partial_t \left[ e^{iE_p t} \int d^3x e^{-i\vec{p} \cdot \vec{x}} (i\partial_t + E_p) \phi(x) \right] \quad (2.12.13)$$

So obviously we can evaluate the time integral at its boundaries in the far future and past. Since the interactions have been turned off at those times, we can use free field expressions to compute the spatial part

$$\begin{aligned} & \int d^3x e^{-i\vec{p} \cdot \vec{x}} (i\partial_t + E_p) \phi(x) \\ &= \int d^3x e^{-i\vec{p} \cdot \vec{x}} (i\partial_t + E_p) \int \frac{d^3k}{(2\pi)^3} \frac{1}{\sqrt{2E_k}} \left( a_{\vec{p}}(\pm\infty) e^{ip \cdot x} + a_{\vec{p}}^\dagger(\pm\infty) e^{-ip \cdot x} \right) \\ &= \int \frac{d^3k}{(2\pi)^3} \int d^3x \left[ \left( \frac{E_p + E_k}{\sqrt{2E_k}} \right) a_k(\pm\infty) e^{-i\vec{p} \cdot \vec{x} - ik \cdot x} + \left( \frac{E_p - E_k}{\sqrt{2E_k}} \right) a_k^\dagger(\pm\infty) e^{-i\vec{p} \cdot \vec{x} + ik \cdot x} \right] \end{aligned} \quad (2.12.14)$$

where (crucially) we used the fact that  $a_p(\pm\infty)$  are time independent. The  $\vec{x}$  integral gives a delta function, which allows us to evaluate the  $\vec{k}$  integral, giving

$$\int d^3x e^{-i\vec{p} \cdot \vec{x}} (i\partial_t + E_p) \phi(x) = \sqrt{2E_p} a_p(\pm\infty) e^{-iE_p t} \quad (2.12.15)$$

so we have shown that

$$i \int d^4x e^{ip \cdot x} (\square + m^2) \phi(x) = \sqrt{2E_p} [a_p(\infty) - a_p(-\infty)] \quad (2.12.16)$$

as desired. We can also take the complex conjugate of this result to get creation operators.

We wanted

$$\begin{aligned} \langle f|S|i \rangle &= 2^{\frac{k}{2}} \sqrt{E_1 \cdots E_k} \langle \Omega | a_{p_k}(\infty) \cdots a_{p_3}(\infty) a_{p_1}^\dagger(-\infty) a_{p_2}^\dagger(-\infty) | \Omega \rangle \\ &= 2^{\frac{k}{2}} \sqrt{E_1 \cdots E_k} \langle \Omega | T \{ a_{p_k}(\infty) \cdots a_{p_3}(\infty) a_{p_1}^\dagger(-\infty) a_{p_2}^\dagger(-\infty) \} | \Omega \rangle \end{aligned} \quad (2.12.17)$$

where  $T$ , the *time-ordering symbol*, just demands that later times are to the left of earlier times. Since  $\infty > -\infty$ , we can replace e.g.  $a(\infty)$  with  $a(\infty) - a(-\infty)$ , and conclude that we have derived the LSZ formula.

### 2.12.2 LSZ is Very General

As you saw above, our proof used very little information about  $\phi(x)$ . So what did we actually use?

1. The operator  $\phi(x)$  is a smooth function of spacetime which can be decomposed into Fourier modes.
2.  $\phi(x)$  can create and destroy one-particle states at early and late times.

That's really it!

So the one-particle states that  $\phi(x)$  creates need not be 'elementary particles'. For example, say we have a field  $\psi_p$  that creates protons and a field  $\psi_e$  that creates electrons. Then we might hope that  $\mathcal{O}(x) \equiv \psi_e(x)\psi_p(x)$  acts as

$$\langle \Omega | \mathcal{O}(x) | p_\mu \text{ hydrogen atom} \rangle = Z e^{-ipx} \quad (2.12.18)$$

so in other words,  $\mathcal{O}(x)$  creates and destroys hydrogen atoms. We can use the LSZ formula with  $\mathcal{O}(x)$  in order to look at the scattering amplitudes for hydrogen atoms! Note that

$$m_{\text{hydrogen}} = m_p + m_e - E_{\text{binding}} \quad (2.12.19)$$

and we would need to put this full  $m_{\text{hydrogen}}$  into the LSZ formula. I mention this just to make it clear that the 'masses' appearing in the LSZ formula really must be physical masses, not just some abstract parameters that appear in the Lagrangian.

Of course in this formulation we do not have a 'hydrogen atom' field in the Lagrangian – the hydrogen atom is created as a composite of a proton and an electron. Similarly, in QCD we have many different bound states of quarks and gluons. Thus we see that *there do not have to be elementary fields in the Lagrangian for each type of particle state in order to calculate the S-Matrix.*

### 2.12.3 Time Ordering and the Feynman Propagator

We have seen 'the propagator' many times already in this course. Now let us define the Feynman propagator, which is the time-ordered version. The motivation is that we want to compute scattering amplitudes using LSZ, but for this we need time-ordered correlation functions. The simplest such correlator is the 2-point function in a free theory, which is the Feynman propagator. It will be a fundamental building block for perturbation theory based on the Feynman rules, which we will discuss next.

First let us look at the Wightman propagator

$$\langle 0 | \phi(x_1) \phi(x_2) | 0 \rangle = \int \frac{d^3 k_1}{(2\pi)^3} \int \frac{d^3 k_2}{(2\pi)^3} \frac{1}{2\sqrt{E_1 E_2}} \langle 0 | a_{k_1} a_{k_2}^\dagger | 0 \rangle e^{-ik_1 x_1 + ik_2 x_2} \quad (2.12.20)$$

The  $a_{k_1} a_{k_2}^\dagger$  gives  $(2\pi)^3 \delta^3(\vec{k}_1 - \vec{k}_2)$  so we get

$$\langle 0 | \phi(x_1) \phi(x_2) | 0 \rangle = \int \frac{d^3 k}{(2\pi)^3} \frac{1}{2E_k} e^{ik \cdot (x_2 - x_1)} \quad (2.12.21)$$

Now we want to calculate the time-ordered or Feynman propagator

$$\begin{aligned} \langle 0 | T \{ \phi(x_1) \phi(x_2) \} | 0 \rangle &= \theta(t_1 - t_2) \langle 0 | \phi(x_1) \phi(x_2) | 0 \rangle + \theta(t_2 - t_1) \langle 0 | \phi(x_2) \phi(x_1) | 0 \rangle \\ &= \int \frac{d^3 k}{(2\pi)^3} \frac{1}{2E_k} (\theta(t_1 - t_2) e^{ik \cdot (x_2 - x_1)} + \theta(t_2 - t_1) e^{ik \cdot (x_1 - x_2)}) \end{aligned} \quad (2.12.22)$$

Now we can take  $\vec{k} \rightarrow -\vec{k}$  in the first term and define  $\tau = t_1 - t_2$  to get

$$\langle 0|T\{\phi(x_1)\phi(x_2)\}|0\rangle = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_k} e^{-i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)} [e^{iE_k\tau}\theta(-\tau) + e^{-iE_k\tau}\theta(\tau)] \quad (2.12.23)$$

The first and second term here are the retarded and advanced propagators that we saw in OFPT.

Now we want to prove and then utilize the identity

$$e^{iE_k\tau}\theta(-\tau) + e^{-iE_k\tau}\theta(\tau) = \lim_{\epsilon \rightarrow 0} \frac{-2E_k}{2\pi i} \int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 - E_k^2 + i\epsilon} e^{i\omega\tau} \quad (2.12.24)$$

This is based on a simpler identity

$$i\theta(\tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1}{\omega + i\epsilon} e^{-i\omega\tau} \quad (2.12.25)$$

We can prove this by noting that when  $\tau > 0$ , we must close the contour in the lower half plane, while when  $\tau < 0$  we must close it in the upper half plane. In the former case we enclose a pole, while in the latter we don't, so we get 1 when  $\tau > 0$  and 0 when  $\tau < 0$ . Note the factor of  $i$  we pick up from the contour integral.

The more complicated version is then based on

$$\begin{aligned} \frac{1}{\omega^2 - E_k^2 + i\epsilon} &= \frac{1}{(\omega - (E_k - i\epsilon))(\omega + (E_k - i\epsilon))} \\ &= \frac{1}{2E_k} \left[ \frac{1}{\omega - E_k + i\epsilon} - \frac{1}{\omega + E_k + i\epsilon} \right] \end{aligned} \quad (2.12.26)$$

where we used  $2E_k\epsilon \sim \epsilon$  since  $E_k > 0$  and all that matters is the sign of  $\epsilon$ , since we are taking it to zero at the end of the calculation.

Going over the algebra and putting it together, we have

$$D_F(x_1, x_2) = \langle 0|T\{\phi(x_1)\phi(x_2)\}|0\rangle = \int \frac{d^4k}{(2\pi i)^4} \frac{i}{k^2 - m^2 + i\epsilon} e^{ik_\mu(x_1^\mu - x_2^\mu)} \quad (2.12.27)$$

which is the Feynman propagator. Note how the time-ordering theta functions, once they are Fourier represented as above, make the propagator manifestly Lorentz invariant, with a pole at  $k^2 = m^2$ , as in the LSZ formula.

Some points that are worth noting:

- In this form of the propagator,  $k_0$  is an integration variable – it is not set to  $E_{\vec{k}}$  – and so the propagator can be off-shell, representing ‘virtual particle’ propagation.
- There’s a factor of  $i$  from the contour integral, even though the  $\phi$  field is real.
- The  $i\epsilon$  is just a reminder of the ‘pole prescription’, that is, where the pole of  $k^2 - m^2$  sits with respect to the contour of integration over purely real  $k_\mu$ . It comes from the time ordering.



- Ignoring the  $i\epsilon$ , the Feynman propagator looks just like the classical propagator for the Klein-Gordon equation, the object we've called  $\frac{1}{\square}$ . That's correct – the Feynman propagator just computes classical propagation in a complicated way.

Recall that the Feynman propagator was, in effect, what we got in OFPT by summing up retarded and advanced contributions, which arise universally.

## 2.13 Feynman Rules as a Generalization of Classical Physics

We have seen that we can compute scattering amplitudes from time-ordered correlation functions, and we already computed the simplest example of such a correlator, namely the Feynman propagator. Now it's time to learn how to compute time-ordered correlators in perturbation theory using the Feynman rules.

### 2.13.1 Derivation from the Lagrangian, as a Generalization of Classical Rules

We saw above how to derive Feynman rules for classical perturbation theory. This first derivation of the quantum Feynman rules will be a direct generalization of those methods.

In the classical case, we had an action

$$S = \int d^4x \left[ \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - V(\phi) \right] \quad (2.13.1)$$

and we used perturbative methods to solve the differential equation

$$(\square + m^2)\phi = -J - \frac{\partial V(\phi)}{\partial\phi} \quad (2.13.2)$$

by taking advantage of the Green's function  $\frac{1}{\square + m^2}$ .

The full quantum mechanical operator  $\phi(x)$  also satisfies this Euler-Lagrange equation. But it also satisfies the canonical commutation relations

$$[\phi(t, \vec{x}), \dot{\phi}(t, \vec{y})] = i\hbar\delta^3(\vec{x} - \vec{y}) \quad (2.13.3)$$

where I put in a factor of  $\hbar$  to remind you that this commutator vanishes in the classical theory. Now let's derive the *Schwinger-Dyson* equations, which apply the Euler-Lagrange equation directly to time-ordered correlators in QFT.

The idea is extremely simple. Let's just compute the derivative as it applies to a time ordered correlator, in order to show that

$$(\square_x + m^2)\langle\Omega|T\{\phi(x)\phi(y)\}|\Omega\rangle = \langle\Omega|T\{(\square_x + m^2)\phi(x)\phi(y)\}|\Omega\rangle - i\hbar\delta^4(x - y) \quad (2.13.4)$$

in the full interacting theory. The last term wouldn't be present in the classical theory, as it comes from a commutator.

Most of the derivatives just pass through into the correlator. The only issue is that the time derivative does not commute with the  $T$  (time-ordering) symbol. So we should carefully compute

$$\begin{aligned}\partial_{t_1} \langle \Omega | T \{ \phi(x_1) \phi(x_2) \} | \Omega \rangle &= \partial_{t_1} [\theta(t_1 - t_2) \langle 0 | \phi(x_1) \phi(x_2) | 0 \rangle + \theta(t_2 - t_1) \langle 0 | \phi(x_2) \phi(x_1) | 0 \rangle] \\ &= \langle \Omega | T \{ \partial_{t_1} \phi(x_1) \phi(x_2) \} | \Omega \rangle + \delta(t_1 - t_2) \langle \Omega | [\partial_t \phi(x_1), \phi(x_2)] | \Omega \rangle \\ &= \langle \Omega | T \{ \partial_{t_1} \phi(x_1) \phi(x_2) \} | \Omega \rangle\end{aligned}\quad (2.13.5)$$

But now if we take another time derivative, we get

$$\begin{aligned}\partial_{t_1}^2 \langle \Omega | T \{ \phi(x_1) \phi(x_2) \} | \Omega \rangle &= \langle \Omega | T \{ \partial_{t_1}^2 \phi(x_1) \phi(x_2) \} | \Omega \rangle + \delta(t_1 - t_2) \langle \Omega | [\partial_t^2 \phi(x_1), \phi(x_2)] | \Omega \rangle \\ &= \langle \Omega | T \{ \partial_{t_1}^2 \phi(x_1) \phi(x_2) \} | \Omega \rangle - i\hbar \delta^4(x_1 - x_2)\end{aligned}\quad (2.13.6)$$

via the canonical commutation relations. Note that in a free theory this 2-point correlator is just the Feynman propagator, and so we see that

$$(\square + m^2) D_F(x, y) = -i\hbar \delta^4(x - y) \quad (2.13.7)$$

as expected, since  $(\square + m^2)\phi = 0$  in a free theory (but this doesn't hold in an interacting theory, where there are quantum corrections to the free 2-point correlator!).

This method generalizes (as you should check) to multipoint correlators; introducing the notation  $\langle \dots \rangle = \langle \Omega | T \{ \dots \} | \Omega \rangle$ , we find

$$\begin{aligned}\square_x \langle \phi(x) \phi(x_1) \dots \phi(x_n) \rangle &= \langle \square_x \phi(x) \phi(x_1) \dots \phi(x_n) \rangle \\ &\quad - i\hbar \sum_j \delta^4(x - x_j) \langle \phi(x_1) \dots \phi(x_{j-1}) \phi(x_{j+1}) \dots \phi(x_n) \rangle\end{aligned}\quad (2.13.8)$$

We can immediately use the EoM  $\square\phi = -\frac{\partial V}{\partial\phi}$  to write this as

$$\begin{aligned}\square_x \langle \phi(x) \phi(x_1) \dots \phi(x_n) \rangle &= \left\langle -\frac{\partial V(\phi(x))}{\partial\phi(x)} \phi(x_1) \dots \phi(x_n) \right\rangle \\ &\quad - i\hbar \sum_j \delta^4(x - x_j) \langle \phi(x_1) \dots \phi(x_{j-1}) \phi(x_{j+1}) \dots \phi(x_n) \rangle\end{aligned}\quad (2.13.9)$$

These are the Schwinger-Dyson equations. The delta function terms are called *contact interactions*, they make the difference between the classical and quantum theories.

Now we can apply the Schwinger-Dyson equations to compute time-ordered correlators in perturbation theory.

For efficiency we define

$$\delta_{ij} \equiv \delta^4(x_i - x_j) \quad (2.13.10)$$

$$D_{ij} = D_{ji} \equiv D_F(x_i, x_j) \quad (2.13.11)$$

so for example

$$\square_x D_{x1} = -i\delta_{x1} \quad (2.13.12)$$

demonstrates that the Feynman propagator is a Green's function.

Now let us use the Schwinger-Dyson equations to expand a simple correlator in perturbation theory. But first, in a free theory, note that

$$\langle \phi_1 \phi_2 \rangle_{free} = \int d^4x D_{x1} \delta_{x2} = D_{12} \quad (2.13.13)$$

and for a 4-point correlator we get

$$\begin{aligned} \langle \phi_1 \phi_2 \phi_3 \phi_4 \rangle &= i \int d^4x D_{x1} \square_x \langle \phi_x \phi_2 \phi_3 \phi_4 \rangle \\ &= \int d^4x D_{x1} (\delta_{x2} \langle \phi_3 \phi_4 \rangle + \delta_{x3} \langle \phi_2 \phi_4 \rangle + \delta_{x4} \langle \phi_2 \phi_3 \rangle) \end{aligned} \quad (2.13.14)$$

Evaluating using the Feynman propagator for these 2-pt functions gives

$$\langle \phi_1 \phi_2 \phi_3 \phi_4 \rangle = D_{12} D_{34} + D_{13} D_{24} + D_{14} D_{23} \quad (2.13.15)$$

which has a diagrammatic expansion, where  $D_{ij}$  are lines and the  $x_i$  are points.

Now let's add interactions. We'll use a Lagrangian

$$L = \frac{1}{2}(\partial_\mu \phi)^2 + \frac{g}{3!} \phi^3 \quad (2.13.16)$$

and we are imagining that  $g$  is small, so that we will treat it perturbatively. We can write

$$\langle \phi_1 \phi_2 \rangle = \int d^4x \delta_{x1} \langle \phi_x \phi_2 \rangle = i \int d^4x (\square_x D_{x1}) \langle \phi_x \phi_2 \rangle = i \int d^4x D_{x1} \square_x \langle \phi_x \phi_2 \rangle \quad (2.13.17)$$

and this is useful because now we can apply what we learned above. Applying the S-D equations to the 2-pt function in the interacting theory gives

$$\begin{aligned} \langle \phi_1 \phi_2 \rangle &= i \int d^4x D_{1x} (\langle \square_x \phi_x \phi_2 \rangle - i \delta_{x2}) \\ &= i \int d^4x D_{1x} \left( \left\langle -\frac{\partial V(\phi_x)}{\partial \phi_x} \phi_2 \right\rangle - i \delta_{x2} \right) \\ &= i \int d^4x D_{1x} \left( \frac{g}{2} \langle \phi_x^2 \phi_2 \rangle - i \delta_{x2} \right) \end{aligned} \quad (2.13.18)$$

What we have seen is that applying the S-D equations gives us terms that are one higher order in perturbation theory in  $g$ . After applying the S-D equations  $k$  times, we can then stop and evaluate using the free field correlators. Note, however, that if we stop with the result above, it will vanish, since  $\langle \phi_x^2 \phi_2 \rangle = 0$  in a free theory.

So let's apply S-D once more. We find

$$\begin{aligned} \langle \phi_1 \phi_2 \rangle &= D_{12} - \frac{g}{2} \int d^4x d^4y D_{x1} D_{y2} \square_y \langle \phi_x^2 \phi_y \rangle \\ &= D_{12} - \frac{g^2}{4} \int d^4x d^4y D_{x1} D_{y2} \langle \phi_x^2 \phi_y^2 \rangle + ig \int d^4x D_{1x} D_{2x} \langle \phi_x \rangle \end{aligned} \quad (2.13.19)$$

If we are only interested in working to order  $g^2$ , then we need only expand the  $\langle\phi_x\rangle$  term once more, giving

$$\langle\phi_x\rangle = i \int d^4y D_{xy} \square_y \langle\phi_y\rangle = i \frac{g}{2} \int d^4y D_{xy} \langle\phi_y^2\rangle \quad (2.13.20)$$

We can evaluate all that remains using the free field correlators, giving

$$\langle\phi_1\phi_2\rangle = D_{12} - g^2 \int d^4x d^4y \left( \frac{1}{2} D_{1x} D_{xy}^2 D_{y2} + \frac{1}{4} D_{1x} D_{xx} D_{yy} D_{y2} + \frac{1}{2} D_{1x} D_{2x} D_{xy} D_{yy} \right) \quad (2.13.21)$$

These can be displayed as Feynman diagrams.

With this, and quite a bit of inspiration, we can guess the Feynman rules.

1. Draw a point for each  $x_i$  in the correlator, and line coming from that point.
2. A line can either connect to an existing point, giving a Feynman propagator, or it can split into 2 lines at a new point  $x$ . A split gives a term proportional to  $-iV'[\phi]$ , and the new point  $x$  gets integrated over.
3. At a given order in perturbation theory, we just write down all diagrams with all lines contracted, and add them up.

These are the Feynman rules in position space; soon we'll derive an equivalent version in momentum space, which allow for a more direct computation of scattering amplitudes.

We write interactions as

$$\frac{g}{3!}\phi^3, \quad \frac{\lambda}{n!}\phi^n, \quad \frac{\kappa}{3!5!2!}\phi_1^3\phi_2^5\phi_3^2 \quad (2.13.22)$$

because the factorials tend to get canceled. The  $n$  of  $n!$  is canceled by the derivative of  $V'[\phi]$ , and then the  $(n-1)!$  will be canceled (generically) by different possible permutations of the lines coming out of a vertex. Occasionally these factors aren't fully canceled, due to 'symmetry factors' of the diagram, which is what actually happened in our example above. Boo hoo. Symmetry factors rarely appear in relevant physical theories, but they're something to be aware of.

## 2.14 The Hamiltonian Formalism for Perturbation Theory

Now let us derive the Feynman rules in what's probably the most standard, systematic way. There's still one more derivation possible, the Path Integral derivation, but we won't cover it until much later in the semester.

The first step is to write the Hamiltonian in perturbation theory as

$$H = H_0 + V \quad (2.14.1)$$

where  $H_0$  can be solved exactly, and for us it will always be the Hamiltonian for a free QFT. The interaction might be e.g.

$$V(t) = \int d^3x \frac{g}{3!} \phi^3 \quad (2.14.2)$$

as we studied before using the Schwinger-Dyson equations. This is all in the Heisenberg Picture. Note that in the Heisenberg picture, the operator  $H$  is time independent, in that

$$i\partial_t H = [H, H] = 0 \quad (2.14.3)$$

so it is the same at all times.

Ultimately we would like to calculate time-ordered correlation functions of Heisenberg Picture fields. However, to facilitate computation in perturbation theory in QFT, we must switch to a different scheme, where both the fields and the states evolve, but in different ways. This is called the *Interaction Picture*, where we make the choice that *fields*  $\phi_0$  *evolve according to*  $H_0$ . We label the interaction picture fields as  $\phi_0(t, \vec{x})$  since they are identical to the free fields.

Why? Because the best we can do, in perturbation theory, is evaluate various correlators with the free fields  $\phi_0(t, \vec{x})$ , so we need to express all computations in terms of their matrix elements. In particular

- Heisenberg picture operators  $\phi(t, \vec{x})$  must be expressed in terms of  $\phi_0(t, \vec{x})$
- The exact vacuum state  $|\Omega\rangle$  must be related to the free vacuum  $|0\rangle$
- We would like the end result to be *manifestly Lorentz invariant*

We can do all of this using the interaction picture fields.

### 2.14.1 Interaction Picture Formalism – Heisenberg Fields from Free Fields

In the Hamiltonian formalism, and in the Heisenberg picture, where the exact vacuum  $|\Omega\rangle$  is time-independent, but operators depend on time... the field satisfies the exact EoM

$$-i\partial_t \phi = [H, \phi] \quad (2.14.4)$$

This can be given a formal solution

$$\phi(t, \vec{x}) = S^\dagger(t, t_0) \phi(t_0, \vec{x}) S(t, t_0) \quad (2.14.5)$$

where  $S$  is the time evolution operator in the Heisenberg picture, or the S-Matrix, satisfying

$$i\partial_t S(t, t_0) = H(t) S(t, t_0), \implies -i\partial_t S^\dagger(t, t_0) = S^\dagger(t, t_0) H(t) \quad (2.14.6)$$

We can see directly that this solves the EoM for  $\phi$ .

We can write the full Heisenberg picture fields in terms of the interaction picture fields as

$$\begin{aligned}\phi(t, \vec{x}) &= S^\dagger(t, t_0) e^{-iH_0(t-t_0)} \phi_0(t, \vec{x}) e^{iH_0(t-t_0)} S(t, t_0) \\ &\equiv U^\dagger(t, t_0) \phi_0(t, \vec{x}) U(t, t_0)\end{aligned}\tag{2.14.7}$$

so that  $U(t, t_0) = e^{iH_0(t-t_0)} S(t, t_0)$  relates the Heisenberg picture fields to the free fields at the same time  $t$ . Note that the placeholder  $t_0$  is a fixed time where all pictures (Heisenberg, Schrodinger, Interaction) coincide.

It's simple enough to write  $U$  in the Heisenberg picture, this is

$$U(t_2, t_1) = e^{iH_0(t_2-t_0)} e^{-iH(t_2-t_1)} e^{-iH_0(t_1-t_0)}\tag{2.14.8}$$

but we would like an expression in the interaction picture, ie in terms of the  $\phi_0$  fields.

To do this, we can derive a differential equation for  $U$ , and then solve it. Note that

$$\begin{aligned}i\partial_t U(t, t_0) &= i(\partial_t e^{iH_0(t-t_0)}) S(t, t_0) + e^{iH_0(t-t_0)} i\partial_t S(t, t_0) \\ &= e^{iH_0(t-t_0)} (H - H_0) e^{-iH_0(t-t_0)} e^{iH_0(t-t_0)} S(t, t_0) \\ &= V_I(t) U(t, t_0)\end{aligned}\tag{2.14.9}$$

where  $V_I(t) = e^{iH_0(t-t_0)} V(t_0) e^{-iH_0(t-t_0)}$  is, by definition, the interaction picture potential. Since at the time  $t_0$  all of the 'pictures' are identical, we have that

$$V_I(t) = V[\phi_0(t, \vec{x})]\tag{2.14.10}$$

so for example for a  $\phi^3$  potential, we have

$$V_I(t) = e^{iH_0(t-t_0)} \left[ \int d^3x \frac{g}{3!} \phi_0(t_0, \vec{x})^3 \right] e^{-iH_0(t-t_0)} = \int d^3x \frac{g}{3!} \phi_0(t, \vec{x})^3\tag{2.14.11}$$

since  $H_0$  is Hermitian and so  $e^{iH_0(t-t_0)}$  is Unitary.

If everything commuted we would write

$$S \sim e^{-i \int_{t_0}^t H dt'} \implies U \sim e^{-i \int_{t_0}^t V(t') dt'}\tag{2.14.12}$$

but we cannot assume that these operators commute; specifically we cannot assume that  $V(t_1)$  and  $V(t_2)$  commute. But the right answer is similar, it is

$$U(t, t_0) = T \left\{ \exp \left[ -i \int_{t_0}^t dt' V_I(t') \right] \right\}\tag{2.14.13}$$

where  $T$  is the time ordering symbol. This is defined via the series expansion

$$U(t, t_0) = 1 - i \int_{t_0}^t dt' V_I(t') - \frac{1}{2} \int_{t_0}^t dt' \int_{t_0}^t dt'' T \{ V_I(t') V_I(t'') \} + \dots\tag{2.14.14}$$

and is often called a Dyson series.

To see that it solves the differential equation, note that we can instead write an integral equation

$$U(t, t_0) = 1 - i \int_{t_0}^t dt' V(t') U(t', t_0) \quad (2.14.15)$$

which is automatically satisfied if we differentiate. Now we can just iterate to see that

$$U = 1 - i \int_{t_0}^t dt' V(t') \left[ 1 - i \int_{t_0}^{t'} dt'' V(t'') U(t'', t_0) \right] \quad (2.14.16)$$

The last term is clearly time ordered, so we can re-write

$$\int_{t_0}^t dt' \int_{t_0}^{t'} dt'' V_I(t') V_I(t'') = \frac{1}{2} \int_{t_0}^t dt' \int_{t_0}^t dt'' T \{ V_I(t') V_I(t'') \} \quad (2.14.17)$$

Proceeding in this way to higher orders gives the desired solution.

In general, it's useful notationally to write

$$U_{21} = U(t_2, t_1) = U^\dagger(t_1, t_2) = U^{-1}(t_1, t_2) \quad (2.14.18)$$

since  $U$  is unitary. We also have that

$$U_{32} U_{21} = U_{31} \quad (2.14.19)$$

where we always have later times on the left. We can use this to write the Heisenberg picture field at time  $t_1$  as

$$\phi(t_1, \vec{x}_1) = U_{10}^\dagger \phi_0(t_1, \vec{x}_1) U_{10} = U_{01} \phi_0(x_1) U_{10} \quad (2.14.20)$$

in terms of the interaction picture field at  $t_1$ .

### 2.14.2 Exact Vacuum and Heisenberg Correlators in the Interaction Picture

Now that we have that formalism available, we need to figure out how to write the exact vacuum  $|\Omega\rangle$  in the interaction picture.

We can get  $|\Omega\rangle$  from the following argument, based on the fact that it's the lowest energy state. Let's imagine evolving

$$e^{-iH(T+t_0)}|0\rangle = e^{-iE_0(T+t_0)}|\Omega\rangle\langle\Omega|0\rangle + \sum_{n \neq 0} e^{-iE_n(T+t_0)}|n\rangle\langle n|0\rangle \quad (2.14.21)$$

Now since  $|\Omega\rangle$  is the lowest energy state, if we send

$$T \rightarrow \infty(1 - i\epsilon) \quad (2.14.22)$$

then only  $\Omega$  will survive, so we can write the formal expression

$$\begin{aligned} |\Omega\rangle &= \mathcal{N}_i \lim_{T \rightarrow \infty(1-i\epsilon)} e^{-iH(t_0-(-T))} e^{-iH_0(-T-t_0)} |0\rangle \\ &= \mathcal{N}_i [U(t_0, -T)|0\rangle] \end{aligned} \quad (2.14.23)$$

where  $\mathcal{N}_i$  is a normalization related to  $\langle\Omega|0\rangle$  which will drop out of all our computations. We obtain a similar relation for  $\langle\Omega|$ .

The book gives an alternative argument, based on the idea that since at  $t_0$  the Heisenberg and interaction picture annihilation operators are the same, the states

$$\lim_{t \rightarrow -\infty} e^{iH_0(t-t_0)} |0\rangle \propto \lim_{t \rightarrow -\infty} S(t, t_0) |\Omega\rangle \quad (2.14.24)$$

and the proportionality constant is just a normalization.

Now we can compute the Heisenberg picture correlators that we wanted using the interaction picture. Let's assume for a moment that  $t_1 > t_2 > \dots > t_n$ , so that

$$\langle\Omega|T\{\phi_1(x_1) \dots \phi_n(x_n)\}|\Omega\rangle = \langle\Omega|\phi_1(x_1) \dots \phi_n(x_n)|\Omega\rangle \quad (2.14.25)$$

Then we can rewrite this as

$$\begin{aligned} &= \mathcal{N}_i \mathcal{N}_f \langle 0|U_{\infty 0} U_{01} \phi_0(x_1) U_{10} U_{02} \phi_0(x_2) U_{20} U_{03} \dots U_{0n} \phi_0(x_n) U_{n0} U_{n-\infty} |0\rangle \\ &= \mathcal{N}_i \mathcal{N}_f \langle 0|U_{\infty 1} \phi_0(x_1) U_{12} \phi_0(x_2) U_{23} \dots U_{n-1,n} \phi_0(x_n) U_{n-\infty} |0\rangle \end{aligned} \quad (2.14.26)$$

Finally we can just write this using the time ordering symbol as

$$\langle\Omega|T\{\phi(x_1) \dots \phi(x_n)\}|\Omega\rangle = \frac{\langle 0|T\{\phi_0(x_1) \dots \phi_0(x_n) U_{\infty, -\infty}\} |0\rangle}{\langle 0|U_{\infty, -\infty} |0\rangle} \quad (2.14.27)$$

where in the numerator the time ordering symbol insists that both the  $\phi(x_i)$  and the  $U$  factors must be time ordered together.

Let's note something simple, namely that

$$\begin{aligned} U_{\infty, -\infty} &= T \left\{ \exp \left[ -i \int_{-\infty}^{\infty} dt' V_I(t') \right] \right\} \\ &= T \left\{ \exp \left[ i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x) \right] \right\} \end{aligned} \quad (2.14.28)$$

where

$$\mathcal{L}_{Int} = \mathcal{L} - \mathcal{L}_{free} \quad (2.14.29)$$

because we derived the interaction  $V$  from the Lagrangian. So we can write

$$\langle\Omega|T\{\phi(x_1) \dots \phi(x_n)\}|\Omega\rangle = \frac{\langle 0|T\left\{ \phi_0(x_1) \dots \phi_0(x_n) e^{i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)} \right\} |0\rangle}{\langle 0|T\left\{ e^{i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)} \right\} |0\rangle} \quad (2.14.30)$$

Finally, we have a manifestly Lorentz invariant result!



## 2.15 Feynman Rules from the Hamiltonian Formalism

Now that we have a beautiful Lorentz invariant formula for the time ordered product, we just expand in  $\mathcal{L}_{Int}$ . So for example in a theory with a  $\phi^3$  interaction, we have the computation

$$\begin{aligned} \langle 0|T \left\{ \phi_0(x_1)\phi_0(x_2)e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)} \right\} |0\rangle &= \langle 0|T \{ \phi_0(x_1)\phi_0(x_2) \} |0\rangle \\ &+ i\frac{g}{3!} \int d^4x \langle 0|T \{ \phi_0(x_1)\phi_0(x_2)\phi^3(x) \} |0\rangle \\ &+ \frac{1}{2} \left( i\frac{g}{3!} \right)^2 \int d^4x d^4y \langle 0|T \{ \phi_0(x_1)\phi_0(x_2)\phi^3(x)\phi^3(y) \} |0\rangle + \dots \end{aligned} \quad (2.15.1)$$

So how do we evaluate this?

### 2.15.1 Evaluating Vacuum Expectation Values with Wick's Theorem

It's helpful to write  $\phi_0 = \phi_+ + \phi_-$  where

$$\phi_+ = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} a_p^\dagger e^{ipx}, \quad \phi_- = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} a_p e^{-ipx} \quad (2.15.2)$$

because then

$$\phi_-|0\rangle = 0, \quad \langle 0|\phi_+ = 0 \quad (2.15.3)$$

Then when we compute

$$\begin{aligned} &\langle 0|T \{ \phi_0(x_1)\phi_0(x_2)\phi^3(x)\phi^3(y) \} |0\rangle \\ &= \langle 0|T \{ [\phi_+(x_1) + \phi_-(x_1)][\phi_+(x_2) + \phi_-(x_2)][\phi_+(x) + \phi_-(x)]^3[\phi_+(y) + \phi_-(y)]^3 \} |0\rangle \end{aligned}$$

we must have exactly 4 of the  $\phi_+$  and  $\phi_-$  to get a non-vanishing result.

Let's look at the case of a 2-point correlator, where we assume  $x_0 > y_0$ :

$$\begin{aligned} T \{ \phi_0(x)\phi_0(y) \} &= \phi_-(x)\phi_-(y) + \phi_+(x)\phi_-(y) + \phi_-(x)\phi_+(y) + \phi_+(x)\phi_+(y) \\ &= \phi_-(x)\phi_-(y) + \phi_+(x)\phi_-(y) + \phi_+(y)\phi_-(x) + \phi_+(x)\phi_+(y) \\ &\quad + [\phi_-(x), \phi_+(y)] \end{aligned} \quad (2.15.4)$$

The expectation value of the first line is 0, because it has been *normal ordered*, meaning that we have moved all of the annihilation operators to the right of all the creation operators, introducing a bunch of commutators along the way. We conventionally denote the normal ordered version of a product of operators as

$$: \phi(x_1)\phi(x_2)\cdots\phi(x_n) : \quad (2.15.5)$$

These have vanishing vacuum expectation values, so it's the contraction piece that's interesting.

Each pair *contraction* of  $\phi_0$  with  $\phi_0$  in a time ordered correlator produces

$$\theta(x_0 - y_0)[\phi_-(x), \phi_+(y)] + \theta(y_0 - x_0)[\phi_-(y), \phi_+(x)] \quad (2.15.6)$$

which is just a Feynman propagator

$$D_F(x_1, x_2) = \langle 0 | T \{ \phi_0(x_1) \phi_0(x_2) \} | 0 \rangle \quad (2.15.7)$$

which, of course, we already know how to evaluate.

There is a simple result called Wick's theorem that says that

$$T \{ \phi_0(x_1) \phi_0(x_2) \cdots \phi_0(x_n) \} = : \phi_0(x_1) \cdots \phi_0(x_n) + \text{sum of all contractions of pairs} : \quad (2.15.8)$$

where by all possible contractions we mean terms with 1 contracted pair, terms with 2 contracted pairs, etc until as many pairs as possible are contracted. For example this means

$$\begin{aligned} T \{ \phi_1 \phi_2 \phi_3 \phi_4 \} &= : \phi_1 \phi_2 \phi_3 \phi_4 + D_{12} \phi_3 \phi_4 + D_{13} \phi_2 \phi_4 + D_{14} \phi_2 \phi_3 \\ &\quad + D_{23} \phi_1 \phi_4 + D_{24} \phi_1 \phi_3 + D_{34} \phi_1 \phi_2 \\ &\quad + D_{12} D_{34} + D_{13} D_{24} + D_{14} D_{23} : \end{aligned} \quad (2.15.9)$$

We already proved it for  $n = 2$ . To prove it in general, one proceeds by induction. As usual in any proof involving the time ordering symbol  $T$ , we assume that  $t_1 > t_2 > \cdots > t_n$  and then simply note that the result will be true for any ordering, since WLOG we could have assumed that ordering from the beginning. So we have

$$\begin{aligned} T \{ \phi_1 \cdots \phi_n \} &= \phi_1 \cdots \phi_n \\ &= (\phi_1^+ + \phi_1^-) : \phi_2 \cdots \phi_n + (n-1) \text{ contractions} : \end{aligned} \quad (2.15.10)$$

Now the  $\phi_1^+$  term is already in normal order, so we don't have to do anything with it. To normal order  $\phi_1^-$  we need to move it through past all of the other operators. Thus every term with any  $\phi_i^+$  can be contracted with  $\phi_1^-$  as it passes through. But this gives all possible contractions of the  $n-1$   $\phi_i$  contracted in all possible ways with  $\phi_1$ . So we've demonstrated the statement with  $n$  operators, and by induction we have Wick's theorem.

With Wick's theorem, we have an extremely systematic way to evaluate our time-ordered correlators. When we evaluate

$$\langle 0 | T \left\{ \phi_0(x_1) \cdots \phi_0(x_n) e^{i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)} \right\} | 0 \rangle \quad (2.15.11)$$

we just expand the exponential to some given order, giving some number of powers of the coupling, and a proliferation of fields. We then evaluate the vacuum expectation value using Wicks theorem, providing a propagator  $D_{xy}$  for each contraction, integrating over interactions points, and labeling these points with the coupling.

### 2.15.2 Feynman Diagrams, Systematically

But to get the Feynman rules right in detail, we need to understand some additional subtleties. For example

$$\begin{aligned} \langle 0|T \{ \phi_0(x_1)\phi_0(x_2)\phi_0^3(x)\phi_0^3(y) \} |0\rangle &= D_{12} (6D_{xy}^3 + 3^2 D_{xx}D_{xy}D_{yy}) \\ &+ 18D_{1x}D_{2x}D_{xy}D_{yy} + 9D_{1x}D_{2y}D_{xx}D_{yy} + 18D_{1x}D_{2y}D_{xy}^2 \\ &+ 18D_{1y}D_{2y}D_{xy}D_{xx} + 9D_{1y}D_{2x}D_{xx}D_{yy} + 18D_{1y}D_{2x}D_{xy}^2 \end{aligned} \quad (2.15.12)$$

We need to understand the combinatorial factors multiplying each term, and the special disconnected or ‘bubble diagrams’ on the first line.

We choose interactions as, for example

$$\frac{\lambda_n}{n!} \phi^n \quad (2.15.13)$$

because for each interaction, generically there are  $n!$  different permutations among the possible contractions.

The only exceptions come from ‘symmetry factors’, and in fact every term in our example above has these factors. These reduce the coefficient of the term because the  $1/n!$  is incompletely canceled.

First let’s consider a diagram without a symmetry factor, namely

$$\langle 0|T \{ \phi_0(x_1)\phi_0(x_2)\phi_0(x_3)\phi_0^3(x)\phi_0^3(y)\phi_0^3(z) \} |0\rangle \supset (3!)^3 D_{1x}D_{2y}D_{3z}D_{xy}D_{yz}D_{zx} \quad (2.15.14)$$

Here we have  $3!$  ways of pairing up each of the  $x$ ,  $y$ , and  $z$  triplets of fields. So there is *no symmetry factor* here, because the  $1/(3!)^3$  has been completely canceled.

Next consider

$$9D_{xx}D_{xy}D_{yy} = \langle 0|\phi_-(x)\phi_-(x)\phi_+(x)\phi_-(y)\phi_+(y)\phi_+(y)|0\rangle \quad (2.15.15)$$

Here we have 3 choices of the  $\phi_0(x)$  and three choices of the  $\phi_0(y)$  to contract, but once we’ve made those choices there are no more possibilities. This means that the symmetry factor is  $(3!)^2/9 = 4$ . We get a factor of 2 from  $D_{xx}$  and from  $D_{xy}$ .

Finally consider

$$6D_{xy}^3 = \langle 0|\phi_-(x)\phi_-(x)\phi_-(x)\phi_+(y)\phi_+(y)\phi_+(y)|0\rangle \quad (2.15.16)$$

Here we have  $3 \times 2 \times 1$  ways of pairing/contracting the  $\phi(x)$  with the  $\phi(y)$ , where contractions come from commutators. Thus the symmetry factor here is  $1/(3!)$ .

There is a general, simple, diagrammatic rule. If, with the points fixed, we can perform an operation that exchanges identical propagators with identical endpoints, then there is a symmetry factor associated with the number of such exchanges.

In most physical theories, there are no symmetry factors, and so the  $n!$  from the definition of the interaction will be completely canceled.

Next let us eliminate that pesky factor of

$$\langle 0|T\{e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\}|0\rangle \quad (2.15.17)$$

from the denominator in our formula for the exact time ordered correlator. If we evaluate it in our pet theory, we get

$$\begin{aligned} \langle 0|T\{e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\}|0\rangle &= \langle 0|0\rangle + \int d^4x d^4y \left(i\frac{g}{3!}\right)^2 \langle 0|T\{\phi_0^3(x)\phi_0^3(y)\}|0\rangle \\ &+ \int d^4w d^4x d^4y d^4z \left(i\frac{g}{3!}\right)^4 \langle 0|T\{\phi_0^3(w)\phi_0^3(x)\phi_0^3(y)\phi_0^3(z)\}|0\rangle + \dots \end{aligned} \quad (2.15.18)$$

If we expand this, we get lots and lots of ‘bubble diagrams’.

Note that every single bubble diagram that we get here also arises from

$$\langle 0|T\left\{\phi_0(x_1)\cdots\phi_0(x_n)e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\right\}|0\rangle \quad (2.15.19)$$

if we expand the exponential and decide not to contract any of the terms from it with the external  $\phi(x_i)$  in the correlator. To be precise, it contains some terms with *connected components that do not involve any of the external points  $x_i$* . These terms always factor out from the rest of the graph, and cancel with the denominator term. Thus we conclude

$$\begin{aligned} \langle \Omega|T\{\phi(x_1)\cdots\phi(x_n)\}|\Omega\rangle &= \frac{\langle 0|T\left\{\phi_0(x_1)\cdots\phi_0(x_n)e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\right\}|0\rangle}{\langle 0|T\{e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\}|0\rangle} \\ &= \langle 0|T\left\{\phi_0(x_1)\cdots\phi_0(x_n)e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\right\}|0\rangle_{\text{no bubbles}} \end{aligned} \quad (2.15.20)$$

This simplifies our task significantly.

Now we have provided a systematic derivation of the Feynman rules. One must draw all diagrams, associate points with interactions and lines with propagators, and divide by the symmetry factor, and then take the sum. This gives the time ordered correlation function.

### 2.15.3 Momentum Space Rules and the S-Matrix

As a final task, let us transform the rules we have derived to momentum space. We will see that this makes it trivial to write down rules for the S-Matrix itself, using the LSZ formula, and not just for the time-ordered correlation functions.

Let’s consider a simple example like

$$\langle 0|T\left\{\phi_0(x_1)\phi_0(x_2)\phi_0(x_3)e^{i\int_{-\infty}^{\infty} d^4x \mathcal{L}_{Int}(x)}\right\}|0\rangle = 0 + igD_{1x}D_{2x}D_{3x} + \dots \quad (2.15.21)$$

It’s natural to write each propagator directly in momentum space as

$$D_{xy} = \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} e^{ik\cdot(x-y)} \quad (2.15.22)$$

Thus to transform the correlation function to momentum space we must compute

$$ig \int d^4x_1 d^4x_2 d^4x_3 e^{-i(p_1x_1+p_2x_2+p_3x_3)} \int d^4x e^{ik_1(x-x_1)+ik_2(x-x_2)+ik_3(x-x_3)} \int \frac{d^4k_1}{(2\pi)^4} \frac{d^4k_2}{(2\pi)^4} \frac{d^4k_3}{(2\pi)^4} \frac{i}{k_1^2 - m^2 + i\epsilon} \frac{i}{k_2^2 - m^2 + i\epsilon} \frac{i}{k_3^2 - m^2 + i\epsilon} \quad (2.15.23)$$

The integral over  $x$  yields a factor of

$$(2\pi)^4 \delta^4(k_1 + k_2 + k_3) \quad (2.15.24)$$

while the integral over  $x_i$  for  $i = 1, 2, 3$  yields

$$(2\pi)^4 \delta^4(p_i - k_i) \quad (2.15.25)$$

This enables us to immediately do the  $k_i$  integrals, giving

$$\langle \phi_0(p_1) \phi_0(p_2) \phi_0(p_3) \rangle = (2\pi)^4 \delta^4(p_1 + p_2 + p_3) (ig) \frac{i}{p_1^2 - m^2} \frac{i}{p_2^2 - m^2} \frac{i}{p_3^2 - m^2} \quad (2.15.26)$$

What happened? Each propagator gave us a factor of

$$\frac{i}{p^2 - m^2} \quad (2.15.27)$$

and the (each) internal vertex gave us a 4-momentum conserving delta function. Also, we obtained an overall factor of

$$(2\pi)^4 \delta^4 \left( \sum_i p_i \right) \quad (2.15.28)$$

which explains why we defined

$$S = 1 + i(2\pi)^4 \delta^4 \left( \sum_i p_i \right) \mathcal{M}(p_i) \quad (2.15.29)$$

when we studied decay rates and cross sections. These are general features, and so if we are interested in scattering we can simply work with  $\mathcal{M}$ .

Furthermore, note that each external propagator contributes exactly

$$i(p^2 - m^2)^{-1} = i(\square + m^2)^{-1} \quad (2.15.30)$$

so the external propagators are exactly cancelled by the LSZ prescription, which instructs us to multiply by

$$-i(p^2 - m^2) = -i(\square + m^2) \quad (2.15.31)$$

This means that all of those funny factors in LSZ simply cancel the external propagators when we use the Feynman rules to compute scattering amplitudes!

So the rules are

- Momentum must be assigned to each line, and it must be conserved at each vertex.
- Propagators get factors of

$$D_p = \frac{i}{p^2 - m^2 + i\epsilon} \quad (2.15.32)$$

External propagators are canceled in the S-Matrix by the LSZ formula, although they remain in time-ordered correlators.

- Vertices get their usual factors of  $i$  times the coupling, e.g.  $ig$  in our example.
- All internal momenta must be integrated over. For tree diagrams they will all be determined by external momenta, so the integrals will be evaluated on delta functions, rendering them trivial. But for any diagrams with internal loops there will be momentum integrals left over, from which quantum effects arise.

#### 2.15.4 Connected vs Disconnected Graphs

It's worth noting that some diagrams for a given scattering amplitude are connected, while others break up into multiple disconnected pieces. The latter have physical effects, but they are always simply given as the product of some connected graphs (obviously), and moreover, they cannot interfere with the connected graphs because they have extra delta function singularities. So for most of the course we will tend to focus explicitly on the connected graphs.

#### 2.15.5 Some Examples

Let us illustrate all this formalism with some examples.

First we can consider the classic example of

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{m^2}{2}\phi^2 - \frac{\lambda_n}{n!}\phi^n \quad (2.15.33)$$

Here we just have vertices

$$-i\lambda_n \quad (2.15.34)$$

and the usual propagators. In particular, with  $\lambda_3 = -g$ , we can compute the 2-to-2 scattering amplitude

$$\begin{aligned} \mathcal{M} &= ig \frac{i}{(p_1 + p_2)^2 - m^2} ig + ig \frac{i}{(p_1 - p_3)^2 - m^2} ig + ig \frac{i}{(p_1 - p_4)^2 - m^2} ig \\ &= \frac{-ig^2}{s - m^2} + \frac{-ig^2}{t - m^2} + \frac{-ig^2}{u - m^2} \end{aligned} \quad (2.15.35)$$

where  $s = (p_1 + p_2)^2$  etc. We can use this to compute an actual scattering cross section from

$$\begin{aligned}\frac{d\sigma}{d\Omega} &= \frac{1}{64\pi^2 E_{CM}^2} |\mathcal{M}|^2 \\ &= \frac{g^4}{64\pi^2 E_{CM}^2} \left| \frac{1}{s - m^2} + \frac{1}{t - m^2} + \frac{1}{u - m^2} \right|^2\end{aligned}\quad (2.15.36)$$

This is a true non-trivial differential cross section.

Now let us consider a theory such as

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{(4!)} \frac{1}{\Lambda^4} (\partial\phi)^4 \quad (2.15.37)$$

What about the derivatives? In position space we will get derivatives of propagators from computations like

$$\langle \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)(\partial\phi(x))^4 \rangle = \partial_\mu D_{1x} \partial^\mu D_{2x} \partial_\nu D_{3x} \partial^\nu D_{4x} + \text{perms} \quad (2.15.38)$$

When we transform to momentum space, these simply become factors if  $ip_\mu$ , or more specifically

$$(ip_1 \cdot ip_2)(ip_3 \cdot ip_4) + (ip_1 \cdot ip_3)(ip_2 \cdot ip_4) + (ip_1 \cdot ip_4)(ip_2 \cdot ip_3) \quad (2.15.39)$$

where here I have taken all momenta incoming, so  $p_1 + p_2 + p_3 + p_4 = 0$ . This means that derivative interactions lead to Feynman rules where  $\partial \rightarrow ip_\mu$  applied to the relevant field. Note that there is an important sign issue here, because

$$D_{xy} = \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} e^{ik \cdot (x-y)} \quad (2.15.40)$$

so that

$$\partial_x D_{xy} = -\partial_x D_{yx} \quad (2.15.41)$$

This is a consequence of the physical fact that

$$\phi(p) \propto a_p e^{-ipx} + a_p^\dagger e^{ipx} \quad (2.15.42)$$

and so the sign of the momentum flips depending on whether we create or destroy a particle.

If we take the simple example of a massless theory, so  $(p_1 + p_2)^2 = 2p_1 \cdot p_2 = s$ , then we can write the result as

$$\mathcal{M} = \frac{1}{4\Lambda^4} (s^2 + t^2 + u^2) \quad (2.15.43)$$

for the scattering amplitude in such a theory. This means the cross section is

$$\begin{aligned}\frac{d\sigma}{d\Omega} &= \frac{1}{64\pi^2 E_{CM}^2} |\mathcal{M}|^2 \\ &= \frac{1}{1024\pi^2 E_{CM}^2 \Lambda^4} |s^2 + t^2 + u^2|^2\end{aligned}\quad (2.15.44)$$

Now let's compare this interaction to the one above.

Note that in the  $g\phi^3$  example, we had explicit propagators with poles. Those poles are *physical resonances*, which can be experimentally observed, and are necessary for long range forces. In contrast, the derivative interaction above vanishes at small  $s, t, u$ , and does not give rise to long range forces. Instead it is a 'contact interaction', or a very short range force.

But we see that at large energies the cross section from the derivative interaction grows rapidly, as expected from the dimension of the operator

$$\frac{1}{\Lambda^4}(\partial\phi)^4 \quad (2.15.45)$$

which is unimportant at low energies, but large at high energies.

Finally, it's worth taking this further – what if we had a series of interactions like

$$\frac{1}{\Lambda^4}(\partial\phi)^4 + \frac{1}{\Lambda^8}[(\partial_\mu\partial_\nu\phi)(\partial^\mu\partial^\nu\phi)]^2 + \dots \quad (2.15.46)$$

with more and more derivatives? In this case the amplitude would be of order

$$\left(\frac{E}{\Lambda}\right)^4 + \left(\frac{E}{\Lambda}\right)^8 + \dots \quad (2.15.47)$$

So in the limit that  $E \ll \Lambda$  the higher order terms are negligible, but when  $E \gtrsim \Lambda$  the expansion breaks down! In other words, at the short-distance, high-energy cutoff scale  $\Lambda$  our effective field theory description fails. Note that in a theory with only the interaction

$$g\phi^3 \quad (2.15.48)$$

it appears (and is true in  $d \leq 6$  dimensions) that the theory makes sense up to arbitrarily high energies. So maybe it does.

As a final and more complicated example, we can use an

$$\mathcal{L}_{int} = \lambda\phi_1\partial_\mu\phi_2\partial^\mu\phi_3 \quad (2.15.49)$$

Since all fields are different we don't include any combinatorial factors.

If we have  $12 \rightarrow 3 \rightarrow 12$  with  $k_3 = p_1 + p_2 = p'_1 + p'_2$  then we get

$$\begin{aligned} i\mathcal{M} &= (i\lambda)^2(-ip_2^\mu)(ik_{3\mu})\frac{i}{k_3^2}(ip_2^\nu)(-ik_{3\nu}) \\ &= -i\lambda^2 \frac{(p_1 \cdot p_2 + (p_2)^2)(p'_1 \cdot p'_2 + (p'_2)^2)}{(p_1 + p_2)^2} \end{aligned} \quad (2.15.50)$$

One can check that integrating by parts does not alter this result, due to momentum conservation at each vertex.



## 2.16 Particles with Spin 1

Particles are defined as states that transform under irreducible, unitary representations of the Poincaré group.

Irreducibility means that no subset of the states transforms only amongst themselves, in any basis. Unitarity means that

$$|\psi\rangle \rightarrow \mathcal{P}|\psi\rangle \quad \text{with} \quad \langle\psi_1|\psi_2\rangle = \langle\psi_1|\mathcal{P}^\dagger\mathcal{P}|\psi_2\rangle \quad (2.16.1)$$

where  $\mathcal{P}$  is some symmetry (Lorentz, in this case) transformation. This is important for the probabilistic interpretation of quantum mechanics.

We know some obvious representations of the Poincaré group, such as scalars, vectors, tensors, etc, but all but scalars are not unitary. In fact, there are no finite dimensional unitary representations of the Poincaré group. Let us see this in some simple examples, where unitarity and Lorentz invariance clash.

### 2.16.1 Unitarity vs Lorentz Invariance

Let's try to use a 4-vector as a quantum mechanical state. This means we have

$$|\psi\rangle = c_0|V_0\rangle + c_1|V_1\rangle + c_2|V_2\rangle + c_3|V_3\rangle \quad (2.16.2)$$

and the norm of the state is

$$\langle\psi|\psi\rangle = |c_0|^2 + |c_1|^2 + |c_2|^2 + |c_3|^2 > 0 \quad (2.16.3)$$

The problem here is that this isn't just  $c_\mu c^\mu$ , which has both  $+$  and  $-$  signs, since the norm has to be positive definite. But this means that it isn't Lorentz invariant, so for example if we start with

$$|\psi\rangle = |V_0\rangle \quad (2.16.4)$$

then under a boost, we get

$$|\psi_\beta\rangle = \cosh \beta |V_0\rangle + \sinh \beta |V_1\rangle \quad (2.16.5)$$

with a norm

$$\langle\psi_\beta|\psi_\beta\rangle = \cosh^2 \beta + \sinh^2 \beta \neq 1 \quad (2.16.6)$$

This is because the boost matrix is not unitary.

If we modify the norm to be Lorentz invariant, so that

$$\langle\psi|\psi\rangle = c_\mu c^\mu = |c_0|^2 - |c_1|^2 - |c_2|^2 - |c_3|^2 \quad (2.16.7)$$

then we inevitably have both positive and negative norm states, and we lose a sensible probabilistic interpretation.

There are two things we can fix

- It turns out that a four vector  $V_\mu$  contains both a spin 0 and a spin 1 part, with 1 and 3 degrees of freedom, respectively, so that can help with the norm.
- We can also improve this example by choosing a basis of states that is not independent of spacetime, but instead has polarization vectors  $\epsilon_\mu(p)$ , so that the basis of states depends on the momentum of the particles.

This will give the *infinite dimensional representations* we need. But one of the easiest ways to discover what we need is to just study free fields with spin.

### 2.16.2 Massive Spin 1 Particles

Even at the classical level, we can demand both Poincaré invariance and *positivity of energy*. So we can start by studying field theories on these terms.

We already know that our free scalar field theory gives us positive energies, and describes scalar (spinless) particles, which are bosons. Let's move on to vectors  $A_\mu$ , which are the simplest we can study. The simplest thing one might guess is

$$\mathcal{L} = -\frac{1}{2}\partial_\nu A_\mu \partial^\nu A^\mu + \frac{1}{2}m^2 A_\mu A^\mu \quad (2.16.8)$$

The equations of motion are

$$(\square + m^2)A_\mu = 0 \quad (2.16.9)$$

In fact this is just 4 copies of the original scalar equation of motion! We have just found 4 decoupled scalars, or the representation  $1 \oplus 1 \oplus 1 \oplus 1$ .

Note that the energy density of this theory is actually sick, because it is

$$E = -\frac{1}{2} \left[ (\partial_t A_0)^2 + (\vec{\nabla} A_0)^2 + m^2 A_0^2 \right] + \frac{1}{2} \left[ (\partial_t A_i)^2 + (\vec{\nabla} A_i)^2 + m^2 A_i^2 \right] \quad (2.16.10)$$

Thus this theory is not physically sensible, because the energy density is not positive definite.

So are these really 4 scalars or one 4-vector? It's worth pausing to note that we don't get to decide what symmetries the theory has. The theory knows (ie the symmetries are determined by the structure of the Lagrangian, not by what we were thinking when we wrote that Lagrangian down), and its EoM and matrix elements will have any and all symmetries that are present. We don't have to know the symmetries ahead of time – for example, the development of special relativity was all about *realizing* that Electrodynamics had the Lorentz symmetries.

Fortunately this isn't the end of the story; as you know, there's another term that we can consider, giving us the general lagrangian

$$\mathcal{L} = \frac{a}{2} A_\mu \square A^\mu + \frac{b}{2} A_\mu \partial^\mu \partial^\nu A_\nu + \frac{1}{2} m^2 A_\mu A^\mu \quad (2.16.11)$$

Note that we have  $\partial_\mu$  contracted with  $A^\mu$ . Since we know that  $\partial_\mu$  must transform as a 4-vector, this should force  $A_\mu$  to transform that way too. This means we should have the degrees of freedom in a  $3 \oplus 1$  representation of spin.

The equations of motion are

$$a\Box A_\mu + b\partial_\mu\partial_\nu A^\nu + m^2 A_\mu = 0 \quad (2.16.12)$$

Note that if we take  $\partial^\mu$  of this we get

$$[(a+b)\Box + m^2](\partial_\mu A^\mu) = 0 \quad (2.16.13)$$

Now we see that something special happens when  $a = -b$ , because then we have

$$\partial_\mu A^\mu = 0 \quad (2.16.14)$$

as long as  $m^2 \neq 0$ . This is a Lorentz-invariant condition that removes the scalar degree of freedom. Thus the Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu \quad (2.16.15)$$

The EoM imply

$$(\Box + m^2)A_\mu = 0 \quad \text{and} \quad \partial^\mu A_\mu = 0 \quad (2.16.16)$$

and we have only a single massive spin 1 particle (although we haven't made that obvious yet).

Now let's look at positivity of energy. The energy momentum tensor is

$$\begin{aligned} T_{\mu\nu} &= \frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\alpha)}\partial_\nu A_\alpha - g_{\mu\nu}\mathcal{L} \\ &= -F_{\mu\alpha}\partial_\nu A^\alpha + g_{\mu\nu}\left(\frac{1}{4}F_{\alpha\beta}^2 - \frac{1}{2}m^2 A_\alpha^2\right) \end{aligned} \quad (2.16.17)$$

To simplify this we can use the fact that

$$-\frac{1}{4}F_{\alpha\beta}^2 = \frac{1}{2}(E^2 - B^2) \quad (2.16.18)$$

Then the energy is

$$\begin{aligned} T_{00} &= -(\partial_0 A_\alpha - \partial_\alpha A_0)\partial_0 A_\alpha - \frac{1}{2}(E^2 - B^2) - \frac{1}{2}m^2 A_\alpha^2 \\ &= \frac{1}{2}(E^2 + B^2) + \partial_i A_0(\partial_0 A_i - \partial_i A_0) - \frac{1}{2}m^2 A_0^2 + \frac{1}{2}m^2 A_i^2 \end{aligned} \quad (2.16.19)$$

which can be written as

$$\begin{aligned} T_{00} &= \frac{1}{2}(E^2 + B^2) + \frac{1}{2}m^2(A_0^2 + A_i^2) \\ &\quad + A_0\partial_0(\partial_\mu A^\mu) - A_0(\Box + m^2)A_0 + \partial_i A_0 F_0^i \end{aligned} \quad (2.16.20)$$

The first line is positive. The first two terms on the second line vanish on the EoM, and the last term is a total derivative, so in fact we see that the energy is positive. Thus the ‘Proca Lagrangian’ provides a classical field theory with positive energy.

Now we can find some explicit solutions to the classical EoM. As usual, we work in Fourier space with

$$A_\mu(x) = \sum_i \int \frac{d^3p}{(2\pi)^3} \tilde{a}_i(\vec{p}) \epsilon_\mu^i(p) e^{ipx} \quad (2.16.21)$$

where  $p_0 = E_p = \sqrt{m^2 + \vec{p}^2}$ . We want to choose  $\epsilon_\mu^i(p)$  so that  $A_\mu$  automatically satisfies its EoM, which means

$$\partial^\mu A_\mu = p^\mu \epsilon_\mu^i = 0 \quad (2.16.22)$$

which means there are three polarization vectors labeled by  $i = 1, 2, 3$ . We can choose a canonical basis where

$$p^\mu = (E, 0, 0, p_z) \quad (2.16.23)$$

and so there are obvious transverse polarization vectors

$$\epsilon_\mu^1 = (0, 1, 0, 0), \quad \epsilon_\mu^2 = (0, 0, 1, 0) \quad (2.16.24)$$

along with the longitudinal polarization

$$\epsilon_\mu^L = \left( \frac{p_z}{m}, 0, 0, \frac{E}{m} \right) \quad (2.16.25)$$

All of the polarizations are normalized so that

$$\epsilon_\mu \epsilon^\mu = -1 \quad (2.16.26)$$

Together these three polarizations form an irreducible representation of the Poincaré group. Note that the polarizations depend non-trivially on  $p_\mu$ , and there are an infinite number of possible momenta, so this is an infinite dimensional representation.

The structure we have found really does provide the description of massive spin 1 particles such as the W and Z bosons and the  $\rho$  meson. It’s worth noting that at high energies, the longitudinal polarization looks like

$$\epsilon_\mu^L \sim \frac{E}{m} (1, 0, 0, 1) \quad (2.16.27)$$

so if we have a cross section that goes as

$$d\sigma \sim g^2 (\epsilon_L)^2 \sim g^2 \frac{E^2}{m^2} = \frac{E^2}{\left(\frac{m}{g}\right)^2} \quad (2.16.28)$$

then the scattering cross section will blow up at high energies. This is bad and means that perturbation theory is breaking down, much as we have seen with interactions like

$$\frac{1}{\Lambda^4}(\partial\phi)^4 \quad (2.16.29)$$

in a scalar theory. Here we see that the role of  $\Lambda \sim \frac{m}{g}$ . This is an extremely rough version of the argument for why the electroweak theory needed a Higgs boson, or some other UV completion, in order to make sense at high energies.

### Massless Spin 1 Particles

It's natural to obtain a theory for massless spin 1 particles by just taking  $m \rightarrow 0$ , giving the very familiar free Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 \quad (2.16.30)$$

But the question is, what happened to the longitudinal polarization

$$\epsilon_\mu^L = \left( \frac{p_z}{m}, 0, 0, \frac{E}{m} \right) \quad (2.16.31)$$

which seems to diverge when  $m \rightarrow 0$ ? Or, if we factor out an  $m$ , we at least find that  $\epsilon_\mu^L = p_\mu$  exactly in the massless limit. Relatedly, note that we had

$$m^2 \partial_\mu A^\mu = 0 \quad (2.16.32)$$

so that for  $m \neq 0$  we had a constraint, but this also seems to disappear when  $m = 0$ .

To understand what happened, we notice that unlike when  $m \neq 0$ , the Lagrangian now has a *gauge redundancy* under

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \alpha(x) \quad (2.16.33)$$

Two different choices of  $A_\mu$  that differ in this way are physically equivalent. This is often called a ‘gauge symmetry’, although ‘redundancy’ is a better word in many respects.

The equations of motion are the usual

$$\square A_\mu - \partial_\mu(\partial_\nu A^\nu) = 0 \quad (2.16.34)$$

which can be written out more explicitly as

$$-\partial_j^2 A_0 + \partial_t \partial_j A_j = 0 \quad (2.16.35)$$

$$\square A_i - \partial_i(\partial_t A_0 - \partial_j A_j) = 0 \quad (2.16.36)$$

To count the number of physical degrees of freedom we must *eliminate the redundancy*, a procedure more commonly known as gauge-fixing. We can choose  $\partial_j A_j = 0$ , which is Coulomb gauge. Then the  $A_0$  EoM is

$$\partial_j^2 A_0 = 0 \quad (2.16.37)$$

Coulomb gauge is preserved by gauge transformations with  $\partial_j^2 \alpha = 0$ . Since  $A_0$  satisfies  $\partial_j^2 A_0 = 0$ , we can choose  $\alpha$  with  $\partial_t \alpha = -A_0$  so that in this gauge

$$A_0 = 0 \quad (2.16.38)$$

So we've eliminated one DoF.

Now let's eliminate another. The remaining DoF satisfy

$$\square A_i = 0 \quad (2.16.39)$$

which seems to have three modes, but  $\partial_i A_i = 0$ , which means in Fourier space that

$$A_\mu(x) = \int \frac{d^3 p}{(2\pi)^3} \epsilon_\mu(p) e^{ipx} \quad (2.16.40)$$

with  $p^2 = 0$ , and  $p_i \epsilon_i = 0$  and  $\epsilon_0 = 0$  from the gauge choice. The only possibilities in a frame with

$$p^\mu = (E, 0, 0, E) \quad (2.16.41)$$

are obvious transverse polarization vectors

$$\epsilon_\mu^1 = (0, 1, 0, 0), \quad \epsilon_\mu^2 = (0, 0, 1, 0) \quad (2.16.42)$$

which represent linearly polarized light. One can also use the linear combinations

$$\epsilon_\mu^R = (0, 1, i, 0), \quad \epsilon_\mu^L = (0, 1, -i, 0) \quad (2.16.43)$$

which represent circular polarization. These are eigenstates of something called *helicity*, which is the spin projected onto the direction that the particle's moving.

Note that if instead we'd chosen Lorenz gauge with  $\partial_\mu A^\mu = p_\mu \epsilon^\mu = 0$  then we'd have

$$\epsilon_\mu^1 = (0, 1, 0, 0), \quad \epsilon_\mu^2 = (0, 0, 1, 0), \quad \epsilon_\mu^f = (1, 0, 0, 1) \quad (2.16.44)$$

but the last forward polarization is unphysical, as it's pure gauge – it's  $p_\mu \phi$  for some scalar  $\phi$  – and it's also non-normalizable, since  $\epsilon_\mu \epsilon^\mu = 0$ .

For both massive and massless cases, we found a basis of polarization vectors  $\epsilon_\mu(p)$ , making them both infinite dimensional representations. The method we used is called that of *induced representations* because the full representation is induced by the representation of the subgroup that keeps  $p_\mu$  fixed. That subgroup is called *the little group*, and it has finite dimensional representations. In the massive case the little group is  $SO(3)$ , while in the massless case it's  $ISO(2)$ , the isometry group of the 2d Euclidean plane.

## 2.17 Covariant Derivatives and Scalar QED

If we want to continue to study massless spin 1 particles, *and we want to write down local interactions for them*, then we need to preserve the gauge redundancy when we write down Lagrangians.

First we can consider theories involving only photons. We cannot use  $A_\mu$  by itself, except in the gauge invariant combination  $F_{\mu\nu}$ . Thus we can write down an interaction (in 4d)

$$\frac{1}{\Lambda^4} (F_{\mu\nu}^2)^2 \quad (2.17.1)$$

You might have tried to write  $F_{ab}F^{bc}F^{ad}\eta_{cd}$  but it vanishes. This interaction is irrelevant and becomes very weak at long distances; this isn't QED, although this is how  $\gamma\gamma \rightarrow \gamma\gamma$  at low energies. We use  $\gamma$  to denote photons.

What about coupling the photon to other kinds of particles? To do it in a gauge invariant way, it seems that we cannot use  $A_\mu$ . We could try to write

$$\frac{1}{\Lambda^2} F^{ab} \partial_a \phi \partial_b \phi \quad (2.17.2)$$

but this also vanishes.

What we would really like is to couple directly to  $A_\mu$ . The trick is to imagine that when  $A_\mu \rightarrow A_\mu + \partial_\mu \alpha$  we also transform our other field  $\phi$ . It turns out that we need more than one scalar, which is easiest to write as a complex scalar with

$$\phi(x) \rightarrow e^{-i\alpha(x)} \phi(x) \quad (2.17.3)$$

Note that we cannot do this with a single real scalar, because this transformation rotates the two real components of  $\phi$  into each other. With this transformation rule, modified to  $A_\mu \rightarrow A_\mu + \frac{1}{e} \partial_\mu \alpha$ , the *gauge covariant derivative*

$$D_\mu \phi = (\partial_\mu + ieA_\mu) \phi \quad (2.17.4)$$

transforms to

$$(\partial_\mu + ieA_\mu + i\alpha(x)) e^{-i\alpha(x)} \phi(x) = e^{-i\alpha(x)} (\partial_\mu + ieA_\mu) \phi \quad (2.17.5)$$

because the derivative of the exponential cancels with the shift of  $A_\mu$ . The whole point of a covariant derivative is that it transforms in the same way as the original field. This means that the lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + (D_\mu \phi^\dagger)(D^\mu \phi) - m^2 \phi^\dagger \phi \quad (2.17.6)$$

is gauge covariant, and all of the interactions are classically dimensionless. This is the lagrangian for *scalar QED*. For example, this is how charged pions interact with the photon.

In general, a field with charge  $q_n$  transforms under gauge transformations as

$$\phi_n \rightarrow e^{iq_n \alpha(x)} \phi_n \quad (2.17.7)$$

and so they couple with a covariant derivative

$$D_\mu \phi_n = (\partial_\mu + iq_n e A_\mu) \phi_n \quad (2.17.8)$$

The overall interaction strength is  $eq_n$ ; above we chose a charge of 1 for our scalar field. For our electromagnetic force, the famous parameter

$$\alpha \equiv \frac{e^2}{4\pi} \quad (2.17.9)$$

which is approximately  $\frac{1}{137}$ , although as you know, this depends on the distance scale at which we measure it.

A final point that's worth mentioning – you might have guessed that a good way to get a gauge invariant Lagrangian is simply to write

$$\mathcal{L}_{int} = e A_\mu J^\mu \quad (2.17.10)$$

This is a good idea because under a gauge transformation we have

$$e A_\mu J^\mu \rightarrow e A_\mu J^\mu + e \partial_\mu \alpha J^\mu = e A_\mu J^\mu - e \alpha \partial_\mu J^\mu \quad (2.17.11)$$

and so if the current  $J^\mu$  is conserved then the interaction is gauge invariant. This is a good idea, but note that adding this term to the Lagrangian changes the current, and so (in the case of scalar QED) we have to include higher order terms, namely a  $\phi^\dagger \phi A_\mu A^\mu$  type term. In spinor QED (the usual QED) the  $A_\mu J^\mu$  coupling is all we need.

### 2.17.1 Quantum Massive Spin 1

Now let's quantize our theories of massive and massless particles (well, at least we'll write down the answer). The quantum field is

$$A_\mu(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \sum_{j=1}^3 \left[ \epsilon_\mu^j(p) a_{\vec{p},j} e^{-ip \cdot x} + \epsilon_\mu^{j\dagger}(p) a_{\vec{p},j}^\dagger e^{ip \cdot x} \right] \quad (2.17.12)$$

where the sum is over the 3 types of polarization vector, and the 3 corresponding creation and annihilation operators for the 3 degrees of freedom (polarization vectors). In other words, to specify the asymptotic particle states we write

$$a_{p,j}^\dagger |0\rangle = \frac{1}{\sqrt{2E_p}} |p, \epsilon^j\rangle \quad (2.17.13)$$

or in other words

$$\langle 0 | A_\mu(x) | p, \epsilon^j \rangle = \epsilon_\mu^j e^{-ip \cdot x} \quad (2.17.14)$$

so that  $A_\mu$  creates a particle with an appropriate linear combination of polarization vectors.



This quantization of  $A_\mu(x)$  is correct because it obeys the correct EoM and has the correct canonical commutation relations.

Note that the fourth polarization, orthogonal to  $\epsilon^1$ ,  $\epsilon^2$ , and  $\epsilon^L$  would be

$$\epsilon_S = \frac{p^\mu}{m} \propto \partial_\mu \alpha(x) \quad (2.17.15)$$

where the last was written in position space, for some scalar function  $\alpha$ . But this hasn't been included because we do not want to include a scalar degree of freedom in our vector particle. Note that this polarization never mixes with the others under Lorentz transformations because in any Lorentz frame it is just the derivative of a scalar.

We can see that the three polarizations we have included do in fact mix with each other by choosing say  $p^\mu = (m, 0, 0, 0)$ , and looking at the little group, namely Lorentz transformations that leave  $p^\mu$  invariant. These are just the  $SO(3)$  rotations. One can also try this in a more confusing frame, where the particle isn't at rest, but it still works.

In QFT we will be computing correlation functions and scattering amplitudes (or quantum mechanical matrix elements) with the field  $A_\mu$ ; these will depend on the polarization by

$$\mathcal{M} = \epsilon_\mu(p) M^\mu \quad (2.17.16)$$

where  $\epsilon_\mu(p)$  is any linear combination of polarization vectors.  $M^\mu$  must therefore transform as a Lorentz 4-vector, so that  $\mathcal{M}$  is Lorentz invariant. In general,  $\mathcal{M}$  must be Lorentz invariant on the restricted space of polarization vectors.

### 2.17.2 Quantum Massless Spin 1

In the massless case we only have the two transverse polarizations

$$A_\mu(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \sum_{j=1}^2 \left[ \epsilon_\mu^j(p) a_{\vec{p},j} e^{-ip \cdot x} + \epsilon_\mu^{j\dagger}(p) a_{\vec{p},j}^\dagger e^{ip \cdot x} \right] \quad (2.17.17)$$

It's worth writing out the momentum, polarizations, and orthogonal vectors

$$p^\mu = (E, 0, 0, E) \quad (2.17.18)$$

and

$$\epsilon_\mu^1 = (0, 1, 0, 0), \quad \epsilon_\mu^2 = (0, 0, 1, 0) \quad (2.17.19)$$

and the orthogonal vectors

$$\epsilon_f^\mu = (1, 0, 0, 1), \quad \epsilon_b^\mu = (1, 0, 0, -1) \quad (2.17.20)$$

where these are 'forward and backward' vectors.

The reason we wrote these out is that there's a problem with Lorentz transformations. In the massive case we had mixing between the two transverse polarization states and the longitudinal

polarization  $\epsilon_f$ . But nothing very discontinuous happens when we take  $m \rightarrow 0$ , so in the massless case *the transverse polarizations mix with  $\epsilon_f \propto p$  under Lorentz transformations*. In general

$$\epsilon_\mu^1(p) \rightarrow c_{11}(\Lambda)\epsilon_\mu^1(p) + c_{12}(\Lambda)\epsilon_\mu^2(p) + c_{13}(\Lambda)p_\mu \quad (2.17.21)$$

and similarly for  $\epsilon_\mu^2(p)$  under the transformation  $\Lambda$ , where  $c_{ij}(\Lambda)$  are numbers associated with the representation.

This is very important, so it's worthwhile to be explicit, so for example

$$\Lambda_\nu^\mu = \begin{pmatrix} \frac{3}{2} & 1 & 0 & -\frac{1}{2} \\ 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 1 & 0 & \frac{1}{2} \end{pmatrix} \quad (2.17.22)$$

is a Lorentz transformation with  $\Lambda_\nu^\mu p^\nu = p^\mu$ , so it is in the little group which preserves  $p^\mu = (E, 0, 0, E)$ . Note that under this transformation

$$\Lambda_\nu^\mu \epsilon_1^\nu = (1, 1, 0, 1) = \epsilon_1^\mu + \frac{1}{E}p^\mu \quad (2.17.23)$$

and so we see that a polarization vector for a transverse polarization shifts by a term proportional to  $p^\mu$ . Nothing like this happens for massive spin 1. *This fact leads to many constraints on the interactions of massless particles with spin.*

Now consider a matrix element for a massless spin 1 particle. It transforms under Lorentz transformations as

$$\mathcal{M} = \epsilon_\mu M^\mu \rightarrow \epsilon'_\mu M'^\mu + c(\Lambda)p_\mu M'^\mu \quad (2.17.24)$$

for some  $c(\Lambda)$ , where  $M'^\mu = \Lambda_\nu^\mu M^\nu$  and  $\epsilon'$  is a combination of transverse polarizations. So we have a problem unless

$$p_\mu M^\mu = 0 \quad (2.17.25)$$

which is required to guarantee that we don't produce unphysical particle states while preserving Lorentz invariance. This is known as a *Ward Identity*, and it's a very important constraint on physical amplitudes.

Note that the transformation of polarization states is closely related to gauge invariance, since under a Lorentz transformation

$$\epsilon_\mu \rightarrow \epsilon'_\mu + c p_\mu \quad \sim \quad A_\mu(p) \rightarrow A_\mu(p) + p_\mu \alpha(p) \quad (2.17.26)$$

The arguments that one gives in gauge theories (such as QED) for the Ward identity are based on gauge invariance.

On a more general level, the Ward identity is a constraint that arises due to the tension between *Quantum Mechanical Unitarity, Lorentz Invariance, and Locality*. These are all physical principles

that actually mean something; we introduce the unphysical gauge redundancy to make it easier to satisfy the requirements of all three at once.

Note that we could completely get rid of the gauge redundancy by just using the electric and magnetic fields directly as observables, so that  $F_{\mu\nu}$  was the ‘elementary quantum field’. But then we would have to write the interactions in terms of

$$A_\mu \sim \frac{1}{\square} \partial_\nu F_\mu^\nu \quad (2.17.27)$$

because the interactions that give rise to long-range (electromagnetic) forces involve this object. Then the theory would appear *non-local*, because  $1/\square \sim 1/p^2$  is dominated at long distances (e.g. on a lattice it’s not a nearest neighbor type interaction), and it would be very difficult to analyze our theory. So we use a gauge redundant description so that we can simultaneously have locality along with Unitarity and Lorentz invariance. Some contemporary research involves only studying scattering processes directly, in which case locality isn’t as manifest, and one can avoid using gauge redundancy.

### 2.17.3 (Guessing) The Photon Propagator

To compute Feynman diagrams involving the photon, e.g. in QED and scalar QED, we will need to know the photon propagator, defined as

$$\langle 0|T\{A^\mu(x)A^\nu(y)\}|0\rangle = i \int \frac{d^4p}{(2\pi)^4} e^{ip \cdot (x-y)} \Pi^{\mu\nu}(p) \quad (2.17.28)$$

and evaluated in the free theory.

This is made complicated by the Lorentz structure, and conceptually non-trivial due to the fact that the photon only has 2 polarization states, while  $\Pi^{\mu\nu}$  is a 4-by-4 Lorentz tensor. Note that the momentum space EoM are

$$(-p^2 g_{\mu\nu} + p_\mu p_\nu) A^\mu = J_\nu \quad (2.17.29)$$

and so we would like to write

$$\Pi \sim \frac{1}{-p^2 g_{\mu\nu} + p_\mu p_\nu} \quad (2.17.30)$$

but in fact that tensor is not invertible!

This problem is solved (in a way that can be proven to be correct using path integrals) by eliminating the gauge invariance. One could just choose a gauge, but then one has to keep track of the gauge constraint, so an easier way is to add to the Lagrangian a gauge fixing term

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 - \frac{1}{2\xi} (\partial_\mu A^\mu)^2 - J_\mu A^\mu \quad (2.17.31)$$

where  $\xi$  is just some parameter, which, if we wanted, we could view as a lagrange multiplier to be integrated over. With this term, the EoM are

$$\left( -p^2 g_{\mu\nu} + \left( 1 - \frac{1}{\xi} \right) p_\mu p_\nu \right) A^\mu = J_\nu \quad (2.17.32)$$

This can be inverted to give a propagator

$$i\Pi^{\mu\nu} = -i \frac{g_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2}}{p^2 + i\epsilon} \quad (2.17.33)$$

which we can verify simply by multiplication. This is the momentum space Feynman propagator in ‘covariant’ or ‘ $R_\xi$ -gauge’. The signs have been chosen so that there is a

$$\frac{i}{p^2 + i\epsilon} \quad (2.17.34)$$

for the spatial components  $A_i$  of the field, which are the physical polarization states.

There are various convenient choices of the parameter  $\xi$  we have introduced, called gauge choices, such as  $\xi = 1$  (Feynman - 't Hooft),  $\xi = 0$  (Lorenz), and  $\xi \rightarrow \infty$  (Unitarity). But the final answer must be gauge invariant, which means it must be independent of  $\xi$  when we add up all the various Feynman diagrams for a physical process (as we’ll do soon).

## 2.18 Scattering and Ward Identities in Scalar QED

We already wrote down the lagrangian density for scalar QED, namely

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + (D_\mu\phi^\dagger)(D^\mu\phi) - m^2\phi^\dagger\phi \quad (2.18.1)$$

with  $D_\mu\phi \equiv \partial_\mu\phi + ieA_\mu\phi$ .

We already studied the theory of a complex scalar field with a global symmetry

$$\phi \rightarrow e^{-i\alpha}\phi \quad (2.18.2)$$

with constant  $\alpha$ . This is a  $U(1)$  global symmetry charge with a conserved current  $J_\mu$ . Since at first order in the coupling  $e$  we have an interaction  $eA_\mu J^\mu$ , the global symmetry charge is the electric charge.

Of course we could write  $\phi = \phi_1 + i\phi_2$  and quantize  $\phi_1$  and  $\phi_2$  separately, but it makes infinitely more sense to study charge eigenstates, that is

$$\phi = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p e^{-ip \cdot x} + b_p^\dagger e^{ip \cdot x}) \quad (2.18.3)$$

and

$$\phi^\dagger = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^\dagger e^{ip \cdot x} + b_p e^{-ip \cdot x}) \quad (2.18.4)$$

These fields transform nicely under the global symmetry, which is proportional to the gauge redundancy

$$A_\mu(x) \rightarrow A_\mu(x) + \frac{1}{e}\partial_\mu\alpha(x), \quad \phi(x) \rightarrow e^{-i\alpha(x)}\phi(x) \quad (2.18.5)$$

The point here is that  $a_p^\dagger$  and  $b_p^\dagger$  create two different species of particles with identical mass and opposite charge, aka particles and anti-particles.

Just as we derived Feynman rules for our various scalar theories, we obtain analogous rules for scalar QED. Let's study the rules for the S-Matrix. As usual, we have the scalar propagator

$$\langle \phi^\dagger(p) \phi(-p) \rangle = \frac{i}{p^2 - m^2 + i\epsilon} \quad (2.18.6)$$

and as we guessed/derived above, we have the photon propagator

$$\frac{-i}{p^2 + i\epsilon} \left[ g_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right] \quad (2.18.7)$$

where  $\xi$  parameterizes the set of covariant gauges.

The simplest interaction to understand is

$$\mathcal{L}_{int} \supset e^2 A_\mu A^\mu \phi^\dagger \phi \quad (2.18.8)$$

and this contributes a Feynman rule

$$2ie^2 g_{\mu\nu} \quad (2.18.9)$$

where the factor of 2 comes from  $A_\mu A^\mu$ .

The slightly subtler interaction is the one with the electromagnetic current

$$\mathcal{L}_{int} \supset -ie A_\mu [\phi^\dagger (\partial^\mu \phi) - (\partial^\mu \phi^\dagger) \phi] \quad (2.18.10)$$

The subtlety comes from the signs associated with the  $\partial_\mu \rightarrow ip_\mu$  on  $\phi$  versus  $\phi^\dagger$  when they create or destroy particles, respectively. The point is that *the sign is associated with the charge of the electromagnetic current and its direction of propagation*. Obviously all vertices must conserve charge. This means that:

- Annihilate a  $\pi^-$  and create a  $\pi^-$  gives

$$ie(-p_\mu^1 - p_\mu^2) \quad (2.18.11)$$

We draw particle flow arrows and momentum flow arrows for this, where particle flow indicates how charge goes around the diagram.

- Annihilate a  $\pi^+$  and create a  $\pi^+$

$$ie(p_\mu^1 + p_\mu^2) \quad (2.18.12)$$

As expected, this better have the opposite sign as the case above, since the charge has flipped sign.

- Annihilate  $\pi^-$  and  $\pi^+$  gives

$$ie(-p_\mu^1 + p_\mu^2) \quad (2.18.13)$$

- Create  $\pi^+$  and  $\pi^-$

$$ie(p_\mu^1 - p_\mu^2) \quad (2.18.14)$$

If we write the momentum as it flows with respect to the charge flow arrows then it's always

$$-ie(p_\mu^1 + p_\mu^2) \quad (2.18.15)$$

because we view the negatively charged particles as ‘particles’, since we view electrons as ‘particles’ and not ‘anti-particles’. There’s an overall sign here that’s purely conventional.

We also have to deal with the external states. For the scalars the external states just get a factor of 1, as usual. For external photons we have the field

$$A_\mu(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \sum_{j=1}^2 \left[ \epsilon_\mu^j(p) a_{\vec{p},j} e^{-ip \cdot x} + \epsilon_\mu^{j\dagger}(p) a_{\vec{p},j}^\dagger e^{ip \cdot x} \right] \quad (2.18.16)$$

This is the same thing as a scalar field, except with some polarization vectors

$$\epsilon_\mu^i(p) \quad (2.18.17)$$

included in the field, so the LSZ prescription just gets modified by including a factor of this polarization vector for incoming photons, and

$$\epsilon_\mu^{i\dagger}(p) \quad (2.18.18)$$

for the outgoing photons.

As an example one can write down a diagram contributing to the  $\pi^- \gamma \rightarrow \pi^- \gamma$  scattering amplitude

$$i\mathcal{M} = (-ie)\epsilon_\mu^1(p_2^\mu + k^\mu) \frac{i}{k^2 - m^2 + i\epsilon} (-ie)(p_3^\nu + k^\nu) \epsilon_\nu^{4\dagger} \quad (2.18.19)$$

where  $p_2$  and  $p_3$  are the scalar incoming and outgoing momenta and  $k = p_1 + p_2$ .

### 2.18.1 Example: ‘Moller Scattering’ of $\pi^- \pi^- \rightarrow \pi^- \pi^-$

Here there are two diagrams, a  $t$ -channel and a  $u$ -channel contribution. There’s no  $s$ -channel contribution because two negatively charged particles cannot annihilate into a photon, due to charge conservation.

The  $t$ -channel diagram is

$$i\mathcal{M}_t = (-ie)(p_1^\mu + p_3^\mu) \frac{-i \left[ g_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right]}{k^2} (-ie)(p_2^\mu + p_4^\mu) \quad (2.18.20)$$

where  $k = p_3 - p_1$ . Note that

$$k \cdot (p_1 + p_3) = p_3^2 - p_1^2 = 0 \quad (2.18.21)$$

and so this can be simplified to give

$$i\mathcal{M}_t = e^2 \frac{(p_1 + p_3) \cdot (p_2 + p_4)}{t} \quad (2.18.22)$$

Note that the  $\xi$  dependence has already vanished in this one diagram. That only had to happen after we summed up both diagrams, but since they have quite different kinematics the  $\xi$  dependence had to cancel independently in each.

The other diagram gives

$$\mathcal{M}_u = e^2 \frac{(p_1 + p_4) \cdot (p_2 + p_3)}{u} \quad (2.18.23)$$

The cross section can also be easily computed as

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{e^4}{64\pi^2 E_{CM}^2} |\mathcal{M}_t + \mathcal{M}_u|^2 \\ &= \frac{\alpha^2}{4s} \left[ \frac{s-u}{t} + \frac{s-t}{u} \right]^2 \end{aligned} \quad (2.18.24)$$

### 2.18.2 More Comments on Gauge Invariance and the Ward Identity

A crucial aspect of scattering amplitudes and other physical observables in QED is that they are gauge invariant. This has at least two guises; one is the Ward identity and another is the fact that even the photon propagator is

$$i\Pi_{\mu\nu} = \frac{-i \left[ g_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right]}{k^2} \quad (2.18.25)$$

actually all amplitudes are independent of  $\xi$ . This means that if we have some internal photon propagator then we could write the amplitude as

$$\Pi^{\mu\nu} M_{\mu\nu} \quad (2.18.26)$$

and so gauge invariance requires that

$$p^\mu p^\nu M_{\mu\nu} = 0 \quad (2.18.27)$$

We can give a general proof of gauge invariance and the Ward identity using the path integral, but for now let's see how it (the Ward identity) arises for a particular example where we have  $\pi^+ \pi^- \rightarrow \gamma \gamma$ .

One diagram is

$$i\mathcal{M}_t = (-ie)^2 \frac{i(2p_1^\mu - p_3^\mu)(p_4^\mu - 2p_2^\mu)}{(p_1 - p_3)^2 - m^2} \epsilon_{3\mu}^\dagger \epsilon_{4\nu}^\dagger \quad (2.18.28)$$

which can be simplified (using only that the pions are on-shell, but not the photons, and also not using any gauge condition) to

$$\mathcal{M}_t = e^2 \frac{(p_3 \cdot \epsilon_3^\dagger - 2p_1 \cdot \epsilon_3^\dagger)(p_4 \cdot \epsilon_4^\dagger - 2p_2 \cdot \epsilon_4^\dagger)}{p_3^2 - 2p_3 \cdot p_1} \quad (2.18.29)$$

Similarly there is a  $u$ -channel diagram

$$\mathcal{M}_u = e^2 \frac{(p_3 \cdot \epsilon_3^\dagger - 2p_2 \cdot \epsilon_3^\dagger)(p_4 \cdot \epsilon_4^\dagger - 2p_1 \cdot \epsilon_4^\dagger)}{p_3^2 - 2p_3 \cdot p_2} \quad (2.18.30)$$

Something nice happens when we check the Ward identity, which means replacing  $\epsilon_3 \rightarrow p_3$ . In this limit

$$\mathcal{M}_t \rightarrow e^2(p_4 \cdot \epsilon_4^\dagger - 2p_2 \cdot \epsilon_4^\dagger) \quad (2.18.31)$$

and similarly for the  $u$ -channel, so we get

$$\mathcal{M}_t + \mathcal{M}_u = e^2(p_4 \cdot \epsilon_4^\dagger - 2p_2 \cdot \epsilon_4^\dagger + p_4 \cdot \epsilon_4^\dagger - 2p_1 \cdot \epsilon_4^\dagger) = 2e^2 \epsilon_4^\dagger \cdot (p_4 - p_2 - p_1) \quad (2.18.32)$$

This isn't zero, but we've left out the 4-pt vertex, which gives

$$i\mathcal{M}_4 = 2ie^2 \epsilon_3^\dagger \cdot \epsilon_4^\dagger \quad (2.18.33)$$

Thus adding it in with  $\epsilon_3 \rightarrow p_3$  we find the total

$$\mathcal{M}_t + \mathcal{M}_u + \mathcal{M}_4 = 2e^2 \epsilon_4^\dagger \cdot (p_4 + p_3 - p_2 - p_1) = 0 \quad (2.18.34)$$

due to momentum conservation.

We did not use that photons are on-shell or massless! Thus the identity we just derived would work even if the external photons weren't physical particle states. So we can put this identity into any larger set of Feynman diagrams and prove  $\xi$ -independence. One can use such arguments to prove gauge invariance diagrammatically, but it's tedious.

## 2.19 Spinors and QED

We've mentioned the fact that the Lorentz transformations form a group, and in fact this is a Lie Group. The book has some nice discussion of the background for this in chapter 10. We've mostly discussed Lorentz transformations as  $4 \times 4$  matrices that act on 4-vectors like  $p_\mu, x_\mu, A_\mu, \partial_\mu$  etc. However, we can view the group in an abstract form, as

$$g(\theta, \beta) = \exp [i\theta_i J^i + i\beta_i K^i] \quad (2.19.1)$$

This is an abstract Lie group element – a general finite dimensional Lorentz transformation – but the infinitesimal generators  $J^i$  and  $K^i$  in the exponent are elements of the Lie algebra. In order to generate the group, they must obey the Lie algebra

$$[J_i, J_j] = i\epsilon_{ijk} J_k \quad (2.19.2)$$

$$[J_i, K_j] = i\epsilon_{ijk} K_k \quad (2.19.3)$$

$$[K_i, K_j] = -i\epsilon_{ijk} J_k \quad (2.19.4)$$



In any *representation* of the abstract Lorentz group, these are just the commutation relations of the matrices associated with the representation. The usual  $4 \times 4$  matrix representation is the fundamental representation, but there are many others. Note that

$$V^{\mu\nu} = \begin{pmatrix} 0 & K_1 & K_2 & K_3 \\ -K_1 & 0 & J_3 & -J_2 \\ -K_2 & -J_3 & 1 & J_1 \\ -K_3 & J_2 & -J_1 & 0 \end{pmatrix} \quad (2.19.5)$$

and we can write the commutation relations directly in these terms as

$$[V^{\mu\nu}, V^{\rho\sigma}] = i(g^{\nu\rho}V^{\mu\sigma} + g^{\mu\sigma}V^{\nu\rho} - g^{\nu\sigma}V^{\mu\rho} - g^{\mu\rho}V^{\nu\sigma}) \quad (2.19.6)$$

although this isn't especially useful. As an example of a very different representation, note that

$$L^{\mu\nu} = i(x^\mu \partial^\nu - x^\nu \partial^\mu) \quad (2.19.7)$$

generate Lorentz transformations on the infinite dimensional vector space of functions. So these are infinite dimensional matrices.

The most useful way to write the Lorentz algebra is using

$$J_i^+ \equiv \frac{1}{2}(J_i + iK_i), \quad J_i^- \equiv \frac{1}{2}(J_i - iK_i) \quad (2.19.8)$$

In this basis the algebra is just

$$[J_i^\pm, J_j^\pm] = i\epsilon_{ijk}J_k^\pm \quad (2.19.9)$$

$$[J_i^+, J_j^-] = 0 \quad (2.19.10)$$

So we just have two copies of  $so(3) = sl(2, R) = su(2)$ . This can be written as

$$so(1, 3) = su(2) \oplus su(2) \quad (2.19.11)$$

Finding general representations of the Lorentz group is easy, because we already know the representations of  $su(2)$  from quantum mechanics. Note that the fundamental representation of  $su(2)$  just involves the  $2 \times 2$  pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.19.12)$$

which is how we end up with spinor representations of the Lorentz group.

As you know, representations of  $su(2)$  are just labeled by a half-integral spin  $s$ ; this representation has dimension  $2s + 1$ . Since the Lorentz group involves two copies of  $su(2)$ , its representations are labeled by two half-integers  $(A, B)$ , and have dimension  $(2A + 1)(2B + 1)$ , so this is also the number of degrees of freedom included. Note, however, that *neither the  $J^+$  nor the  $J^-$  are the rotation*

*generators*. The rotation generators are  $\vec{J} = \vec{J}^+ + \vec{J}^-$ . So you should not confuse  $A$  or  $B$  with the spin of the particle!

Every representation  $(A, B)$  of the Lorentz group provides (usually reducible) representations of the rotations  $\vec{J}$ . For example

$$\left(\frac{1}{2}, 0\right) = \left(\frac{1}{2}\right)_{\vec{J}} \quad (2.19.13)$$

$$\left(\frac{1}{2}, \frac{1}{2}\right) = (1 \oplus 0)_{\vec{J}} \quad (2.19.14)$$

$$(1, 1) = (2 \oplus 1 \oplus 0)_{\vec{J}} \quad (2.19.15)$$

In fact any  $(A, B)$  involves spins  $j = A + B, A + B - 1, \dots, |A - B|$ .

The relevance of this for QFT is that we construct our lagrangians out of fields such as  $A_\mu(x)$  and  $\psi(x)$  that have definite transformations under the Lorentz group. However, particles transform under irreducible representations of the Poincaré group (Lorentz plus translations), which have spins associated with the little group of transformations that are left over after we fix the momentum. For massive particles, the little group always has algebra  $so(3) = su(2)$ . So the decomposition of  $(A, B)$  under the rotations  $\vec{J}$  tells us the possible spins of particles in a quantum field with a Lorentz group irreducible representation  $(A, B)$ .

It's worth noting that exponentiating the Lie Algebra gives the universal cover of the group, e.g.  $SU(2)$  and not  $SO(3)$ . The universal cover of the Lorentz group is  $SL(2, C)$ .

### 2.19.1 Spinor Representations

There are two complex representations of the Lorentz group, namely  $(\frac{1}{2}, 0)$  and  $(0, \frac{1}{2})$ . They both transform as spin 1/2 under rotations, which act using the Pauli matrices, since these satisfy the correct algebra

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \quad (2.19.16)$$

Another useful and important fact is that

$$\{\sigma_i, \sigma_j\} = \sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij} \quad (2.19.17)$$

where this is an anti-commutator.

Thus we have the representations

$$\left(\frac{1}{2}, 0\right) : \quad \vec{J} = \frac{1}{2}\vec{\sigma}, \quad \vec{K} = i\frac{1}{2}\vec{\sigma} \quad (2.19.18)$$

and

$$\left(0, \frac{1}{2}\right) : \quad \vec{J} = \frac{1}{2}\vec{\sigma}, \quad \vec{K} = -i\frac{1}{2}\vec{\sigma} \quad (2.19.19)$$

Thus the rotations are Hermitian and the boosts are anti-Hermitian, ie  $\vec{K}^\dagger = -\vec{K}$ . Since the generators are complex conjugates of each other in these two representations, we say that they are complex-conjugate representations.

These two representations are called *spinors*, and the  $(1/2, 0)$  case are called *left-handed spinors*, while the  $(0, 1/2)$  case are called *right-handed spinors*. The right handed spinors (which are complex 2-vectors) transform as

$$\psi_R \rightarrow e^{\frac{1}{2}(i\theta_i\sigma_i + \beta_j\sigma_j)}\psi_R \quad (2.19.20)$$

while

$$\psi_L \rightarrow e^{\frac{1}{2}(i\theta_i\sigma_i - \beta_j\sigma_j)}\psi_L \quad (2.19.21)$$

where the  $\theta_i$  and  $\beta_j$  are real numbers. Note that  $\psi_{L/R}^\dagger$  transforms with the opposite sign for  $\theta_i$ , but the same sign for  $\beta_j$ , and with the representation matrices on the other side of the vector (adjointed), as we get simply by daggering these equations.

### 2.19.2 Unitary Representations and Lorentz Invariant Lagrangians

As we just saw, these spinor representations aren't unitary, since  $\vec{K}^\dagger = -\vec{K}$ . This will always happen if we have unitary representations of  $su(2) \oplus su(2)$ , since  $\vec{J}_\pm = \frac{1}{2}(\vec{J} \pm i\vec{K})$ . Thus we see (as claimed before) that there are no finite dimensional unitary representations of the Lorentz group.

We can avoid this problem by studying infinite dimensional unitary representations, where we account for the fact that particles have momenta  $p_\mu$  which transform under the Lorentz group. We used this technique for spin 1, now let's apply it for spin  $\frac{1}{2}$ .

We had fun last time with guess work, so let's try that again. We can start by guessing a Lagrangian like

$$\psi_R^\dagger \square \psi_R + m^2 \psi_R^\dagger \psi_R \quad (2.19.22)$$

But actually this isn't even Lorentz invariant! The problem is that  $\psi_R^\dagger$  and  $\psi_R$  both transform by  $e^{\beta_i\sigma_i}$  and so there's no cancellation.

If we include two spinors, an  $L$  and an  $R$ , then we can write down terms like

$$\psi_L^\dagger \psi_R \quad (2.19.23)$$

With this term we have a cancellation of both the  $e^{i\theta_i\sigma_i}$  terms and also the  $e^{\beta_i\sigma_i}$  terms, since the  $L$  and  $R$  spinors transform oppositely. So we can write down a term called 'Dirac mass' as

$$m \left( \psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L \right) \quad (2.19.24)$$

However, if we try to write down a kinetic term like

$$\mathcal{L} = \psi_L^\dagger \square \psi_R + \psi_R^\dagger \square \psi_L \quad (2.19.25)$$

it's both Lorentz invariant and real, but actually if we write  $\psi_R = (\psi_1, \psi_2)$  we would just find that this is the Lagrangian for a bunch of scalar degrees of freedom! Again, we don't tell the theory what the particle content is (and how it transforms), it tells us, and this is the wrong theory. What we need is some way to force  $\psi$  to transform as a spinor.

To proceed, let's try looking at the fermion bilinear

$$\psi_R^\dagger \sigma_i \psi_R \quad (2.19.26)$$

How does this transform? Well, infinitesimally

$$\delta \psi_R = \frac{1}{2} (i\theta_i \sigma_i + \beta_j \sigma_j) \psi_R \quad (2.19.27)$$

and so we have

$$\begin{aligned} \delta \left( \psi_R^\dagger \sigma_i \psi_R \right) &= \frac{1}{2} \psi_R^\dagger [\sigma_i \sigma_j (i\theta_j + \beta_j) + (-i\theta_j + \beta_j) \sigma_j \sigma_i] \psi_R \\ &= \frac{1}{2} \beta_j \psi_R^\dagger [\sigma_i \sigma_j + \sigma_j \sigma_i] \psi_R + \frac{1}{2} i\theta_j \psi_R^\dagger [\sigma_i \sigma_j - \sigma_j \sigma_i] \psi_R \\ &= \beta_i \psi_R^\dagger \psi_R + \theta_j \epsilon_{ijk} \psi_R^\dagger \sigma_k \psi_R \end{aligned} \quad (2.19.28)$$

Note that the object  $\psi_R^\dagger \psi_R$  transforms as

$$\delta \left( \psi_R^\dagger \psi_R \right) = \beta_i \psi_R^\dagger \sigma_i \psi_R \quad (2.19.29)$$

This means that if we look at

$$\left( \psi_R^\dagger \mathbf{1} \psi_R, \psi_R^\dagger \vec{\sigma} \psi_R \right) \quad (2.19.30)$$

then it transforms in exactly the same way as a 4-vector! We write this by defining

$$\sigma^\mu = (\mathbf{1}, \vec{\sigma}), \quad \bar{\sigma}^\mu = (\mathbf{1}, -\vec{\sigma}) \quad (2.19.31)$$

so that

$$\psi_R^\dagger \sigma^\mu \psi_R, \quad \psi_L^\dagger \bar{\sigma}^\mu \psi_L \quad (2.19.32)$$

transform as 4-vectors. This means that we can write a Lorentz invariant Lagrangian

$$\mathcal{L} = i\psi_R^\dagger \sigma^\mu \partial_\mu \psi_R + i\psi_R^\dagger \bar{\sigma}^\mu \partial_\mu \psi_R - m \left( \psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L \right) \quad (2.19.33)$$

We added an  $i$  because

$$\left( i\psi_R^\dagger \sigma^\mu \partial_\mu \psi_R \right)^\dagger = -i(\partial_\mu \psi_R^\dagger) \sigma^\mu \psi_R = i\psi_R^\dagger \sigma^\mu \partial_\mu \psi_R \quad (2.19.34)$$

where in the last step we had to integrate by parts to get the derivative to act on the  $\psi_R$  again.

### 2.19.3 Weyl Fermions and ‘Majorana Masses’

We saw above that there are two kinds of spinors,  $\psi_L$  and  $\psi_R$ , and we were able to write down a Lorentz invariant and otherwise reasonable-seeming Lagrangian for them, although to give them a mass we had to have both  $\psi_L$  and  $\psi_R$ .

Actually, one can get a mass for a spinor without needing to have separate fields  $\psi_L$  and  $\psi_R$ . And in fact we don’t even need both kinds of spinors, as we can get something that transforms like  $\psi_L$  directly from the complex conjugate of  $\psi_R$ . By convention, therefore, most advanced treatments of QFT only explicitly use  $\psi_L$  spinors.

The term

$$\psi_R^T \sigma_2 \psi_R = -i \psi_R^\alpha \epsilon_{\alpha\beta} \psi_R^\beta \quad (2.19.35)$$

turns out to be Lorentz invariant. But the most natural way to understand this is to realize that

$$\psi_R^T \sigma_2 \quad (2.19.36)$$

actually transforms as

$$\begin{aligned} \delta(\psi_R^T \sigma_2) &= \frac{1}{2}(i\theta_j + \beta_j) \psi_R^T \sigma_j^T \sigma_2 \\ &= \frac{1}{2}(-i\theta_j - \beta_j) \psi_R^T \sigma_2 \sigma_j \end{aligned} \quad (2.19.37)$$

because

$$\sigma_j^T \sigma_2 = -\sigma_2 \sigma_j \quad (2.19.38)$$

Thus we see that  $\psi_R^T \sigma_2$  transforms in the same way as  $\psi_L^\dagger$ , or in other words

$$(\psi_R^T \sigma_2)^\dagger = -\sigma_2 \psi_R^* \quad (2.19.39)$$

transforms as a left-handed spinor.

Anyway, the most important point is that we can just use  $\psi_L$  or  $\psi_R$  spinors by themselves if we like, but it’s also worth noting that

$$m \psi_R^\alpha \epsilon_{\alpha\beta} \psi_R^\beta \quad (2.19.40)$$

is a perfectly good mass term. This isn’t allowed, though, if the spinor  $\psi_R$  has charge, because this type of mass makes  $\psi_R$  its own anti-particle. So this kind of mass isn’t what we use in QED, when we study particles like the electron. But this is a perfectly good candidate for neutrino masses.

### 2.19.4 Dirac Spinors, Dirac Matrices, and QED

Although the 2-component or Weyl fermions are used to discuss the standard model, supersymmetry, and many other contemporary research topics, QED is usually discussed (and QCD can be discussed) using what are known as 4-component or Dirac spinors.

The idea is to combine  $\psi_R$  and  $\psi_L$  so that the Lagrangian

$$\mathcal{L} = i\psi_R^\dagger \sigma^\mu \partial_\mu \psi_R + i\psi_L^\dagger \bar{\sigma}^\mu \partial_\mu \psi_L - m \left( \psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L \right) \quad (2.19.41)$$

can be written more simply, because it is invariant under  $L \leftrightarrow R$ . In particular, we can write a Dirac spinor

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad (2.19.42)$$

and we can also define

$$\bar{\psi} = \left( \psi_R^\dagger \quad \psi_L^\dagger \right) \quad (2.19.43)$$

Then we define the  $4 \times 4$  Dirac matrices

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \text{i.e.} \quad \gamma^0 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad (2.19.44)$$

In these terms the lagrangian can be written

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi = \bar{\psi} (i\not{\partial} - m) \psi \quad (2.19.45)$$

which is the conventional ‘Dirac Lagrangian’ with EoM

$$(i\gamma^\mu \partial_\mu - m) \psi \quad (2.19.46)$$

which is the Dirac equation.

The Dirac matrices satisfy anti-commutation relations

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \quad (2.19.47)$$

These anti-commutation relations define a *Clifford Algebra*, and they are more fundamental than the explicit form of the Dirac matrices. The particular form of the matrices we are using is called the *Weyl representation*.

It’s useful to define the shorthand

$$\sigma^{\mu\nu} \equiv \frac{i}{2} [\gamma^\mu, \gamma^\nu] \quad (2.19.48)$$

When acting on Dirac spinors, the Lorentz generators can be written as

$$S^{\mu\nu} = \frac{1}{2} \sigma^{\mu\nu} \quad (2.19.49)$$

This action of Lorentz transformations will be correct for any representation of the  $\gamma^\mu$  matrices that satisfies the Clifford algebra. One can derive that the commutation relation

$$[V^{\mu\nu}, V^{\rho\sigma}] = i (g^{\nu\rho} V^{\mu\sigma} + g^{\mu\sigma} V^{\nu\rho} - g^{\nu\sigma} V^{\mu\rho} - g^{\mu\rho} V^{\nu\sigma}) \quad (2.19.50)$$

is satisfied by  $V^{\mu\nu} \rightarrow S^{\mu\nu}$ . Note, however that  $S \neq V$ , and in particular  $S$  are complex. The  $V$  or vector representation is irreducible, while the Dirac representation is reducible, since it is  $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ , the sum of the two kinds of spinor representations.

As the conjugation procedure we are using

$$\bar{\psi} = \psi^\dagger \gamma^0 \quad (2.19.51)$$

to construct the Dirac Lagrangian, and all invariants such as

$$\bar{\psi} \gamma_\mu \psi, \bar{\psi} \gamma_\mu \gamma_\nu \psi, \bar{\psi} \partial_\mu \psi \quad (2.19.52)$$

which transform as tensors under the Lorentz group. The reason we need this is because the representation of the Lorentz group that just acts on these spinors (as opposed to acting on particles in spacetime) is not Unitary.

### 2.19.5 QED Lagrangian

We would like to write a Lagrangian coupling  $\psi$  to the photon. As with scalar QED, we need to preserve the gauge redundancy. But that just means we take

$$\psi(x) \rightarrow e^{-i\alpha(x)} \psi(x) \quad (2.19.53)$$

where  $A_\mu \rightarrow A_\mu + \frac{i}{e} \partial_\mu \alpha$ . Note that the gauge transformation is just a Lorentz scalar (just a spacetime dependent complex number, in fact), so it doesn't care about  $\gamma$  matrices, including the  $\gamma^0$  in the definition of  $\bar{\psi}$ .

Thus we have the same sort of covariant derivative

$$D_\mu \psi = (\partial_\mu + ieA_\mu) \psi \quad (2.19.54)$$

The QED Lagrangian is just

$$\mathcal{L}_{QED} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}^2 \quad (2.19.55)$$

We get something interesting if we try to compare the EoM for QED to that for a scalar field. The EoM for QED is just

$$(i\not{D} - e\not{A} - m)\psi = 0 \quad (2.19.56)$$

If we multiply by this operator again with a sign flip for  $m$  we still have a valid equation, and it is

$$\begin{aligned} 0 &= (i\not{D} - e\not{A} + m)(i\not{D} - e\not{A} - m)\psi \\ &= [(i\partial_\mu - eA_\mu)(i\partial_\nu - eA_\nu)\gamma^\mu\gamma^\nu - m^2] \psi \\ &= \left[ \frac{1}{4}\{i\partial_\mu - eA_\mu, i\partial_\nu - eA_\nu\}\{\gamma^\mu, \gamma^\nu\} + \frac{1}{4}[i\partial_\mu - eA_\mu, i\partial_\nu - eA_\nu][\gamma^\mu, \gamma^\nu] - m^2 \right] \psi \end{aligned} \quad (2.19.57)$$

Note that

$$[i\partial_\mu - eA_\mu, i\partial_\nu - eA_\nu] = -ieF_{\mu\nu} \quad (2.19.58)$$

and so we can write this as

$$\left[ (i\partial_\mu - eA_\mu)^2 - \frac{e}{2}F_{\mu\nu}\sigma^{\mu\nu} - m^2 \right] \psi = 0 \quad (2.19.59)$$

We have an extra term compared to what we find for a scalar field. We can write this finding as

$$\not{D}^2 = D_\mu^2 + \frac{e}{2}F_{\mu\nu}\sigma^{\mu\nu} \quad (2.19.60)$$

What does this mean? Recall that the Lorentz generators act as  $\frac{1}{2}\sigma^{\mu\nu}$  on Dirac spinors, and they have the form (in the Weyl representation)

$$S_{ij} = \frac{1}{2}\epsilon_{ijk} \begin{pmatrix} \sigma^k & 0 \\ 0 & \sigma^k \end{pmatrix}, \quad S_{0i} = -\frac{i}{2} \begin{pmatrix} \sigma^i & 0 \\ 0 & -\sigma^i \end{pmatrix} \quad (2.19.61)$$

These relations are relevant because

$$F_{0i} = E_i, \quad F_{ij} = -\epsilon_{ijk}B_k \quad (2.19.62)$$

This means that we get

$$\left[ D_\mu^2 + m^2 - e \begin{pmatrix} (\vec{B} + i\vec{E}) \cdot \vec{\sigma} & 0 \\ 0 & (\vec{B} - i\vec{E}) \cdot \vec{\sigma} \end{pmatrix} \right] \psi = 0 \quad (2.19.63)$$

This corresponds to a magnetic dipole moment. We have obtained the prediction that

$$\mu_B = \frac{e}{2m_e} \quad (2.19.64)$$

for an electron! We will calculate quantum corrections to this later.

## 2.20 Quantization and Feynman Rules for QED

Just as we found polarization vectors in QED, we need to find spinor solutions to the Dirac equation in order to quantize the field  $\psi(x)$  for the electron (or whatever we call the charged fermion).

### 2.20.1 Solutions to the Dirac Equation

We know that in the free limit, like the other fields we have studied,  $\psi$  obeys the Klein gordon equation, so we can write it as

$$\psi_s(x) = \int \frac{d^3p}{(2\pi)^3} u_s(p) e^{ip \cdot x} \quad (2.20.1)$$



where we take  $p^0 > 0$ . We could just obtain anti-particle solutions by taking  $p_0 < 0$ . But instead, we will write the anti-particle solutions using a different spinor

$$\chi_s(x) = \int \frac{d^3p}{(2\pi)^3} v_s(p) e^{ip \cdot x} \quad (2.20.2)$$

with  $p_0 > 0$ , so that the spinors for particles and anti-particles are  $u_s(p)$  and  $\bar{v}_s(p)$ . As with the vector case, we need only find the solutions for fixed  $p^\mu$ , using the little group that keeps  $p^\mu$  invariant. We also know that

$$p^2 = m^2 \quad (2.20.3)$$

because the free spinors satisfy the Klein-Gordon equation (after squaring the Dirac equation).

The Dirac equation in the Weyl basis is

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (2.20.4)$$

which becomes

$$\begin{pmatrix} -m & p \cdot \sigma \\ p \cdot \bar{\sigma} & -m \end{pmatrix} u_s(p) = \begin{pmatrix} -m & -p \cdot \sigma \\ -p \cdot \bar{\sigma} & -m \end{pmatrix} v_s(p) = 0 \quad (2.20.5)$$

In the rest frame

$$p^\mu = (m, 0, 0, 0) \quad (2.20.6)$$

and so the equations of motion are

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} u_s(p) = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} v_s(p) = 0 \quad (2.20.7)$$

Thus the solutions are

$$u_s = \begin{pmatrix} \xi_s \\ \xi_s \end{pmatrix}, \quad v_s = \begin{pmatrix} \eta_s \\ -\eta_s \end{pmatrix} \quad (2.20.8)$$

for general 2-component (Weyl) spinors  $\xi_s$  and  $\eta_s$ . We can list four linearly independent solutions ('polarizations') if we like. Naively the Dirac spinor has 4 complex components, or 8 real degrees of freedom, but the EoM reduce that to 4 real degrees of freedom, corresponding to spin up or down for particles and antiparticles.

One can show that in a general frame, we can write the solutions as

$$u_s(p) = \begin{pmatrix} \sqrt{p \cdot \sigma} \xi_s \\ \sqrt{p \cdot \bar{\sigma}} \xi_s \end{pmatrix}, \quad v_s(p) = \begin{pmatrix} \sqrt{p \cdot \sigma} \eta_s \\ -\sqrt{p \cdot \bar{\sigma}} \eta_s \end{pmatrix} \quad (2.20.9)$$

If  $p_\mu$  is purely in the  $z$  direction, these square roots of matrices just have components  $\sqrt{E \pm p_z}$ .

To figure out the normalization, we can compute the Lorentz invariant inner product

$$\begin{aligned}
\bar{u}_s(p)u_{s'}(p) &= u_s^\dagger(p)\gamma_0 u_{s'}(p) = \begin{pmatrix} \sqrt{p \cdot \sigma} \xi_s \\ \sqrt{p \cdot \bar{\sigma}} \xi_s \end{pmatrix}^\dagger \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{p \cdot \sigma} \xi_{s'} \\ \sqrt{p \cdot \bar{\sigma}} \xi_{s'} \end{pmatrix} \\
&= \begin{pmatrix} \xi_s \\ \xi_s \end{pmatrix}^\dagger \begin{pmatrix} \sqrt{(p \cdot \sigma)(p \cdot \bar{\sigma})} & \\ 0 & \sqrt{(p \cdot \sigma)(p \cdot \bar{\sigma})} \end{pmatrix} \begin{pmatrix} \xi_{s'} \\ \xi_{s'} \end{pmatrix} \\
&= 2m\delta_{s,s'}
\end{aligned} \tag{2.20.10}$$

Clearly in the rest frame this result holds; you can easily check that it also works with a momentum in the  $z$ -direction, since then the inner products are  $\sqrt{E - p_z}\sqrt{E + p_z}$ . So by rotation invariance this normalization works in any frame; the normalization is in fact Lorentz invariant. Similarly

$$\bar{v}_s(p)v_{s'}(p) = -2m\delta_{s,s'} \tag{2.20.11}$$

and

$$\bar{v}_s u_{s'} = \bar{u}_s v_{s'} = 0 \tag{2.20.12}$$

as one would want for orthonormal solutions.

### 2.20.2 Quantizing the Spinor Field

In analogy with the scalar and vector field, we write

$$\psi(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^s u_p^s e^{-ipx} + b_p^{s\dagger} v_p^s e^{ipx}) \tag{2.20.13}$$

and

$$\bar{\psi}(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^{s\dagger} \bar{u}_p^s e^{ipx} + b_p^s \bar{v}_p^s e^{-ipx}) \tag{2.20.14}$$

In this language  $\psi(x)$  annihilates incoming electrons and  $\bar{\psi}(x)$  annihilates incoming positrons.

Given that the Lagrangian for our theory was

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi \tag{2.20.15}$$

we have that the canonical conjugate to  $\psi$  is

$$\frac{\partial \mathcal{L}}{\partial \dot{\psi}} = i\bar{\psi} \tag{2.20.16}$$

and so one might think that we should impose canonical commutation relations

$$[\psi_a(t, \vec{x}), \psi_b^\dagger(t, \vec{y})] = \delta^3(\vec{x} - \vec{y})\delta_{ab} \tag{2.20.17}$$

but in fact this is not correct, and we need impose the anti-commutation relations

$$\{\psi_a(t, \vec{x}), \psi_b^\dagger(t, \vec{y})\} = \delta^3(\vec{x} - \vec{y})\delta_{ab} \quad (2.20.18)$$

We will study the reason in more detail soon (well, next semester), but for now let us consider the issue of stability, which we also discussed in the case of spin 1 particles.

The energy momentum tensor for our fermion field is

$$T_{\mu\nu} = i\bar{\psi}\gamma_\mu\partial_\nu\psi - g_{\mu\nu}(i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi) \quad (2.20.19)$$

and so the energy density (or Hamiltonian density) is

$$\begin{aligned} T_{00} &= \bar{\psi}(i\gamma^i\partial_i + m)\psi \\ &= i\bar{\psi}\gamma^0\partial_t\psi \end{aligned} \quad (2.20.20)$$

Now we can write that using our quantized version of the Diract field, and see what happens if we choose commutation vs anticommutation relations, ie

$$[a_p^s, a_q^{s'\dagger}] = \delta^3(\vec{p} - \vec{q})\delta_{s,s'} \quad \text{OR} \quad \{a_p^s, a_q^{s'\dagger}\} = \delta^3(\vec{p} - \vec{q})\delta_{s,s'} \quad (2.20.21)$$

This means that the energy is

$$\begin{aligned} E &= \int d^3x T_{00} \\ &= i \int d^3x \int \frac{d^3p d^3q}{(2\pi)^6} \frac{1}{2\sqrt{E_q E_k}} \sum_{s,s'} \left( e^{ipx} a_p^{s'\dagger} \bar{u}_p^{s'} + e^{-ipx} b_p^{s'} \bar{v}_p^{s'} \right) \\ &\quad \times \gamma^0 \partial_t (e^{-iqx} a_q^s u_q^s + e^{iqx} b_q^{s\dagger} \bar{v}_q^s) \end{aligned} \quad (2.20.22)$$

As usual, the integral over  $x$  forces  $\vec{q} = \pm\vec{p}$  in the various terms. Note that since

$$\bar{u}_s(p)\gamma^0 u_{s'}(p) = u_s^\dagger(p) u_{s'}(p) = 2E_p \delta_{ss'} \quad (2.20.23)$$

and

$$u_s^\dagger(p) v_{s'}(-p) = v_s^\dagger(p) u_{s'}(-p) = 0 \quad (2.20.24)$$

Using these relations, we find

$$E = \sum_s \int \frac{d^3p}{(2\pi)^3} E_p (a_p^{s\dagger} a_p^s - b_p^s b_p^{s\dagger}) \quad (2.20.25)$$

Now we see what happens with commutators versus anticommutators. With the latter we get

$$E = E_0 + \sum_s \int \frac{d^3p}{(2\pi)^3} E_p (a_p^{s\dagger} a_p^s + b_p^{s\dagger} b_p^s) \quad (2.20.26)$$

but if we had used commutators instead, we would have had *negative energy from the antiparticles!* To avoid having an energy that is unbounded from below, we must use anticommutation relations.

At an operational level, what happened? Why was this different from the scalar case (or the vector boson case)? Basically, we got a different result because the kinetic term for  $\psi$  involves only one time derivative, and the extra factor of energy was hidden in  $\bar{u}u$ . We'll discuss this in various other ways next semester.

There are several other problems with violating this *spin and statistics* relation; in particular theories with the wrong choice of commutators vs anticommutators violate causality and cannot have a Lorentz invariant S-Matrix. We will revisit these issues next semester.

### 2.20.3 Feynman Rules for QED

To finish up our limited discussion of QED, to be resumed next semester, we would like to determine the Feynman rules.

We have quantized the free Dirac field with

$$\psi(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^s u_p^s e^{-ipx} + b_p^{s\dagger} v_p^s e^{ipx}) \quad (2.20.27)$$

and

$$\bar{\psi}(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^{s\dagger} \bar{u}_p^s e^{ipx} + b_p^s \bar{v}_p^s e^{-ipx}) \quad (2.20.28)$$

where we have creation and annihilation operators that anticommute. The non-vanishing anti-commutators are

$$\{a_p^s, a_q^{s'\dagger}\} = \{b_p^s, b_q^{s'\dagger}\} = \delta_{ss'} (2\pi)^3 \delta^3(\vec{p} - \vec{q}) \quad (2.20.29)$$

and we have a basis for spinors

$$u_s(p) = \begin{pmatrix} \sqrt{p \cdot \sigma} \xi_s \\ \sqrt{p \cdot \bar{\sigma}} \xi_s \end{pmatrix}, \quad v_s(p) = \begin{pmatrix} \sqrt{p \cdot \sigma} \eta_s \\ -\sqrt{p \cdot \bar{\sigma}} \eta_s \end{pmatrix} \quad (2.20.30)$$

with  $\xi_1 = \eta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\xi_2 = \eta_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . The spinors satisfy outer-product sum rules

$$\sum_{s=1}^2 u_s(p) \bar{u}_s(p) = \not{p} + m \quad (2.20.31)$$

$$\sum_{s=1}^2 v_s(p) \bar{v}_s(p) = \not{p} - m \quad (2.20.32)$$

The Feynman propagator for our (free) spinors is

$$\langle \bar{\psi}(p) \psi(-p) \rangle = \frac{i(\not{p} + m)}{p^2 - m^2 + i\epsilon} \quad (2.20.33)$$

The easiest way to understand this is based on the fact that it must satisfy the Dirac equation.

The QED Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi \quad (2.20.34)$$

We already know the propagators, and the external line factors for photons. Just as photons get factors of  $\epsilon_\mu(p)$ , for our Dirac fermions the LSZ formula gives

$$u^s(p), \bar{u}^s(p), \bar{v}^s(p), v^s(p) \quad (2.20.35)$$

We conventionally write arrows pointing into the diagram to represent initial particles and out of the diagram to represent final state particles, and the reverse for anti-particles.

Note that external spinors represent on-shell particles, which means that they obey the EoM (Dirac equation)

$$(\not{p} - m)u^s(p) = \bar{u}^s(\not{p} - m) = 0 \quad (2.20.36)$$

$$(\not{p} + m)v^s(p) = \bar{v}^s(\not{p} + m) = 0 \quad (2.20.37)$$

This can be useful for simplifying calculations.

QED is simpler than scalar QED in that there is only a single interaction

$$\mathcal{L}_{int} = -e\bar{\psi}\gamma^\mu\psi A_\mu \quad (2.20.38)$$

This leads to a single vertex

$$-ie\gamma^\mu \quad (2.20.39)$$

that connects to either the photon polarization or to a photon propagator.

As an example, we can consider  $e^-\gamma \rightarrow e^-\gamma$ . One diagram that contributes gives

$$(-ie)^2 \bar{u}(p_3)\gamma^\mu \frac{i(\not{p}_1 + \not{p}_2 + m)}{(p_1 + p_2)^2 - m^2} \gamma^\nu u(p_1) \epsilon_\mu^2(p_2) \epsilon_\nu^4(p_4) \quad (2.20.40)$$

Note that the gamma matrices from the vertices and propagator get multiplied together in a series along the electron line.

## 2.21 Overview of ‘Renormalization’

We would now like to move on to study genuinely quantum mechanical effects in QFT. In perturbation theory, that means studying ‘loop diagrams’ and their effects on various observables, particularly scattering amplitudes, decay rates, etc.

Computing loop diagrams means that we will be integrating over intermediate particle momenta, for example

$$\int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2} \frac{i}{(p - k)^2 - m^2} \quad (2.21.1)$$

Note that this is equivalent to a position space integral

$$\int d^4x d^4y D_F(x_1, x) [D_F(x, y)]^2 D_F(y, x_2) \quad (2.21.2)$$

Roughly speaking, the region where  $k \rightarrow \infty$  comes from  $y \rightarrow x$  in position space, whereas  $k \rightarrow 0$  is associated with  $x - y \rightarrow \infty$ .

If we are doing some experiment at a length scale  $L_{exp}$ , then it's important to consider how the momentum  $k$  in the loop compares to  $L_{exp}$ . In particular

- When  $kL_{exp} \gg 1$  these are UV (ultraviolet), high-energy, or short distance effects. These can usually be incorporated in small changes in the description of the theory at scales of order  $L_{exp}$ . Sometimes, taking  $k \rightarrow \infty$  gives unphysical infinities that arise because we have naively assumed that the theory does not change at all at arbitrarily short distances, but these always cancel out of physical observables.
- In the case  $kL_{exp} \sim 1$  the effect comes from distance scales of order the experimental scale, and will generically be important and worth keeping track of in physical observables.
- When  $kL_{exp} \ll 1$  we are talking about IR (infrared), low-energy, or long distance effects. Usually these are small, but sometimes these give naive infinities if we make the mistake of assuming that our experiment has arbitrarily precise resolution in energy. The IR is where long-distance forces act, so a careful consideration of IR divergences tells us a tremendous amount about basic features of reality, such as which particles can possibly give rise to long-range potentials.

One should always think carefully about the length/energy scales of relevance in a given calculation. In fact, these comments are all related to ways in which, by computing without thinking, one can obtain non-sense answers:

- One can insist that a particular QFT makes sense down to arbitrarily short distances. Sometimes this is a possibility, for example in  $g\phi^3$  theory in  $d \leq 6$  dimensions or in  $\lambda\phi^4$  theory in  $d < 4$  dimensions. But many other times it is not. We already saw an example where it's a disaster, namely when one has an interaction like

$$\frac{(\partial\phi)^4}{\Lambda^4} \quad (2.21.3)$$

in any number of dimensions. This sort of theory can't possibly make sense (in perturbation theory) at distances shorter than  $1/\Lambda$ .

- We can (accidentally) setup experiments that cannot be performed. How many photons of wavelength  $10^{20}$  meters do we produce when we wave our hands around? If you cannot measure it, then you probably shouldn't calculate a scattering cross section that *presumes* that the answer is definite... for example, you might have naively presumed that the answer is zero, but QFT knows better.

So one must be careful in translating from formal calculations to physics – if the latter doesn’t make sense, the former won’t either.

Why the name ‘renormalization’? Consider the example of some scalar theory with

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m_0^2\phi^2 - \frac{\lambda_0}{4!}\phi^4 \quad (2.21.4)$$

When we studied this theory, we noted that when  $\lambda = 0$ , the states look like a Fock space of particles with mass

$$m_{phys} = m_0 \quad (2.21.5)$$

where  $m_0$  is just a parameter in the Lagrangian. But there is no reason why this must be maintained when we include quantum effects – such effects can change the physical mass! They can also change the normalization of the 1-particle states; that is, when we compute quantum effects we might find

$$\langle\psi_p|\phi(x)|\Omega\rangle = Ze^{ip\cdot x} \quad (2.21.6)$$

where  $Z \neq 1$  is some constant. We need to take account of this by ‘renormalizing’ the field  $\phi$ .

Similarly, one might wonder how to think about the coupling constant  $\lambda_0$ , but a natural interpretation is via

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 E_{CM}^2} \lambda_0^2 \quad (2.21.7)$$

the differential cross section for  $\phi$  particle scattering at tree level. But this will receive quantum mechanical corrections with a detailed dependence on the kinematics and energy scale  $E_{CM}$ , in which case it becomes unclear what  $\lambda_0$  really means. Again, the parameter in the Lagrangian will not connect so directly to the physics, and it’s worth asking how we fix it using experiments.

Thus we need to *Renormalize* the various parameters in the Lagrangian *to connect them to physical observables*. Hence the name.

Finally, we would like to take the idea of effective field theory very seriously, and consider the sensitivity of our theory to the short-distance, UV parameters defined at the very high energy scale

$$\Lambda \quad (2.21.8)$$

which is the energy cutoff. In particular, we would like to ask how the long-distance parameters and observables change with  $\Lambda$  and the short-distance parameters that define the theory. This leads to the ‘renormalization group’, or more aptly, the ‘renormalization flow’, which tells us how the parameters change with  $\Lambda$ . This flow is intimately connected with the phenomenology of quantum mechanical effects, which can cause forces to change (logarithmically) with distance.

## 2.22 Casimir Effect

The Casimir effect is a fun example of something other than scattering that we can compute in QFT. It also has some of the features that we’ll find in other QFT situations – naive infinities from

unphysical short-distance extrapolations, and the necessity to ask physical questions in order to get physical answers.

We already saw that the naive vacuum energy is

$$E = \langle 0|H_0|0\rangle = \int \frac{d^3k}{(2\pi)^3} \frac{E_k}{2} = \frac{1}{4\pi^2} \int^\Lambda k^3 dk \sim \frac{\Lambda^4}{4\pi^2} \quad (2.22.1)$$

if we employ a ‘hard cutoff’ on the momenta. If we take the hard cutoff to infinity we get an infinite answer. Recall that in our balls and springs model there was something more complicated than a hard cutoff, because we had  $E_k \sim \Lambda \sin(k/\Lambda)$ , but roughly speaking we would have obtained the same scaling behavior.

Let’s set up our 1-d Casimir box, beginning at  $x = 0$  and extending to  $x = a$ . We will employ an IR cutoff at  $L \gg a$  so that we can sum the energy of modes between  $x = 0$  and  $x = a$  and those between  $x = a$  and  $x = L$ .

Modes in a box of size  $r$  are discretized with

$$E_n = \frac{\pi}{r} n \quad (2.22.2)$$

which turns the integral (which gave the zero point energy) into a sum. But we still get an energy, even in 1d, which is

$$E_{tot}(a) = E_{box} + E_{outside} = \left( \frac{1}{a} + \frac{1}{L-a} \right) \frac{\pi}{2} \sum_{n=1}^{\infty} n \quad (2.22.3)$$

and this is still infinite. We can differentiate it to obtain a force

$$F(a) = -\frac{dE_{tot}}{da} = \left( \frac{1}{a^2} - \frac{1}{(L-a)^2} \right) \frac{\pi}{2} \sum_{n=1}^{\infty} n \quad (2.22.4)$$

but it’s still infinite. The problem is that the walls are way too powerful – they can stop modes with arbitrarily large momentum and energy! So we need to account for the fact that the walls won’t stop modes with such high momentum and energy (for example, you might be aware that gamma rays can penetrate walls that stop visible light).

Going back to our hard cutoff, we note that

$$E_{max} = \pi\Lambda = \frac{\pi n}{r} \quad (2.22.5)$$

implies that we must have

$$n_{max} = \frac{r\Lambda}{\pi} \quad (2.22.6)$$

Thus we find that

$$E(r) = \frac{\pi}{2r} \sum_{n=1}^{n_{max}} n = \frac{\pi}{2r} \frac{n_m(n_m + 1)}{2} = \frac{\pi}{4r} (\Lambda r)(\Lambda r + 1) = \frac{\pi}{4} (\Lambda^2 r + \Lambda) \quad (2.22.7)$$



This means that

$$E_{tot}(a) = \frac{\pi}{4}(\Lambda^2 L + 2\Lambda) \quad (2.22.8)$$

This has no dependence on  $a$ , so it seems that the Casimir energy and force are zero! This is the right answer to leading order in large  $\Lambda$ , but our coarse approximation has missed a finite piece.

The problem is that  $n_{max}$  is a discontinuous function of  $a$  and  $L$ , since modes are either included or not. Instead we can cutoff our sum with a weighting

$$n \rightarrow ne^{-\frac{n}{n_{max}}} \quad (2.22.9)$$

which shuts off continuously for  $n > n_{max} = \Lambda r$ . If we compute  $E(r)$  now, we find

$$\begin{aligned} E(r) &= \frac{\pi}{2} \sum_n n e^{-\frac{n}{r\Lambda}} \\ &\approx \frac{\pi}{2} \Lambda^2 r - \frac{\pi}{24r} + \dots \end{aligned} \quad (2.22.10)$$

This gives

$$E_{tot}(a) = \frac{\pi}{2} \Lambda^2 L - \frac{\pi}{24a} - \frac{\pi}{24(L-a)} \quad (2.22.11)$$

But now we want to take the limit that  $L \gg a$ , so the last term drops out, and we find a force

$$F(a) = -\frac{dE_{tot}(a)}{da} = -\frac{\pi}{24a^2} \quad (2.22.12)$$

This is an attractive Casimir force.

### Regulator (In)Dependence

But our exponential regulator was totally ad hoc! How do we know that the result wasn't due entirely to the choice of  $e^{-n/n_{max}}$ ? Well let's just try a general regulator, by defining

$$E(r) = \frac{\pi}{2} \sum_n n f\left(\frac{n}{r\Lambda}\right) \quad (2.22.13)$$

where  $f(x)$  is just some regulator function. In this case we find

$$E(L-a) = \frac{\pi}{2} (L-a) \Lambda^2 \sum_n \frac{n}{(L-a)^2 \Lambda^2} f\left(\frac{n}{(L-a)\Lambda}\right) \quad (2.22.14)$$

In the limit that  $L \rightarrow \infty$  we can really just take the continuum (integral) limit of the sum, which gives

$$E(L-a) = \frac{\pi}{2} L \Lambda^2 \int_0^\infty dx x f(x) - \frac{\pi}{2} a \Lambda^2 \int_0^\infty dx x f(x) \quad (2.22.15)$$

where  $x = \frac{n}{(L-a)\Lambda}$ .

For fixed  $a$  we must be more careful, so if we define the energy density as

$$\rho = \frac{\pi}{2} \Lambda^2 \int_0^\infty dx x f(x) \quad (2.22.16)$$

then the first term above is just  $\rho L$ . The second term can be re-written as

$$\frac{\pi}{2} a \Lambda^2 \int_0^\infty dx x f(x) = \frac{\pi}{2a} \int dn n f\left(\frac{n}{a\Lambda}\right) \quad (2.22.17)$$

to make it look very similar to the  $E(a)$  term. Thus we can combine  $E(a)$  and  $E(L-a)$  to write

$$E_{tot} = \rho L + \frac{\pi}{2a} \left[ \sum_n n f\left(\frac{n}{a\Lambda}\right) - \int dn n f\left(\frac{n}{a\Lambda}\right) \right] \quad (2.22.18)$$

The term in brackets is the difference between an infinite sum and an infinite integral. In complete generality, this is an Euler-Maclaurin series

$$\sum_{n=1}^N F(n) - \int_0^N F(n) dn = \frac{F(0) + F(N)}{2} + \frac{F'(N) - F'(0)}{12} + \dots + B_j \frac{F^{(j-1)}(N) - F^{(j-1)}(0)}{j!} + \dots$$

where  $B_j$  are the Bernoulli nubers, which vanish for odd  $j > 1$ , and note we have used  $B_2 = 1/6$ .

In our computation we have

$$F(n) = n f\left(\frac{n}{a\Lambda}\right) \quad (2.22.19)$$

so assuming that

$$\lim_{x \rightarrow \infty} x f^{(j-1)}(x) = 0 \quad (2.22.20)$$

we have

$$E_{tot} = \rho L - \frac{\pi f(0)}{24a} - \frac{B_4}{4!} \frac{3\pi}{2a^3 \Lambda^2} f''(0) + \dots \quad (2.22.21)$$

Note that as  $\Lambda \rightarrow \infty$  only the first two terms survive, and as long as

$$f(0) = 1 \quad (2.22.22)$$

we always get the same, consistent result for the Casimir force. The two requirements on  $f(x)$  insure that

- The ultra high energy, short distance modes are irrelevant for the physics.
- The short distance regulator function  $f(x)$  does not change the modes at very long distances, where the Casimir effect actually arises.

## Sub-Leading Terms and Counterterms

The sub-leading terms at large  $\Lambda$  aren't exactly zero, and they do depend on  $f(x)$ . If you want to measure the Casimir effect very, very accurately, you really do need to know what happens at short distances, and the regulator function  $f(x)$  would be observable – you would need to compute it by modeling the interaction of the modes with the walls of the box. But it's also worth noting that to obtain the first few terms in a power series in

$$\frac{1}{\Lambda a} \ll 1 \quad (2.22.23)$$

we only need to know *the first few derivatives* of  $f(x)$ , which is a finite amount of information about the short-distance corrections to the long distance physics. That's all that survives when we do experiments with finite accuracy.

You might also wonder why we needed to include the IR regulator  $L$ . Physically it's clear that we need to account for forces on both sides of the wall, but there's a more formal point to be made. If instead of the free field Lagrangian we had used

$$\mathcal{L} \rightarrow \mathcal{L} + \rho_c \quad (2.22.24)$$

then the energy density would have been

$$\rho_0 \sim \Lambda^2 \rightarrow \rho_0 - \rho_c \quad (2.22.25)$$

In other words, by adding a constant (a cosmological constant, it turns out) to the Lagrangian, we can shift the zero point energy density.

This is our first example where 'bare' terms in the Lagrangian combine with calculable effects to produce the 'physical' or 'renormalized' quantity. In more old-fashioned terms, we would say that  $\rho_c$  is a term that can be added to cancel infinities in intermediate steps of a calculation, although in any case, the physical effects we calculate will always be finite.

## 2.23 The Exact 2-pt Correlator – Kallen-Lehmann Representation

Lately, we have mostly been performing computations in perturbation theory. But there are some exact results that we can obtain in QFT. One example, which will be useful to keep in mind when we study renormalization, is the *exact* or non-perturbative time ordered correlator of two fields, namely

$$D(x - y) = \langle \Omega | T \{ \phi(x) \phi(y) \} | \Omega \rangle \quad (2.23.1)$$

where these are genuine Heisenberg picture fields. In the free case this is the Feynman propagator, but in general it could be much more complicated.

In this section we'll compute exactly what form it must take, obtaining the Kallen-Lehmann representation for it. We'll see that in a completely generally theory, the propagator gets 'smeared out in mass', corresponding to the fact that it can create and destroy a continuum of possible states.

The idea of the argument is simple – we just want to insert a complete set of states between  $\phi(x)$  and  $\phi(y)$ , and then use general principle to constrain the result.

We may as well choose intermediate states that have definite momentum  $k$ . We can do this (as always) because the momentum operator  $P_\mu$  always commutes with the Hamiltonian  $H = P_0$ . Thus for any state with momentum  $k$  we can write

$$\langle \Omega | \phi(x) | \psi_k \rangle = e^{ik \cdot x} \langle \Omega | \phi(0) | \psi_k \rangle \quad (2.23.2)$$

This follows because the momentum  $P_\mu$  is the generator of translations, so we can move  $x$  to 0 at the cost of a phase factor. Similarly

$$\langle \psi_k | \phi(y) | \Omega \rangle = e^{-ik \cdot y} \langle \psi_k | \phi(0) | \Omega \rangle \quad (2.23.3)$$

where I have assumed that  $\phi = \phi^\dagger$ . Let's assume WLOG that  $x_0 > y_0$ , as usual, so we don't have to worry about time ordering. Then

$$\begin{aligned} D(x-y) &= \langle \Omega | T \{ \phi(x) \phi(y) \} | \Omega \rangle \\ &= \sum_{\psi_k} e^{ik \cdot (x-y)} |\langle \Omega | \phi(0) | \psi_k \rangle|^2 \end{aligned} \quad (2.23.4)$$

where the sum is over all states in the QFT with momentum  $k$ . This includes one particle states, million particle states, and everything else in the Hilbert space.

We've already used translation symmetry. Now let's use Lorentz invariance. The object

$$|\langle \Omega | \phi(0) | \psi_k \rangle|^2 \quad (2.23.5)$$

is a Lorentz scalar that only depends on a single momentum  $k_\mu$ , so it can only depend on  $k^2$  and, if  $k_\mu$  is time-like, the sign of  $k_0$ . In fact  $k$  must be time-like, since it must have positive energy if  $|\Omega\rangle$  is the zero-energy vacuum, so we must have

$$\sum_{\psi_k} \delta^4(k-p) |\langle \Omega | \phi(0) | \psi_k \rangle|^2 = \frac{1}{(2\pi)^3} \theta(p_0) \rho(p^2) \quad (2.23.6)$$

for the *spectral function*  $\rho(x)$ , where  $\rho(x) = 0$  for  $x < 0$ . More importantly, we have that

$$\rho(\mu^2) \geq 0 \quad (2.23.7)$$

so the spectral function is non-negative. Thus we can write

$$\begin{aligned} D(x-y) &= \int \frac{d^4 p}{(2\pi)^3} \rho(p^2) \theta(p_0) e^{ip \cdot (x-y)} \\ &= \int \frac{d^4 p}{(2\pi)^3} \theta(p_0) e^{ip \cdot (x-y)} \int_0^\infty d\mu^2 \rho(\mu^2) \delta(p^2 - \mu^2) \end{aligned} \quad (2.23.8)$$

Now we can interchange the order of integration and incorporate the time ordering symbol to write this as

$$D(x-y) = \int_0^\infty d\mu^2 \rho(\mu^2) D_F(x-y, \mu^2) \quad (2.23.9)$$

or in momentum space

$$D(p) = \int_0^\infty d\mu^2 \rho(\mu^2) \frac{i}{p^2 - \mu^2 + i\epsilon} \quad (2.23.10)$$

Thus we see that the exact propagator is just an integral over the Feynman (free theory) propagator with a positive spectral function depending on the mass  $\mu^2$ . Note that this provides a strong constraint on its high energy behavior – it can never vanish faster than  $1/p^2$  as  $p \rightarrow \infty$ .

We can use one more ingredient to derive a sum rule for  $\rho(\mu^2)$ . Note that even in the exact theory we have the commutation relation

$$[\dot{\phi}(t, \vec{x}), \phi(t, \vec{y})] = -i\delta^3(\vec{x} - \vec{y}) \quad (2.23.11)$$

Furthermore, at equal times  $x_0 = y_0$  we have

$$\langle 0 | [\dot{\phi}_0(x), \phi_0(y)] | 0 \rangle = -i\delta^3(\vec{x} - \vec{y}) \quad (2.23.12)$$

This leads us to conclude that in any theory, no matter how strongly interacting, we must have

$$\int_0^\infty \rho(\mu^2) d\mu^2 = 1 \quad (2.23.13)$$

since the exact commutator is just an integral of  $\rho(\mu^2)$  times the free theory commutator. However, as we will see, this requires that the fields are correctly normalized. Note that the free theory is just the case where

$$\rho(\mu^2) = \delta(\mu^2 - m^2) \quad (2.23.14)$$

So we see that the integrated spectral weight is always the same, but interactions can spread it out over a range of  $\mu^2$  values. This is a strong constraint, since  $\rho(\mu^2)$  must always be positive.

Finally, let's stop and think about what we should expect to find in a perturbative QFT. In the free theory, we have a particle of mass  $m$ , and then starting at energy

$$E \geq 2m \quad (2.23.15)$$

we have a continuum of 2 particle states, then at  $E \geq 3m$  we have a continuum of 3 particle states, and so on. Once we include perturbative loop corrections, there will (generically) be mixing between all of these states. But in perturbation theory, this is small, so in particular the 2-particle states will only be present with

$$E^2 - \vec{p}^2 > 4m^2 + \mathcal{O}(g) \quad (2.23.16)$$

where  $g$  is the coupling. This means that we should expect

$$\rho(\mu^2) \approx Z\delta(m^2 - \mu^2) + \rho_{\gtrsim 4m^2}(\mu^2) \quad (2.23.17)$$

where the latter term accounts of  $2, 3, \dots$  particle states, and the former is an isolated delta function with coefficient  $Z < 1$ . In other words,  $D(p)$  still has an isolated pole at the *physical mass*  $m^2$ , with residue  $Z \leq 1$ . It's worth emphasizing that this statement about the magnitude of  $Z$  only follows if we insist that  $\phi$  is a canonical Heisenberg picture operator; in general we can use any operator  $\mathcal{O}$  to create our particles, in which case  $Z$  might take any positive value. In fact, if we like we can define the normalization of our operators so that  $Z = 1$ . We will be computing these effective  $Z$  and  $m^2$  values in terms of 'bare' Lagrangian parameters when we compute loop effects in QFT.

## 2.24 Basic Loop Effects – the 2-pt Correlator

Now let's look at the one-loop correction to the propagator in our favorite theory, which has an action

$$S = \int d^d x \left[ \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + \frac{g}{6}\phi^3 \right] \quad (2.24.1)$$

in  $d$  spacetime dimensions. For various reasons (such as the fact that the dimensions of  $g$  depend on  $d$ ) it's useful to view  $d$  as a parameter.

We already saw that we can exclude bubble diagrams from consideration. We will also exclude disconnected diagrams, as they cannot tell us anything about propagation. There are disconnected diagrams that contribute to  $\langle \Omega | \phi(x) | \Omega \rangle$  but these can always be cancelled. So let's look at connected diagrams contributing to

$$\langle \Omega | T \{ \phi(x) \phi(y) \} | \Omega \rangle \quad (2.24.2)$$

In fact, it turns out that we need only compute the *1-Particle Irreducible* or '1PI' diagrams. These are defined as diagrams that are still connected after any one internal line is cut.

The reason we only need 1PI diagrams is because

$$\begin{aligned} D(k) &= D_F(k) + D_F(k)(i\Pi(k))D_F(k) + D_F(k)(i\Pi(k))D_F(k)(i\Pi(k))D_F(k) + \dots \\ &= \frac{i}{k^2 - m^2 + i\epsilon} \sum_{n=0}^{\infty} \left[ \frac{-\Pi(k)}{k^2 - m^2 + i\epsilon} \right]^n \\ &= \frac{i}{k^2 - m^2 + i\epsilon + \Pi(k)} \end{aligned} \quad (2.24.3)$$

So all non-1PI contributions are just part of the geometric series, and do not need to be computed independently.

So now we need to compute  $i\Pi(k)$ . It is

$$i\Pi(p) = \frac{1}{2}(ig)^2 \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2 + i\epsilon} \frac{i}{(p-k)^2 - m^2 + i\epsilon} \quad (2.24.4)$$

Note that there is a symmetry factor of  $1/2$ , as we found a while ago when we discussed these factors. If this were part of a scattering amplitude we would have  $p^2 = m^2$ , but we will allow  $p$  to be ‘off-shell’, meaning that  $p^2 \neq m^2$ , since this diagram could in general occur as some sub-diagram in a more complicated process. For example, it could appear as a propagator correction in a 2-to-2 scattering amplitude.

The next part is ‘just’ a computation, but there are many points to be made (1) the spacetime dimension dependence, (2) about regularization of the short-distance divergences (infinities), and (3) the general techniques for computing such integrals.

Let’s think for a bit about dimensional analysis. The object  $\Pi(p)$  has dimension  $[Energy]^2$ , in accord with the dimension  $p^2$  and  $m^2$  in the propagator. Part of this comes from the dimensions of  $g$ ; the integral itself has dimension

$$\frac{[Energy]^d}{[Energy]^4} \quad (2.24.5)$$

Thus there are no short-distance (large  $k$ ) divergences at all when  $d < 4$ , but such divergences may appear for  $d \geq 4$ , suggesting that the theory needs to be more carefully defined at short distances. Furthermore, note that when  $d = 6$  we see again the special case where  $g$  is dimensionless, so all units apparently come from the integral.

There are a variety of roughly equivalent techniques for evaluating these kinds of integrals. The most standard is the ‘Feynman parameterization’, where we use the fact that

$$\frac{1}{A_1 A_2 \cdots A_n} = \int_0^1 dx_1 dx_2 \cdots dx_n \delta(x_1 + \cdots x_n - 1) \frac{(n-1)!}{[x_1 A_1 + x_2 A_2 + \cdots x_n A_n]^n} \quad (2.24.6)$$

The version of this that we need is simply

$$\frac{1}{AB} = \int_0^1 dx \frac{1}{(A + (B - A)x)^2} \quad (2.24.7)$$

which you can check instead of paying attention to the next three minutes of lecture.

Now we can apply this with  $A = (p - k)^2 - m^2 + i\epsilon$  and  $B = k^2 - m^2 + i\epsilon$  to write

$$\begin{aligned} A + [B - A]x &= (p - k)^2 - m^2 + i\epsilon + [k^2 - (k - p)^2]x \\ &= [k - (1 - x)p]^2 + p^2 x(1 - x) - m^2 + i\epsilon \end{aligned} \quad (2.24.8)$$

where we completed the square in the second line. This leads to an integral

$$i\Pi(p) = \frac{g^2}{2} \int \frac{d^d k}{(2\pi)^d} \int_0^1 dx \frac{1}{[(k - (1 - x)p)^2 + p^2 x(1 - x) - m^2 + i\epsilon]^2} \quad (2.24.9)$$

We can shift the integration variable to

$$\ell = k + p(1 - x) \quad (2.24.10)$$

so that we have

$$i\Pi(p) = \frac{g^2}{2} \int_0^1 dx \int \frac{d^d \ell}{(2\pi)^d} \frac{1}{[\ell^2 - \Delta + i\epsilon]^2} \quad (2.24.11)$$

with

$$\Delta \equiv m^2 - p^2 x(1-x) \quad (2.24.12)$$

Recall that  $\ell^2 = \ell_0^2 - \vec{\ell}^2$ ; let's consider the  $\ell_0$  integral as a complex contour integral. Assuming  $\Delta > 0$ , there are poles at

$$\ell_0 = \pm \left( \sqrt{\vec{\ell}^2 + \Delta} - i\epsilon \right) \quad (2.24.13)$$

The contour in  $\ell_0$  ran from  $-\infty$  to  $\infty$ , but the placement of the poles means that we can rotate the contour of integration counter-clockwise, so that it runs from  $-i\infty$  to  $i\infty$ . This is often called ‘Wick rotation’ in the context of Feynman integrals, and you can check that it still works when  $\Delta < 0$ .

This means that (in effect) we now have  $\ell_0 = i\ell_E$  – in other words, we are now in Euclidean, as opposed to Lorentzian, space. The integral becomes

$$i\Pi(p) = i \frac{g^2}{2} \int_0^1 dx \int \frac{d^d \ell_E}{(2\pi)^d} \frac{1}{[\ell_E^2 + \Delta]^2} \quad (2.24.14)$$

and we have no more need for  $\epsilon$ , as it's now clear that the denominator never vanishes for  $\Delta > 0$ .

Now we have many choices of what to do, but the basic issue we face is that for  $d \geq 4$ , this integral diverges if we trust it to arbitrarily short distances, aka arbitrarily large  $\ell_E$ . Here are some options

- There's something called the Pauli-Villars regulator you can read about in the book. The idea is to subtract the opposite of a particle at high energy.
- We can evaluate this as a function of  $d$ , imagining that  $d < 4$ , and then *analytically continue* in the dimension  $d$ . That may sound crazy, but it's by far the most popular way of regulating integrals in QFT, because it preserves Lorentz invariance and something called ‘gauge invariance’ that we'll talk about when we study gauge theories.
- We can differentiate the integral once or twice with respect to  $\Delta$ , then do the integral with  $d < 8$ , and then note that all we're missing are some divergent integration constants.
- We can just impose a hard cutoff  $|\ell_E| < \Lambda$  and evaluate, corresponding to a physical situation where there are no more modes above the energy/momentum cutoff.

Let's start with the last and work our way up.



Evaluating with a hard cutoff we have

$$\begin{aligned}
i\Pi(p) &= i\frac{g^2}{2} \int_0^1 dx \int_0^\Lambda \frac{d^4\ell_E}{(2\pi)^4} \frac{1}{[\ell_E^2 + \Delta]^2} \\
&= i\frac{g^2}{2} \int_0^1 dx \int_0^\Lambda \frac{d\ell_E}{(2\pi)^4} \frac{(2\pi^2)\ell_E^3}{[\ell_E^2 + \Delta]^2} \\
&= i\frac{g^2}{2} \int_0^1 dx \int_0^{\Lambda^2} \frac{dx}{(2\pi)^4} \frac{\pi^2 x}{[x + \Delta]^2} \\
&= i\frac{g^2}{32\pi^2} \int_0^1 dx \left( \log \left( 1 + \frac{\Lambda^2}{\Delta} \right) - \frac{\Lambda^2}{\Lambda^2 + \Delta} \right)
\end{aligned} \tag{2.24.15}$$

From this we see explicitly that there's a logarithmic divergence at large energies. This is very important and we'll discuss it more in a moment.

We can also do the  $x$  integral and then expand in large  $\Lambda$ , giving

$$i\Pi(p) \approx i\frac{g^2}{32\pi^2} \left( \log \left( \frac{\Lambda^2}{m^2} \right) - \frac{2\sqrt{p^2 - 4m^2} \tanh^{-1} \left( \frac{p}{\sqrt{p^2 - 4m^2}} \right)}{p} + 1 \right) \tag{2.24.16}$$

We obtain a simpler result in the  $m = 0$  limit, namely

$$i\Pi(p) \approx i\frac{g^2}{32\pi^2} \left( \log \left( \frac{\Lambda^2}{-p^2} \right) + 1 \right) \tag{2.24.17}$$

You can see that the  $+1$  can be removed by changing the cutoff by  $\Lambda \rightarrow \Lambda/\sqrt{e}$ .

But the  $\log(-p^2)$  cannot be removed by any change of  $\Lambda$ , and so it is a true physical effect. This is a very important point – the short distance scale  $\Lambda^2$  can never depend on the long-distance scale at which we do the experiment, namely  $p^2$ . We fix  $\Lambda$  once, when we set the theory.

## Ode to Logarithms

Why are logarithms interesting?

- They come from

$$\log \left( \frac{p^2}{\Lambda^2} \right) \sim \int_p^\Lambda \frac{dk}{k} \tag{2.24.18}$$

This means that *they are not dominated by any particular scale, but get non-trivial contributions at every energy/distance scale*. So the logarithmic effects are those that are always important, for any experiment, at any scale.

- They incorporate a hidden dependence on the short-distance scale which alters dimensional analysis – naively we would have expected that the result when  $m = 0$  could only depend on  $p^2$ , and so it would be totally fixed by scaling. But this ‘dimensional transmutation’ effect means that much more interesting results are possible.
- We must fix our masses and interaction strengths by experiments *at some particular scale*  $\mu$ , and then we make predictions at other scales. In this process, the scale  $\mu$  can appear, and so there will be dependence on  $\mu^2/p^2$  when we do an experiment at the scale set by  $p^2$ .

## Renormalization of a Force

We can immediately use our result within a t-channel Feynman diagram. We could evaluate it as a one-loop correction, or we can just write it with a resummed propagator

$$i\mathcal{M}(p) = \frac{-ig^2}{p^2 + \frac{g^2}{32\pi^2} \log \frac{\Lambda^2}{-p^2}} \quad (2.24.19)$$

If we just view this as a t-channel exchange, and thus a long-range force, then the momentum  $p_\mu$  is spacelike, so that  $p^2 < 0$ . This means it’s natural to write  $Q^2 = -p^2$  and

$$i\mathcal{M}(Q) = \frac{ig^2}{Q^2 - \frac{g^2}{32\pi^2} \log \frac{\Lambda^2}{Q^2}} \quad (2.24.20)$$

It’s also natural to define a dimensionless coupling by

$$\tilde{g}^2(Q) = \frac{g^2}{Q^2} \quad (2.24.21)$$

Now how do we eliminate the dependence on the cutoff  $\Lambda$ ? We want to actually measure this effect at some fixed scale  $Q^2 = \mu^2$ , writing a *renormalization condition*

$$i\mathcal{M}(\mu) \equiv i\tilde{g}_R^2(\mu) = \frac{i\tilde{g}_0^2}{1 - \frac{\tilde{g}_0^2}{32\pi^2} \log \frac{\Lambda^2}{\mu^2}} \quad (2.24.22)$$

in terms of the physical observable  $\mathcal{M}(\mu)$ , where  $\tilde{g}_0$  is the original or ‘bare’ coupling divided by  $Q^2$ . This gives

$$i\mathcal{M}(Q) = \frac{\tilde{g}_R^2(Q)}{1 - \frac{\tilde{g}_R^2(Q)}{32\pi^2} \log \left( \frac{\mu^2}{Q^2} \right)} \quad (2.24.23)$$

Now the scattering amplitude (from these specific diagrams) for any momentum  $Q$  has been written directly in terms of the renormalized coupling  $g_R$  defined at the scale  $\mu$ . Thus physical observables have been written in terms of other observables, and the short-distance cutoff  $\Lambda$  has been completely eliminated!

## UV Sensitivity of Relevant Parameters and the $d = 6$ Case

The results that we have obtained are a bit special in  $d = 4$ , due to the fact that the only interaction was  $g\phi^3$ , and that  $g$  has units of energy. This means that perturbation theory is strongly constrained by dimensional analysis alone.

We can fix this by working in  $d = 6$  spacetime dimensions instead. In that case, we must compute

$$i\Pi(p) = i\frac{g^2}{2} \int_0^1 dx \int \frac{d^6 \ell_E}{(2\pi)^6} \frac{1}{[\ell_E^2 + \Delta]^2} \quad (2.24.24)$$

Here we see that  $g$  is dimensionless, and the  $[energy]^2$  dimensions of  $\Pi(p)$  come from the  $d^6 \ell_E$ .

We already did a computation with a hard cutoff, so as a change of pace let's differentiate the integral with respect to  $\Delta$ . This gives

$$F''(\Delta) = \int \frac{d^6 \ell_E}{(2\pi)^6} \frac{6}{[\ell_E^2 + \Delta]^4} \quad (2.24.25)$$

where

$$i\Pi(p) = i\frac{g^2}{2} \int_0^1 dx F(\Delta) \quad (2.24.26)$$

The second derivative of  $F$  can be computed directly, as it is finite. It's useful to use the fact that the surface area of a  $d - 1$  sphere is

$$A_{d-1} = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \quad (2.24.27)$$

to write  $d^d \ell_E = A_{d-1} \ell_E^{d-1} d\ell_E$ , and so we find

$$F''(\Delta) = \frac{1}{64\pi^3} \frac{1}{\Delta} \quad (2.24.28)$$

This means that

$$F(\Delta) = \Lambda^2 + C\Delta + \frac{1}{64\pi^3} \Delta \log \Delta \quad (2.24.29)$$

$$= \Lambda_1^2 + \frac{1}{64\pi^3} \Delta \log \frac{\Delta}{\Lambda_2^2} \quad (2.24.30)$$

where  $\Lambda_1$  and  $\Lambda_2$  are integration constants (with units of energy) that were not determined by this method. These constants are directly sensitive to the short-distance physics, and would have been fixed if we had chosen a specific cutoff procedure, such as the hard cutoff we used above.

If we integrate over  $x$ , taking the  $m = 0$  limit again for algebraic simplicity, we find

$$i\Pi(p) = i\frac{g^2}{2} \left( \Lambda_1^2 - \frac{p^2}{384\pi^3} \log \left( \frac{-p^2}{\Lambda_2^2} \right) \right) \quad (2.24.31)$$

where I shifted  $\Lambda_2$  by a constant factor to simplify the result.

What does this result mean?

- It's natural to interpret this as a correction to the propagator, so

$$\frac{1}{p^2 - m_0^2 + \frac{g_0^2}{2} \left( \Lambda_1^2 - \frac{p^2}{384\pi^3} \log \left( \frac{-p^2}{\Lambda_2^2} \right) \right)} \quad (2.24.32)$$

- Here in  $d = 6$ , the coupling  $g$  is dimensionless, but  $\Pi(p)$  has units of  $[energy]^2$ , which are made up either by the cutoff  $\Lambda$  or by  $p^2$ . Thus we have a correction to the mass

$$\delta m^2 = \frac{g^2}{2} \Lambda_1^2 \quad (2.24.33)$$

which is of order the high-energy cutoff! This is our second signal of ‘fine tuning’ – in order to get a small physical mass for our  $\phi$  boson, we must carefully tune the bare mass  $m_0$  to cancel the  $\Lambda_1^2$  piece.

This completes our picture of fine-tuning for the Higgs boson mass – not only does dimensional analysis tell us that the Higgs boson mass might as well be huge, but *quantum corrections set the Higgs mass near the cutoff  $\Lambda$  unless we fine tune them away.*

- We also see that there's a renormalization of the  $p^2$  dependence, although this is weaker – it's only logarithmic – which means that it comes equally from all scales, and does not depend with great sensitivity on the short-distance physics.

To really define what we're doing, we need to fix the parameters of the theory by doing some particular experiment at some scale (some choice of  $p_\mu$ ). Then we can extrapolate to any other value of the momentum  $p_\mu$  flowing through the 2-pt correlator.

As a simplified version of this real experiment, we can do a thought experiment where we just fix the parameters at a *renormalization scale*  $p^2 = -\mu^2$ , and then extrapolate from there. We can again define a renormalized, physical mass  $m_R$  and a renormalized field  $\phi_R = \frac{1}{\sqrt{Z}}\phi$ , all at the *renormalization scale*  $p^2 = -\mu^2$ , via

$$\langle \Omega | T \{ \phi_R \phi_R \} | \Omega \rangle = \frac{1}{-\mu^2 - m_R^2} = \frac{1}{Z} \frac{1}{\mu^2 - m_0^2 + \frac{g_0^2}{2} \left( \Lambda_1^2 + \frac{\mu^2}{384\pi^3} \log \left( \frac{\mu^2}{\Lambda_2^2} \right) \right)} \quad (2.24.34)$$

which means that

$$Z = \frac{1}{1 + \frac{g_0^2}{384\pi^3} \log \left( \frac{\mu^2}{\Lambda_2^2} \right)} \quad (2.24.35)$$

$$m_R^2 = Z \left( m_0^2 - \frac{g_0^2}{2} \Lambda_1^2 \right) \quad (2.24.36)$$

in terms of the UV cutoff parameters  $\Lambda_1$ ,  $\Lambda_2$  and the bare parameters  $g_0$  and  $m_0$ .

Now the 2-pt function of the renormalized field at a general scale is

$$\begin{aligned}
\langle \Omega | T \{ \phi_R \phi_R \} | \Omega \rangle &= \frac{1 + \frac{g_0^2}{384\pi^3} \log \left( \frac{\mu^2}{\Lambda_2^2} \right)}{p^2 - m_0^2 + \frac{g_0^2}{2} \left( \Lambda_1^2 - \frac{p^2}{384\pi^3} \log \left( \frac{-p^2}{\Lambda_2^2} \right) \right)} \\
&= \frac{1}{p^2 \frac{1 + \frac{g_0^2}{384\pi^3} \log \left( \frac{-p^2}{\Lambda_2^2} \right)}{1 + \frac{g_0^2}{384\pi^3} \log \left( \frac{\mu^2}{\Lambda_2^2} \right)} - m_R^2} \\
&\approx \frac{1}{p^2 \left( 1 + \frac{g_0^2}{384\pi^3} \log \left( \frac{-p^2}{\mu^2} \right) \right) - m_R^2} \tag{2.24.37}
\end{aligned}$$

where in the last line we only kept the terms of order  $g_0^2$ . Thus we see that the physical renormalized 2-pt function only depends on the renormalization scale  $\mu$  and the renormalized mass  $m_R$ .

Is the mass  $m_R$  the physical mass? Only if we choose  $\mu^2 = -m_R^2$  as our renormalization scale. The physical mass is the place where the 2-pt function has a pole, so for general  $\mu$  it gets shifted from  $m_R$ . There's nothing wrong with this – no one forces us to use the physical mass as the renormalized mass!

Note that we haven't actually renormalized  $g_0$  yet. This is because we haven't considered general one-loop corrections to the interaction. Let's move on to do that.

## 2.25 General Loop Effects, Renormalization, and Interpretation

Let's go back and systematically compute all of the UV sensitive one-loop diagrams in our  $\phi^3$  theory, in order to complete the program of renormalization. *We want to show that all of the hidden dependence on very short distance physics can be encapsulated in a renormalization of  $\phi$ , the mass of  $\phi$ , and the interaction strength  $g$ .*

Another important point, which won't be illustrated with this example, is that *if the action we start with has certain symmetries, we (almost) never need to introduce any counterterms for interactions that would break those symmetries.*

This means writing the action in terms of a renormalized field  $\phi_R$  as

$$S = \int d^d x \left( \left( Z \frac{1}{2} (\partial \phi_R)^2 - \frac{1}{2} m_0^2 \phi_R^2 \right) - \frac{g_0}{6} \phi_R^3 \right) \tag{2.25.1}$$

We'll mostly focus on  $d = 6$  dimensions, where the coupling is marginal  $\equiv$  classically dimensionless. It's conventional, and convenient, to view these parameters as

$$Z = 1 + \delta Z, \quad m_0^2 = m_R^2 + \delta m^2, \quad g_0 = g_R + \delta g \tag{2.25.2}$$

where  $\delta Z, \delta m^2$ , and  $\delta g$  are called *counterterms*. The counterterms get their name because they are supposed to absorb the UV sensitivity (divergences aka infinities) when we do computations.

Also, the counterterms are always higher order in perturbation theory, where perturbation theory is defined as a formal expansion

$$g_0 \text{ or } g_R \quad (2.25.3)$$

Note that since (as we will see)

$$\delta g \sim g_0^3 \sim g_R^3 \quad (2.25.4)$$

doing perturbation theory in  $g_R$  and in  $g_0$  is equivalent. They are both (formally) infinitesimal parameters. We already saw above that

$$\delta Z \sim g^2, \quad \delta m^2 \sim g^2 \quad (2.25.5)$$

where  $g$  could be viewed as either  $g_0$  or  $g_R$ , since it doesn't matter at this order.

We computed the 2-pt function above, with various choices of regulator. What about other loop diagrams? A natural next step is to compute

$$\langle \phi(p_1)\phi(p_2)\phi(p_3) \rangle_{1\text{-loop}, 1PI} = (-ig)^3 \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2} \frac{i}{(k - p_1)^2 - m^2} \frac{i}{(k + p_3)^2 - m^2} \quad (2.25.6)$$

If we look at this diagram at very large  $k$ , its of order

$$\sim \int \frac{d^d k}{k^6} \quad (2.25.7)$$

and so it has a logarithmic divergence in  $d = 6$ , and is UV finite in  $d < 6$ . This also means that the 4-pt diagram with 4 propagators is finite in  $d < 8$ , where we will always stay. Thus this 3-pt diagram is the last divergence we will encounter at one-loop in  $d = 6$  dimensions.

That isn't to say that higher point diagrams aren't interesting and physical! For example there's a diagram

$$\langle \phi\phi\phi\phi \rangle_{1\text{-loop}} \supset (-ig)^4 \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2} \frac{i}{(k - p_1)^2 - m^2} \frac{i}{(k - p_1 - p_2)^2 - m^2} \frac{i}{(k + p_4)^2 - m^2}$$

This 'box diagram' is UV finite in  $d < 8$  dimensions, but that just means that it's completely unambiguous and calculable. It definitely affects e.g. the 2-to-2 scattering of  $\phi$  particles at order  $g^4$ . In other words, *it's dominated by energies and momenta of order the energies and momenta of the process (e.g. scattering) being studied.*

What we will now see (or argue for) is that all of the short-distance dependence can be absorbed into shifts of  $\delta Z, \delta m^2, \delta g$ . In other words, *once we do an experiment that fixes the physical normalization of  $\phi$ , the physical mass of the  $\phi$  particles, and the physical interaction strength, we can predict the result of any other experiment.* We'll prove this at one-loop, but at higher loops it won't be entirely obvious... the Wilsonian procedure we will discuss in a lecture or two will make it more or less 'obvious' at all loops.

To perform our general renormalization analysis, we will use dimensional regularization as our regulator, as it's nice and general and you may as well become familiar with it. For the 2-pt function renormalization, we needed to compute the integral

$$i\Pi(p) = i\frac{g^2}{2} \int_0^1 dx \int_0^\infty \frac{d^d \ell_E}{(2\pi)^d} \frac{1}{[\ell_E^2 + \Delta]^2} \quad (2.25.8)$$

There are a few different  $d$  dependence pieces. The area  $\Omega_d$  of a unit sphere in  $d$  dimensions is

$$\Omega_d = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}, \quad \text{where} \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{n!4^n} \sqrt{\pi} \quad (2.25.9)$$

It's also important to note that the gamma function has poles at

$$\Gamma(-n + x) = \frac{(-1)^n}{n!} \left[ \frac{1}{x} - \gamma + \sum_{k=1}^n \frac{1}{k} + \mathcal{O}(x) \right] \quad (2.25.10)$$

The coupling constant  $g$  will have units that depend on  $d$ , so it's most convenient to take

$$g \rightarrow g\tilde{\mu}^{\frac{6-d}{2}} \quad (2.25.11)$$

so that  $g$  is always dimensionless, and the parameter  $\mu$  with dimensions of energy soaks up the units.

The integral we need is (somewhat generalized, for reference)

$$\int \frac{d^d q}{(2\pi)^d} \frac{(q^2)^a}{(q^2 + \Delta)^b} = \frac{\Gamma(b - a - \frac{d}{2})\Gamma(a + \frac{d}{2})}{(4\pi)^{d/2}\Gamma(b)\Gamma(\frac{d}{2})} \Delta^{a + \frac{d}{2} - b} \quad (2.25.12)$$

In our specific case, for the evaluation of the 2-pt function, with  $\alpha = \frac{g^2}{(4\pi)^3}$  and  $\epsilon = 6 - d$  we have

$$\begin{aligned} \Pi(p) &= \frac{g^2}{2} \int_0^1 dx \int_0^\infty \frac{d^d \ell_E}{(2\pi)^d} \frac{1}{[\ell_E^2 + \Delta]^2} \\ &= \frac{\alpha}{2} \Gamma\left(\frac{\epsilon}{2} - 1\right) \int_0^1 dx \Delta \left(\frac{4\pi\tilde{\mu}^2}{\Delta}\right)^{\frac{\epsilon}{2}} \end{aligned} \quad (2.25.13)$$

Using the expansion for  $\epsilon = 6 - d$  small, we find

$$\begin{aligned} \Pi(p) &= \frac{1}{2}\alpha \int_0^1 dx \Delta \log(\Delta/\mu^2) \\ &\quad - \frac{1}{6}\alpha \left(\frac{1}{\epsilon} + \frac{1}{2}\right) p^2 \\ &\quad - \alpha \left(\frac{1}{\epsilon} + \frac{1}{2}\right) m^2 \end{aligned} \quad (2.25.14)$$

where we wrote

$$\mu = \sqrt{4\pi} e^{-\gamma/2} \tilde{\mu} \quad (2.25.15)$$

for convenience. The short distance divergences have shown up *as poles in  $\epsilon$* . That's how they always appear in dimensional regularization. Furthermore, note that the scale  $\mu$  we introduced to fix the dimensions of  $g$  appears to make up dimensions for logarithms. A feature of dimensional regularization is that it only produces poles in  $\epsilon$  and logarithms; it never produces explicit 'power divergences' such as the  $\Lambda^2$  we found with a hard cutoff. This is often a useful technical advantage, although it doesn't have physical meaning (in particular, the fine tunings associated with the masses of scalar fields do not go away simply because we use dim reg).

It is conventional to view the counter terms as a part of  $\mathcal{L}_{int}$ , making a compensating perturbative contribution to  $\Pi$  of the form

$$-i (\delta Z p^2 + \delta m^2) \quad (2.25.16)$$

Clearly we must take

$$\delta Z \approx -\frac{\alpha}{6\epsilon}, \quad \delta m^2 \approx -\frac{\alpha}{\epsilon} \quad (2.25.17)$$

in the limit that  $\epsilon \rightarrow 0$ , so that we cancel off the short-distance divergence. Aside from this, we can choose  $\delta Z$  and  $\delta m^2$  in whatever self-consistent manner is most convenient; this is called the choice of a *renormalization scheme*. The choice above (with an exact = sign) is called 'modified minimal subtraction'. This is just a common convention; it is not particularly important physically. Note, however, that shifts of  $\mu$  are exactly compensated by shifts of  $\delta Z$  and  $\delta m^2$ .

There is a physically sensible choice (although in more complicated theories and situations modified minimal subtraction is standard) where we demand that the full propagator has a pole at

$$p^2 = m_R^2 \quad (2.25.18)$$

with residue 1. In that case the exact propagator in this scheme takes the form

$$\frac{1}{p^2 - m_R^2 + i\epsilon} \left( \frac{1}{1 - \Pi(p)/(p^2 - m_R^2)} \right) \quad (2.25.19)$$

This is the dimensional regularization version of what we discussed above with differentiation or a hard cutoff.

Now let's consider

$$\langle \phi(p_1) \phi(p_2) \phi(p_3) \rangle_{1-loop, 1PI} = (-ig)^3 \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2} \frac{i}{(k - p_1)^2 - m^2} \frac{i}{(k + p_3)^2 - m^2} \quad (2.25.20)$$

We can use Feynman parameterization, defining

$$\int dF_3 = 2 \int_0^1 dx_1 dx_2 dx_3 \delta(x_1 + x_2 + x_3 - 1) \quad (2.25.21)$$



to write this as

$$g^3 \int dF_3 \int \frac{d^d \ell}{(2\pi)^d} \frac{1}{[\ell^2 + \Delta]^3} \quad (2.25.22)$$

where

$$\Delta = x_1 x_3 p_1^2 + x_2 x_3 p_2^2 + x_1 x_2 p_3^2 - m^2 \quad (2.25.23)$$

We can again use dimensional regularization to evaluate the result, which diverges logarithmically in  $d = 6$ . We find

$$\begin{aligned} & i \frac{1}{2} g \alpha \Gamma\left(\frac{\epsilon}{2}\right) \int dF_3 \left(\frac{4\pi \tilde{\mu}^2}{\Delta}\right)^{\epsilon/2} \\ &= i g \alpha \left[ \frac{1}{\epsilon} - \frac{1}{2} \int dF_3 \log\left(\frac{\Delta}{\mu^2}\right) \right] \end{aligned} \quad (2.25.24)$$

We can now use this result to fix the counter-term for  $\delta g$ , noting that it contributes as

$$-i \delta g \quad (2.25.25)$$

to the 3-pt correlator. This means that

$$\delta g = \frac{g \alpha}{\epsilon} \quad (2.25.26)$$

in modified minimal subtraction, which is conventional. But we could, alternatively, use some other condition that's found to be convenient.

We argued that only one-loop diagrams with 2 or 3 propagators can have divergences in  $d \leq 6$  dimensions. Thus by cancelling these divergences with  $\delta Z, \delta m^2$ , and  $\delta g$  we have eliminated all short-distance divergences at one-loop. It turns out that with these three parameters we can do that to all loop orders, although this is non-trivial to prove directly (due to overlapping divergences).

### Another Example – $\phi^4$ Theory in $d = 4$

As another example, let's consider the theory

$$S = \int d^d x \left( \left( Z \frac{1}{2} (\partial \phi_R)^2 - \frac{1}{2} m_0^2 \phi_R^2 \right) - \frac{\lambda_0}{24} \phi_R^4 \right) \quad (2.25.27)$$

in  $d = 4$  dimensions, where  $\lambda_0$  is classically marginal. We didn't use this theory before because it's not a general example, since, as we will see,  $Z = 1$  at one-loop.

We can first study the 2-pt function at one-loop, but it's just

$$\begin{aligned} \langle \Omega | T \{ \phi \phi \} | \Omega \rangle_{1-loop} &= i \lambda \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 - m^2} \\ &= \frac{i \lambda}{(2\pi)^4} \Lambda^2 \end{aligned} \quad (2.25.28)$$

in four dimensions. This has no non-trivial momentum dependence, and it simply requires us to adjust the counterterm for the mass  $\delta m^2$  to remove the dependence on the high-energy cutoff  $\Lambda$ . So that fixes  $Z = 1$ , and it determines  $\delta m^2$ .

The other diagram to compute is

$$\langle \Omega | T \{ \phi \phi \phi \phi \} | \Omega \rangle_{1-loop} = \frac{1}{2} (i\lambda)^2 \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2 + i\epsilon} \frac{i}{(P - k)^2 - m^2 + i\epsilon} \quad (2.25.29)$$

This is just one of three contributions – we get an  $s$ ,  $t$ , and  $u$ -channel contribution. But the integral is identical to the 2-pt function renormalization integral that we computed in  $\phi^3$  theory, except here we have the momentum  $P = p_1 + p_2$  running in the loop (for the  $s$ -channel, or other permutations for the other channels). We define  $\delta\lambda = \lambda_0 - \lambda_R$  in order to cancel the divergence.

Note that, conspicuously missing from the set of counterterms, is one for a  $g\phi^3$  type interaction. Why didn't we need such a counterterm? The reason is that there is a  $\phi \rightarrow -\phi$  symmetry that forbids such an interaction!

## 2.26 Large Logarithms and Renormalization Flows

We have performed a renormalization analysis of the  $g\phi^3$  theory, and also quickly considered the  $\lambda\phi^4$  theory. Let's look at some physical observables, and then systematize our findings.

A basic point will be that the renormalized field  $\phi_R$ , the renormalized mass  $m_R$ , and the renormalized coupling  $g_R$  must depend on the renormalization scale  $\mu$ , for two reasons, one mundane/technical and one deep/conceptual

- Physical observables cannot depend on  $\mu$ . Clearly the bare parameters do not. So the renormalized parameters must have a  $\mu$  dependence that cancels the explicit  $\mu$  dependence from quantum corrections due to loop diagrams.
- As we've emphasized, logarithms encode a contribution from every distance scale, and these accumulate over large hierarchies in scale  $\mu^2/p^2$ . Thus the effective coupling constant is not  $g$ , because perturbation theory looks like (parametrically)

$$g^2 + g^4 [\log(\mu^2/p^2) + 1] + g^6 [\log^2(\mu^2/p^2) + \log(\mu^2/p^2) + 1] + \dots \quad (2.26.1)$$

and thus perturbation theory breaks down if we don't resum the logarithms. This happens naturally if we define a running  $g(\mu)$  and then set  $\mu \approx p^2$  for our experiment.

### Analyzing a Physical Experiment at Loop Level and Resumming Logs

We will study a few physical observables at loop level. A key theme will be the explicit presence of the renormalization scale ' $\mu$ ', which is a completely arbitrary parameter that must drop out of all physical observables.

Recall that we found

$$\Pi_{\bar{M}\bar{S}}(p) = \frac{1}{2}\alpha \int_0^1 dx \Delta \log(\Delta/\mu^2) - \frac{\alpha}{12} (p^2 + 6m^2) \quad (2.26.2)$$

There is an alternative ‘on-shell’ scheme with a different choice of  $\delta m^2$  and  $\delta Z$  with

$$\Pi_{OS}(p) = \frac{1}{2}\alpha \int_0^1 dx \Delta \log(\Delta/\Delta_0) - \frac{\alpha}{12} (p^2 + m^2) \quad (2.26.3)$$

where

$$\Delta_0 = m^2[1 - x(1 - x)] \quad (2.26.4)$$

in this alternative scheme the propagator has a pole in  $p^2$  at  $m^2$  with residue one.

This means that

$$\begin{aligned} m_{phys}^2 &= m^2 - \frac{1}{2}\alpha \left[ -\frac{5}{6}m^2 + \int_0^1 dx D_0 \log(D_0/\mu^2) \right] \\ &= m^2 \left[ 1 + \frac{5}{12}\alpha \log(\mu^2/m^2) + \frac{34 - 3\pi\sqrt{3}}{15} \right] \end{aligned} \quad (2.26.5)$$

Clearly the physical mass  $m_{phys}$  cannot depend on  $\mu$ . There’s nothing to worry about though, all this means is that *the renormalized mass  $m$  in the modified minimal subtraction scheme does depend on  $\mu$* , in exactly the right way so that the  $\mu$  dependence cancels.

We also see that on a formal level, if  $\mu \gg m$ , then we have a large logarithm appearing, which can be resummed by choosing  $\mu \approx m$ , deriving a differential equation, and integrating it. The equation is

$$\frac{dm_{phys}}{d \log \mu} = 0 \quad (2.26.6)$$

which gives

$$\frac{dm}{d \log \mu} = -\frac{5}{12}\alpha m \quad (2.26.7)$$

This is called an *anomalous dimension* for the mass parameter; it tells us how we have to change the mass parameter when we change the renormalization scale.

We can do some hard work and compute the high-energy limit of the scattering cross section for  $\phi\phi \rightarrow \phi\phi$ , so  $s \gg m^2, \mu^2$ . The squared amplitude in this scheme is

$$|\mathcal{M}|_{obs}^2 \approx |\mathcal{M}_0|^2 \left[ 1 - \frac{3}{2}\alpha \log(s/\mu^2) + \frac{1}{3}\alpha \log(\delta^2) \right] \quad (2.26.8)$$

where  $\mathcal{M}_0$  is the tree-level scattering amplitude. The  $\delta$  term is an IR regulator. The point is, again, that  $\mu$  is a fake parameter, so we must have

$$\begin{aligned} 0 &= \frac{d}{d \log \mu} \log |\mathcal{M}|_{obs}^2 \\ &= \frac{2}{\alpha} \frac{d\alpha}{d \log \mu} + 3\alpha \end{aligned} \quad (2.26.9)$$

so we find that we must have

$$\beta(\alpha) \equiv \frac{d\alpha}{d \log \mu} = -\frac{3}{2}\alpha^2 \quad (2.26.10)$$

for the renormalized coupling. You will check in the homework that, in the simpler  $\phi^4$  theory, there are  $\log^2 s$  terms at higher order in perturbation theory. In this theory, the effective coupling is of order

$$\alpha \log s \quad (2.26.11)$$

and not just  $\alpha$ . Thus we must resum these large logarithms at large energy by setting  $\mu^2 \sim s$ , using this  $\beta$  function differential equation.

This  $\beta$  function means that the strength of the coupling, and of the associated force, changes logarithmically with the energy or distance scale. In this case, in  $\phi^3$  theory in  $d = 6$  dimensions, we have found that

$$\alpha \rightarrow 0 \quad \text{as} \quad \mu \rightarrow \infty \quad (2.26.12)$$

so in fact, the theory becomes more and more weakly coupled at large energies, or short distances. Such theories are referred to as *asymptotically free*, and QCD and the BCS theory of superconductivity are both examples of such theories. These theories have the possibility of being well defined down to arbitrarily short distances.

### Systematics of Schemes and $\beta$ Functions

We do not need to study a true physical observable to compute the  $\beta$  function. Using the action

$$S = \int d^d x \left( Z \frac{1}{2} (\partial \phi_R)^2 - \frac{1}{2} (m^2 + \delta m^2) \phi_R^2 - \frac{g + \delta g}{6} \phi_R^3 \right) \quad (2.26.13)$$

$$= \int d^d x \left( \left( \frac{1}{2} (\partial \phi_0)^2 - \frac{1}{2} m_0^2 \phi_0^2 \right) - \frac{g_0}{6} \phi_0^3 \right) \quad (2.26.14)$$

we have the relations between the bare and renormalized fields

$$\phi_0 = \sqrt{Z} \phi \quad (2.26.15)$$

$$m_0^2 = \frac{m^2 + \delta m^2}{Z} \quad (2.26.16)$$

$$g_0 = \frac{g + \delta g}{Z^{3/2}} \quad (2.26.17)$$

When we use dimensional regularization and the ‘modified minimal subtraction’ scheme, we can write the counterterms as

$$Z = 1 + \sum_{n=1} \frac{a_n(\alpha)}{\epsilon^n} \quad (2.26.18)$$

$$\frac{\delta m^2}{m^2} = \sum_{n=1} \frac{b_n(\alpha)}{\epsilon^n} \quad (2.26.19)$$

$$\frac{\delta g}{g} = \sum_{n=1} \frac{c_n(\alpha)}{\epsilon^n} \quad (2.26.20)$$

For example we found above that

$$a_1(\alpha) = -\frac{1}{6}\alpha + \dots \quad (2.26.21)$$

$$b_1(\alpha) = -\alpha + \dots \quad (2.26.22)$$

$$c_1(\alpha) = -\alpha + \dots \quad (2.26.23)$$

As we have stated several times before, all physical observables and bare parameters must be independent of  $\mu$ .

In the case of bare parameters, the reason is that they are what define or specify the theory at very short distances, so in a sense, they are physical. If we could compute exactly, physical observables could be expressed purely in terms of these bare parameters. So they must be independent of  $\mu$ .

To start with, consider  $g_0$ . We can define

$$\alpha_0 = \frac{g_0^2}{(4\pi)^3} = \alpha \frac{\left(1 + \frac{\delta g}{g}\right)^2}{Z^3} \tilde{\mu}^{6-d} \quad (2.26.24)$$

and its logarithm is

$$\log \alpha_0 = \log \alpha + \epsilon \log \tilde{\mu} + G(\alpha, \epsilon) \quad (2.26.25)$$

where

$$G(\alpha, \epsilon) = \log \left( \left(1 + \frac{\delta g}{g}\right)^2 \frac{1}{Z^3} \right) \quad (2.26.26)$$

Note that this means that  $G(\alpha, \epsilon)$  has an expansion in  $1/\epsilon^n$  starting at  $n = 1$  with

$$G(\alpha, \epsilon) = \sum_{n=1}^{\infty} \frac{G_n(\alpha)}{\epsilon^n} \quad (2.26.27)$$

Now we can differentiate with respect to  $\log \mu$  to get

$$0 = \frac{1}{\alpha} \frac{d\alpha}{d \log \mu} + \epsilon + \frac{\partial G}{\partial \alpha} \frac{d\alpha}{d \log \mu} \quad (2.26.28)$$

regrouping gives

$$0 = \left(1 + \frac{\alpha G'_1}{\epsilon} + \frac{\alpha G'_2}{\epsilon^2} + \dots\right) \frac{d\alpha}{d \log \mu} + \epsilon \alpha \quad (2.26.29)$$

Thus to have a physically sensible result in the limit  $\epsilon \rightarrow 0$ , we must be able to expand the rate of change of  $\alpha$  in  $\epsilon$  and have a finite limit as  $\epsilon \rightarrow 0$ , which means

$$\frac{d\alpha}{d \log \mu} = -\epsilon \alpha + \beta(\alpha) \quad (2.26.30)$$

with

$$\beta(\alpha) = \alpha^2 G'_1(\alpha) \quad (2.26.31)$$

and higher order terms in  $\epsilon$  must cancel, a result that can be checked in perturbation theory. Those cancellations should accord with the resummation of logarithms.

From our previous calculations, this tells us that in our theory

$$\beta(\alpha) = -\frac{3}{2}\alpha^2 \quad (2.26.32)$$

which agrees with what we found by directly studying a physical observable, such as the 2-to-2 scattering amplitude. Note, again, that the coupling goes to zero at large  $\mu$ , or short distances, so our theory is asymptotically free.

But this analysis was much simpler, and only required us to renormalize the logarithmically divergent corrections to the action for our  $\phi^3$  theory. It's crucial that these effects come from logarithms, which, as we have emphasized, accrue contributions to physical observables at every energy or distance scale.

We can repeat this analysis for the other parameters in the theory. For example, defining

$$M(\alpha, \epsilon) = \log \left( \frac{1 + \frac{\delta m}{m}}{\sqrt{Z}} \right) = \sum_{n=1}^{\infty} \frac{M_n(\alpha)}{\epsilon^n} \quad (2.26.33)$$

we find that

$$\gamma_m(\alpha) \equiv \frac{1}{m} \frac{dm}{d \log \mu} = \alpha M'_1(\alpha) \approx -\frac{5}{12}\alpha \quad (2.26.34)$$

as we found above.

Let's perform the same analysis for the propagator. The bare propagator is

$$D_0(p) = Z D(p) \quad (2.26.35)$$

so if we take the logarithm and differentiate, we have

$$\begin{aligned} 0 &= \frac{d}{d \log \mu} \log D_0(p) \\ &= \frac{dZ}{d \log \mu} + \frac{d}{d \log \mu} \log D(p) \\ &= \frac{dZ}{d \log \mu} + \frac{1}{D(p)} \left( \frac{\partial}{\partial \log \mu} + \frac{d\alpha}{d \log \mu} \frac{\partial}{\partial \alpha} + \frac{dm}{d \log \mu} \frac{\partial}{\partial m} \right) D(p) \end{aligned} \quad (2.26.36)$$

We can define an anomalous dimension for the field as

$$\gamma_\phi \equiv \frac{1}{2} \frac{d \log Z}{d \log \mu} \approx \frac{\alpha}{12} \quad (2.26.37)$$

This means that we can write an equation

$$\left( \frac{\partial}{\partial \log \mu} + \beta(\alpha) \frac{\partial}{\partial \alpha} + \gamma_m(\alpha) m \frac{\partial}{\partial m} + 2\gamma_\phi(\alpha) \right) D(p) = 0 \quad (2.26.38)$$

which is called the *Callan-Symanzik equation* for the propagator.

### Another Example – $\phi^4$ Theory

Our other example theory is  $\lambda\phi^4$  theory, which has a dimensionless coupling in  $d = 4$  dimensions. In that case the only non-trivial diagrams are the 2-to-2 diagrams we discussed above, and there are 3, one for each channel, and so we have a divergent part

$$\begin{aligned} \langle \phi\phi\phi\phi \rangle_{div} &= \frac{3}{2} (-i\lambda)^2 \int \frac{d^d k}{(2\pi)^d} \left( \frac{i}{k^2 - m^2 + i\epsilon} \right)^2 \\ &= i \frac{3\lambda^2}{32\pi^2} \Gamma\left(\frac{\epsilon}{2}\right) \left( \frac{\tilde{\mu}^2}{m^2} \right)^{\frac{\epsilon}{2}} \end{aligned} \quad (2.26.39)$$

The counterterm contributes as

$$-i\delta\lambda \quad (2.26.40)$$

and we must choose it so that it cancels the pole in  $\epsilon$ , giving

$$\delta\lambda = \frac{3\lambda^2}{16\pi^2\epsilon} \quad (2.26.41)$$

Now we have that

$$\log \lambda_0 = \log \lambda + \epsilon \log \mu + G(\lambda, \epsilon) \quad (2.26.42)$$

where

$$G(\lambda, \epsilon) = \log \left( 1 + \frac{\delta\lambda}{\lambda} \right) \quad (2.26.43)$$

The coupling  $\lambda_0$  must be independent of the RG scale  $\mu$ . This tells us that

$$0 = \left( 1 + \frac{\lambda G'_1}{\epsilon} + \frac{\lambda G'_2}{\epsilon^2} + \dots \right) \frac{d\lambda}{d \log \mu} + \epsilon \lambda \quad (2.26.44)$$

Working through the algebra order by order in  $\epsilon$ , this says that

$$\beta(\lambda) = \lambda^2 G'_1(\lambda) \quad (2.26.45)$$

where  $G_n(\lambda)$  are the expansions of  $G$  in  $\epsilon^{-n}$ . So we find a beta function

$$\frac{d\lambda}{d\log\mu} = -(4-d)\lambda + \frac{3}{16\pi^2}\lambda^2 \quad (2.26.46)$$

Note that here we have a  $+$  sign in front, so the coupling gets stronger and stronger at high energies. This theory is *infrared free*, because the coupling will get arbitrarily weak at long distances. But such a theory cannot be complete by itself.

## Fixed Points

We have seen that for our  $\phi^3$  theory, the beta function is a non-trivial, non-vanishing function of the coupling  $\alpha$ . What if it were to vanish?

That would mean that the coupling had no scale dependence, so the theory is at a *scale-invariant fixed point*. In fact, interacting, unitarity, Poincaré invariant theories with scale invariance are believed to be fully conformal invariant. So if we find vanishing  $\beta$  functions, we most likely have a *Conformal Field Theory*.

To be clear, we would only have such a theory if we also set all other dimensionful parameters, such as particle masses  $m = 0$ . These parameters are not scale invariant unless they vanish.

We can find an example of this if we view  $d$  as a parameter, for example in our  $\phi^4$  example above, if we take

$$\frac{d\lambda}{d\log\mu} = 0 = -(4-d)\lambda + \frac{3}{16\pi^2}\lambda^2 \quad (2.26.47)$$

then we find a fixed point at  $\lambda = \lambda_*$

$$\lambda_* = \frac{16\pi^2}{3}(4-d) \quad (2.26.48)$$

This is the fixed point in the  $\epsilon$  expansion that analytically continues in  $d$  to the 3d Ising model.

In such a case, note that the Callan-Symanzik equation for the propagator (with vanishing masses) becomes

$$\left( \frac{\partial}{\partial \log \mu} + 2\gamma_\phi(\lambda_*) \right) D(p) = 0 \quad (2.26.49)$$

This is such a simple differential equation that we can solve it immediately. We find

$$D(p) = \frac{C}{p^2} \left( \frac{\mu^2}{p^2} \right)^{-\gamma_\phi(\lambda_*)} \quad (2.26.50)$$

where we used the engineering dimensions of  $D$  to get the  $1/p^2$  in front. We see that although the engineering dimensions must be unchanged, the scaling dimension of  $\phi$  has been altered by the anomalous dimension  $\gamma_\phi$ , explaining its designation. At an interacting fixed point (a CFT), all



operators (generically) get anomalous dimensions. Note that this provides a very general way of obtaining logarithms... since typically

$$\gamma_\phi = a\lambda_* + b\lambda_*^2 + \dots \quad (2.26.51)$$

when we expand in  $\gamma_\phi$  we have

$$\left(\frac{\mu^2}{p^2}\right)^{-\gamma_\phi(\lambda_*)} = 1 - \gamma_\phi \log\left(\frac{\mu^2}{p^2}\right) + \dots \quad (2.26.52)$$

and so logarithms emerge from the perturbative expansion of the anomalous scaling dimension.

Further discussion of CFTs and fixed points is beyond the scope of these notes, this semester, but we may return to this topic at some point.

## 2.27 QM Example of Wilsonian Renormalization

Now we will discuss another elementary example from quantum mechanics, analogous to the Wilsonian framework for renormalization.

The starting point is the ‘Rayleigh-Ritz’ method for studying the effects of a perturbation to the Hamiltonian. Let us consider the QM Hamiltonian

$$H = H_0 + \lambda V \quad (2.27.1)$$

where we have already diagonalized the  $H_0$  term with eigenstates  $|i\rangle$ . Then we can compute the matrix

$$V_{ij} = \langle i|V|j\rangle \quad (2.27.2)$$

and then just diagonalize the full Hamiltonian  $E_i\delta_{ij} + \lambda V_{ij}$ . The eigenstates and eigenvectors will solve the full Schrodinger equation.

As an example, let’s study an anharmonic oscillator, with  $V = x^4$  and  $H_0$  the Hamiltonian of a simple oscillator. This can be solved very simply in Mathematica, I recommend watching these lectures to see how (<http://pirsa.org/displayFlash.php?id=15080045>). The original eigenstates just have oscillator number  $|n\rangle$ , with

$$V = x^4 = \frac{(a + a^\dagger)^4}{4} \quad (2.27.3)$$

It’s straightforward to evaluate  $V_{nm} = \langle n|V|m\rangle$  using oscillators, especially in Mathematica. Numerically diagonalizing the resulting matrix  $H_{mn} = n\delta_{mn} + V_{mn}$  gives a high precision estimate of the energies, and one can also compute the wavefunctions very precisely as linear combinations of the harmonic oscillator wavefunctions.

But we aren’t especially interested in anharmonic oscillators. The reason we have introduced this system is to discuss renormalization. We would like to consider the question *how precise are the energies when we truncate the Hamiltonian to a  $\Lambda \times \Lambda$  matrix, and can we systematically improve*

the results by somehow including the contributions of the energy levels with  $n > \Lambda$ ? For this purpose, we want to break up the Hilbert space into

$$\mathcal{H}_L \oplus \mathcal{H}_H \quad (2.27.4)$$

corresponding to the ‘low energy’ and ‘high energy states, where the dividing line is set by  $\Lambda$ . We will only focus on the properties of the low energy states in  $\mathcal{H}_L$ . We want to systematically study this approximation as a function of  $\Lambda$ . The analogy in QFT will be to a cutoff energy or distance scale  $\Lambda$ . The dependence of couplings on  $\Lambda$  is the Wilsonian renormalization flow.

Thus we have two approximations; we write  $H = H_0 + gV$  and we want to (intelligently) truncate the spectrum at  $\Lambda$ . The original Schrodinger equation is

$$(H_0 + gV) |\psi\rangle = \epsilon |\psi\rangle \quad (2.27.5)$$

However, we will only be interested in  $\psi \in \mathcal{H}_L$ . Although  $H_0$  has already been diagonalized, so that it takes  $\mathcal{H}_L \rightarrow \mathcal{H}_L$ , there will be terms in  $V$  that take  $\mathcal{H}_L \rightarrow \mathcal{H}_H$ , and we want to account for these. We can write a block matrix equation for the exact eigenstates

$$\begin{pmatrix} H_{LL} - \epsilon & gV_{LH} \\ gV_{HL} & H_{HH} - \epsilon \end{pmatrix} \begin{pmatrix} \psi_L \\ \psi_H \end{pmatrix} = 0 \quad (2.27.6)$$

This says that

$$|\psi_H\rangle = \frac{g}{H_{HH} - \epsilon} V_{LH} |\psi_L\rangle \quad (2.27.7)$$

Plugging this relation into the equation for  $|\psi_L\rangle$ , we find an equation for the light modes

$$\left( H_{LL} + V_{LH} \frac{1}{H_{HH} - \epsilon} V_{HL} \right) |\psi_L\rangle = \epsilon |\psi_L\rangle \quad (2.27.8)$$

where we absorbed  $g$  into  $V$ . So far this is an exact relation, which shows how we can ‘integrate out’ the high energy modes and write down an effective Hamiltonian for the low-energy modes.

However, since the high energy modes have large expectation value for  $H_{HH}$ , we should expect that we can consider successive approximations:

$$\begin{aligned} H_{LL} |\psi_L\rangle &\approx \epsilon |\psi_L\rangle \\ \left( H_{LL} + V_{LH} \frac{1}{H_{HH}} V_{HL} \right) |\psi_L\rangle &\approx \epsilon |\psi_L\rangle \end{aligned} \quad (2.27.9)$$

and so on. To compute the inverse, we can just use  $H_{HH,0}$ , ie  $H_{HH}$  without  $V$ . Now we can compute the  $V_{LH} \frac{1}{H_{HH}} V_{HL}$  term in perturbation theory and see how big it is. Specifically, it’s nice to write

$$\left( V_{LH} \frac{1}{\epsilon - H_{HH}} V_{HL} \right)_{ij} = \int_{\Lambda}^{\infty} \frac{M(E)_{ij}}{\epsilon - E} dE \quad (2.27.10)$$

We can analyze  $M$  by writing

$$\begin{aligned} C(\tau)_{ij} &= \langle i|V(\tau/2)V(-\tau/2)|j\rangle \\ &= \int_0^\infty M(E)_{ij} e^{-\left(E - \frac{E_i + E_j}{2}\right)\tau} \end{aligned} \quad (2.27.11)$$

so that the large  $E$  behavior is the small  $\tau$  behavior. Now we can analyze the small  $\tau$  behavior via the  $V(x)V(0)$  OPE. So the high-energy behavior, and thus the error in truncation, can be understood using the OPE of the interaction.

We would like to study what happens when we change the cutoff that divides  $H$  and  $L$  modes. Can we incorporate the effects of the heavy modes that we are ‘integrating out’? The philosophy of Wilsonian renormalization is to see how the parameters of the theory must change as we change the cutoff. Note that to leading order in perturbation theory, we can just neglect the states above the cutoff, since they first appear at order  $V^2$ . What happens if we include these terms?

In the case of the anharmonic oscillator, we have that

$$H = \omega a^\dagger a + \lambda(a^\dagger + a)^4 \quad (2.27.12)$$

In the harmonic oscillator basis  $|n\rangle$ , the perturbation has elements that are only up to 4 units off of the diagonal. This means that  $V_{HL}$  can only connect states that are very close to the cutoff dimension  $\Lambda$ . We can study the size of the  $M(E)$  above explicitly by computing the  $V(\tau)V(-\tau)$  correlator that defines  $C(\tau)$  above. This would be an interesting exercise...

## 2.28 Wilsonian Renormalization Flows

We have discussed the ‘renormalization group equations’, namely the  $\beta$  function equations for the change of couplings with scale. So far we have motivated it in two ways

- Quantum corrections inevitably refer to some specific scale at which observables/parameters are defined, either the high energy cutoff scale  $\Lambda$  or the renormalization scale  $\mu$ , and have a logarithmic dependence on that scale (in perturbation theory). We need to re-absorb those logarithms in case they become large.
- The renormalization scale  $\mu$  is unphysical, and cannot appear in any observables, so we obtain an equation for the dependence of renormalized couplings on  $\mu$  by demanding that the bare parameters (related directly to physical observables) cannot depend on  $\mu$ .

Now we will discuss another idea for renormalization, based on the ‘Wilsonian’ philosophy. Probably the best reference to read about these ideas is Polchinski’s paper ‘Renormalization and Effective Lagrangians’. The idea of the Wilsonian philosophy is to take the cutoff seriously, and consider what happens if we lower it. This is closely related to the Kadanoff block spin RG you derived in the homework. In a continuum QFT, one imagines integrating out all momenta in the range

$$(\Lambda - \delta\Lambda)^2 < p^2 < \Lambda^2 \quad (2.28.1)$$

This is usually evaluated in Euclidean space, where  $p^2$  is positive definite, so that it makes much more sense. One obtains a new theory, ie the action changes as

$$S_\Lambda[\phi] \rightarrow S_{\Lambda-\delta\Lambda}[\phi] \quad (2.28.2)$$

and we change the couplings in a way that leaves the long-distance physics invariant.

Let's see how this works very explicitly. We have some lagrangian with a cutoff  $\Lambda$  written as

$$\begin{aligned} L &= \frac{1}{2}(\partial\phi)^2 - m^2\phi^2 - \frac{\lambda_4}{4!}\phi^4 - \frac{\lambda_6}{6!\Lambda^2}\phi^6 - \frac{\lambda_8}{8!\Lambda^4}\phi^8 - \dots \\ &= \frac{1}{2}(\partial\phi)^2 + L_{int}(\phi) \end{aligned} \quad (2.28.3)$$

We have chosen a ‘toy’ Lagrangian without derivative interactions for simplicity, and we work in  $d = 4$ . Note that purely by dimensional analysis, ignoring all interactions, we must have

$$\beta_n^{exact} \equiv \frac{\partial\lambda_n}{\partial\log\Lambda} \supset (n-4)\lambda_n \quad (2.28.4)$$

simply because we have scaled out factors of  $\Lambda^{n-4}$  from the definition of the couplings. Note that this is in part because we are in  $d = 4$  dimensions. These  $\beta$  functions *will not be the same* as those that we found for the dependence of couplings on the renormalization scale  $\mu$ .

Now we want to integrate out modes with energy-momentum in the range  $(\Lambda - \delta\Lambda)^2 < p^2 < \Lambda^2$ . What do such modes contribute to?

They contribute in tree diagrams and loop diagrams. So for example if

$$p_1 + p_2 + p_3 = p_4 + p_5 + p_6 \quad (2.28.5)$$

then we can have a correction to  $\lambda_6$  where

$$\delta\lambda_6 \sim \frac{\lambda_4^2}{\Lambda^2} \quad (2.28.6)$$

Similarly, there is a tree-level correction

$$\delta\lambda_8 \sim \frac{\lambda_4\lambda_6}{\Lambda^4} \quad (2.28.7)$$

In general, these tree level diagrams lead to a change in the action of the form

$$\frac{\delta S_{tree}}{\delta\Lambda} = \int d^4p (2\pi)^4 \frac{1}{p^2 + m^2} \delta(p^2 - \Lambda^2) \frac{\delta L_{int}}{\delta\phi(p)} \frac{\delta L_{int}}{\delta\phi(-p)} \quad (2.28.8)$$

One can in fact use other softer methods for cutting of the high momentum modes (the book makes a different choice, for example, and Polchinski's paper considers a general function); the physical results don't change.

There are also one-loop (and only one-loop) renormalization corrections. For example

$$\delta\lambda_4 \sim \frac{\lambda_6}{16\pi^2} \quad (2.28.9)$$

In general we get

$$\frac{\delta S_{1-loop}}{\delta\Lambda} = - \int d^4p (2\pi)^4 \frac{1}{p^2 + m^2} \delta(p^2 - \Lambda^2) \frac{\delta^2 L_{int}}{\delta\phi(p)\delta\phi(-p)} \quad (2.28.10)$$

from loop diagrams where we take an interaction and wrap it back around.

Thus we have the Wilson-Polchinski RG equation

$$\frac{\delta S}{\delta\Lambda} = \int d^4p \frac{(2\pi)^4}{p^2 + m^2} \delta(p^2 - \Lambda^2) \left[ \frac{\delta L_{int}}{\delta\phi(p)} \frac{\delta L_{int}}{\delta\phi(-p)} - \frac{\delta^2 L_{int}}{\delta\phi(p)\delta\phi(-p)} \right] \quad (2.28.11)$$

This is called an *exact RG equation* because we have not used any sort of approximations in deriving it. This sounds way too good to be true, but it is true!

The caveat is that even if we start out with many of the  $\lambda_n = 0$ , as soon as we change the cutoff  $\Lambda$ , all of the  $\lambda_n$  are immediately generated with non-zero values. So we have a simple RG equation, but we need to keep track of an infinite number of couplings to use it. This means, again, that the ‘ $\beta$  functions’ from this exact RG equation are *not the same* as those that we derived above using the renormalization scale  $\mu$ .

As a toy example, we can study just the  $\phi^4$  and  $\phi^6$  interactions. It turns out that the exact RG  $\beta$  functions take the form (with coefficients that I’m not putting in yet)

$$\beta_4^{exact} \approx -\frac{\lambda_6}{16\pi^2} \quad (2.28.12)$$

$$\beta_6^{exact} \approx 2\lambda_6 + \lambda_4^2 \quad (2.28.13)$$

An important point here is that the 2 dominates over the perturbative coupling dependence. This means that as  $\Lambda$  decreases,  $\lambda_6$  is driven very quickly to a point where

$$2\lambda_6 + \lambda_4^2 \approx 0 \quad \implies \quad \lambda_6 \approx -\frac{1}{2}\lambda_4^2 \quad (2.28.14)$$

This means that when we lower the cutoff  $\Lambda$ , it isn’t exactly the case that  $\lambda_6 = 0$ , but what is true is that  $\lambda_6$ , *and all of the other irrelevant couplings, like  $\lambda_8$ , are very quickly driven to a value that is determined by the marginal and relevant couplings, such as  $\lambda_4$* . It isn’t that the irrelevant couplings vanish at low energies (in this way of looking at the exact RG), but that they are entirely fixed in terms of the relevant and marginal couplings. The book shows this explicitly with some complicated-looking differential equation solving.

This also means that if we imagine that we fix a lower cutoff  $\Lambda_L$  and define the theory at a larger cutoff  $\Lambda_H$ , the boundary condition on the irrelevant couplings at  $\Lambda_H$  do not matter, so we can freely take  $\Lambda_H \rightarrow \infty$ , and only worry about specifying the relevant and marginal couplings as a boundary condition. The boundary condition for e.g.  $\lambda_6$  is ‘irrelevant’. The only case where that’s not true is when the irrelevant coupling breaks a symmetry of the low-energy Lagrangian, so that it would never be generated – in that case it’s UV value can be important in setting the rate for symmetry breaking processes, e.g. proton decay.

## Connecting Renormalization Scale and the Wilsonian RG<sup>3</sup>

But how does the Wilsonian exact RG equation connect with the  $\beta$  functions that we saw when studying the renormalized coupling, and its dependence on the renormalization scale  $\mu$ ?

The physical point is that *when we change the cutoff, it's possible to make a choice (which is an approximation) so that we always set  $\lambda_6 = 0$  while keeping the low-energy physics invariant*. This is sometimes called ‘the sawtooth’ for reasons that are clear from a graphic depiction of the coupling constant evolution.

Let us now use this idea to derive the usual  $\phi^4$  theory  $\beta$ -function or Callan-Symanzik or Gellman-Low (these are equivalent monikers) equation

$$\frac{\partial \lambda_4}{\partial \log \mu} = \frac{3\lambda_4^2}{16\pi^2} \quad (2.28.15)$$

from the exact RG equation, which looks nothing like this.

If we work it out with explicit coefficients, the exact RG takes the form

$$\beta_4^{exact} = -\frac{\lambda_6}{16\pi^2} \quad (2.28.16)$$

$$\beta_6^{exact} = 2\lambda_6 + 6! \frac{\lambda_4^2}{72} \Lambda \delta(|p_1 + p_2 + p_3| - \Lambda) - \frac{\lambda_8}{16\pi^2} \quad (2.28.17)$$

Note that the combinatorial factors come from

$$\frac{1}{2} \frac{\delta L_{int}}{\delta \phi(p)} \frac{\delta L_{int}}{\delta \phi(-p)} \propto \frac{1}{2} \left( \frac{4}{4!} \right)^2 = \frac{1}{72} \quad (2.28.18)$$

I kept a  $\lambda_8$  dependence for illustrative purposes. Note that  $\lambda_8$  will approach  $\lambda_4^3 \ll \lambda_4^2$  in perturbation theory, so we can drop it if we only want to compute the running of  $\lambda_4$  to leading order. We would need to keep it if we wanted the 2-loop Callan-Symanzik equation.

The idea of the sawtooth is to

1. Start out with some value of  $\lambda_4$ , but with  $\lambda_6 = 0$ .
2. Flow infinitesimally from  $\Lambda \rightarrow \Lambda - d\Lambda$ , changing both couplings.
3. Redefine  $\lambda_4$  so that  $\lambda_6 = 0$ , choosing a new renormalization flow trajectory with the same low-energy physics.

Following the first two steps, we integrate the exact RG equation to get

$$\lambda_6(\Lambda) = 10\lambda_4^2 (\theta(\Lambda - d\Lambda - |p_1 + p_2 + p_3|) - \theta(\Lambda - |p_1 + p_2 + p_3|)) \quad (2.28.19)$$

So when  $\Lambda \rightarrow \Lambda - d\Lambda$ , we have  $\lambda_6$  turning on infinitesimally, and it's negative. Note that the  $\lambda_4$  coupling doesn't change, because  $\lambda_6 = 0$  at  $\Lambda_0$ .

---

<sup>3</sup>The following analysis borrows from a little note written by Matt Baumgart, as well as from Polchinski's paper.

Now we want to compute the 4-pt function

$$G^{(4)} = G_4^{(4)}(\lambda_4, \lambda_6) + G_6^{(4)}(\lambda_4, \lambda_6) = G_4^{(4)}(\lambda'_4, 0) \quad (2.28.20)$$

The point of this equation is that we want to compute the contribution to the 4-pt correlator from both  $\lambda_4$  and  $\lambda_6$ , and find a new value  $\lambda'_4$  so that the contribution purely from the 4-pt function is equal to this sum.

Being careful about the contractions, we can compute

$$\begin{aligned} G_6^{(4)} &= i \frac{\lambda_4^2}{72\Lambda^2} 108 \int_{\Lambda-d\Lambda}^{\Lambda} \frac{d^4p}{(2\pi)^4} \frac{1}{p^2 - m^2} \\ &= i \frac{3\lambda_4^2}{2\Lambda^2} \frac{\Lambda d\Lambda}{8\pi^2} \end{aligned} \quad (2.28.21)$$

The 108 comes from noting that when we close the loop, since  $p_1 + p_2 + p_3 = p_4 + p_5 + p_6 \approx \Lambda$ , we must contract two of the external momenta with 1, 2, 3 and two with 4, 5, 6. This leads to a combinatorial factor of  $3 \times 3 \times 4! \times \frac{1}{2}$ . The  $3 \times 3$  comes from three choices for each end of the loop, the  $4!$  comes from the usual factor we get from 4-pt correlators, and the  $1/2$  comes from the ‘symmetry factor’ or the redundancy in contracting with either side of the 6-pt coupling.

Thus to absorb this into a change in  $\lambda$ , we must choose

$$\lambda_4(\Lambda - d\Lambda) = \lambda_4 - \frac{3\lambda_4^2(\Lambda)d\Lambda}{16\pi^2\Lambda} \quad (2.28.22)$$

which implies the Callan-Symanzik equation we have above. This is how one can relate the Wilsonian RG to the Callan-Symanzik equations for the renormalized couplings as a function of the renormalization scale.

## 2.29 Path Integrals

Some comments on path integrals:

- PIs are an equivalent, independent way of thinking about quantum mechanics, which are intuitively based on a gedanken-experiment limit of the double slit experiment.
- PIs make Lorentz invariance and the classical limit very obvious, but make Unitarity, ‘quantization’, and the Hamiltonian much less obvious.
- From the PI perspective of QFT, it’s also not at all obvious that the relevant asymptotic states are particles, perhaps because in many cases, such as in general CFT, they are not!
- In particular, one can define ‘time’ and ‘forward in time’ however one wishes. Euclidean, as opposed to Minkowski time can be used to study statistical mechanics. The partition function is just a path integral on periodically identified Euclidean time.
- The flexibility in defining ‘time’ also means that one can easily talk about CFTs in ‘radial quantization’, something we may discuss eventually.

- PIs make it clear that perturbation theory generates asymptotic series.
- One can identify non-perturbative effects directly and beautifully in the PI, as alternate classical solutions to the field equations.
- Gauge theories tend to be easier to discuss using the path integral, because they can be quantized in a more manifestly Lorentz-invariant way. We can use the PI to derive the photon propagator, and then to generalize to Non-abelian gauge theories.
- Srednicki's QFT book (available for free as a draft pdf on his website) entirely develops QFT using path integrals, so it might be a useful reference and alternative guide.

### 2.29.1 Gaussian Integrals

As we will see very soon, all path integral computations in perturbation theory involve computing Gaussian integrals. A general 1-dimensional Gaussian integral is

$$I(J) = \int dp e^{-\frac{1}{2}ap^2 + Jp} \quad (2.29.1)$$

Note that by differentiating with respect to  $J$ , we can get arbitrary ‘correlation functions’

$$\left( \frac{d}{dJ} \right)^k I(J) \Big|_{J=0} = \int dp p^k e^{-\frac{1}{2}ap^2} \quad (2.29.2)$$

Thus  $I(J)$  is a generating function for these correlation functions. We will use the same trick in QM and QFT. You saw some of this on an early problem set.

To evaluate this  $I(J)$ , we complete the square

$$I(J) = \int dp e^{-\frac{1}{2}a(p - \frac{J}{a})^2 + \frac{J^2}{2a}} \quad (2.29.3)$$

Since the integral converges, we can now just shift and rescale the integration variable to give

$$I(J) = \frac{1}{\sqrt{a}} e^{\frac{J^2}{2a}} \int dp e^{-\frac{1}{2}p^2} \quad (2.29.4)$$

Now the last integral can be evaluated by squaring it, so that we find

$$I(J) = \frac{2\pi}{\sqrt{a}} e^{\frac{J^2}{2a}} \quad (2.29.5)$$

We can immediately generalize this to matrix valued  $\mathbf{A}$  and vector valued  $\vec{J}$ , giving

$$\int d\vec{p} e^{-\frac{1}{2}\vec{p} \cdot \mathbf{A} \cdot \vec{p} + \vec{J} \cdot \vec{p}} = \frac{(2\pi)^n}{\sqrt{\det A}} e^{\frac{1}{2}\vec{J} \cdot \mathbf{A} \cdot \vec{J}} \quad (2.29.6)$$

where  $n$  is the dimension of the vectors and matrices. When  $n = \infty$  we have a path integral.



### 2.29.2 Path Integrals in QM

In any quantum mechanical system with ‘coordinates’ and ‘momenta’  $Q_a$  and  $P_a$  satisfying the canonical commutation relations

$$[Q_a, P_b] = i\delta_{ab} \quad (2.29.7)$$

with other commutators vanishing, we can obtain a PI description. These could be the  $\hat{x}$  and  $\hat{p}$  of ordinary 1-d NR QM.

We can either diagonalize the  $P_a$  or  $Q_a$ , using states

$$Q_a|q\rangle = q_a|q\rangle \quad (2.29.8)$$

or

$$P_a|p\rangle = p_a|p\rangle \quad (2.29.9)$$

Note that since both are a complete set, we can write

$$1 = \int dq|q\rangle\langle q| = \int dp|p\rangle\langle p| \quad (2.29.10)$$

Simply by virtue of the canonical commutation relations, we have that

$$\langle q|p\rangle = e^{iq_ap^a} \quad (2.29.11)$$

up to a normalization. This follows by noting that  $-i\partial/\partial q_a$  has to act on these wavefunctions as  $p_a$ .

These states should be generalized to Eigenstates of Heisenberg picture operators by choosing

$$Q_a(t)|q;t\rangle = q_a|q;t\rangle \quad (2.29.12)$$

$$P_a(t)|p;t\rangle = p_a|p;t\rangle \quad (2.29.13)$$

Note that these are eigenstates of the Heisenberg picture  $Q_a(t)$ , not the result of allowing  $|q;t_0\rangle$  to evolve with time. The eigenstates evolve with time according to  $e^{iHt}$  instead of  $e^{-iHt}$ .

We can evaluate a matrix element at a final time and an initial time by inserting many complete sets of states

$$\langle f|i\rangle = \int dq_n \cdots dq_1 \langle q_f|e^{-iH(t_f)\delta t}|q_n\rangle \langle q_n|\cdots|q_2\rangle \langle q_2|e^{-iH(t_2)\delta t}|q_1\rangle \langle q_1|e^{-iH(t_1)\delta t}|q_i\rangle \quad (2.29.14)$$

and integrating over them. We suppressed the  $a$  index labeling the different  $q_a$  in favor of different time steps. These can be evaluated by using the momentum eigenstates to write

$$\begin{aligned} \langle q';t+\delta t|q;t\rangle &= \int \frac{dp_a}{2\pi} \langle q';t+\delta t|p_a;t\rangle \langle p_a;t|q;t\rangle \\ &= \int \frac{dp_a}{2\pi} \langle q';t+\delta t|e^{-iH\delta t}|p_a;t+\delta t\rangle \langle p_a;t|q;t\rangle \\ &= \int \frac{dp_a}{2\pi} e^{-iH(q',p)\delta t + i(q'_a - q_a)p^a} \end{aligned} \quad (2.29.15)$$

Note that  $-iq_ap^a$  in the exponent comes from one inner product, while  $iq'_ap^a$  comes from the other.

If the Hamiltonian is general then this is the best we can do. When we take the limit that  $n \rightarrow \infty$  and  $\delta t \rightarrow 0$ , we can label the  $q_a$  and  $p_a$  at different times by the paths

$$q_a(t), p_a(t) \quad (2.29.16)$$

Then we find that

$$\langle q'; t' | q; t \rangle = \int \mathcal{D}q_a(t) \mathcal{D}p_a(t) e^{i \int_{t'}^t dt \dot{q}_a(t) p_a(t) - H(q(t), p(t))} \quad (2.29.17)$$

This is the path integral, although as is, it may look unfamiliar. The exponent is the action written in terms of  $p_a$  and  $q_a$ , whereas you are used to seeing it written in terms of  $q_a$  and  $\dot{q}_a$ .

This expression can be greatly simplified when the Hamiltonian takes the form

$$H(q, p) = \frac{p_a^2}{2m_a} + V(q_a, t) \quad (2.29.18)$$

so that all momentum dependence is explicit and quadratic, we can do the  $dp_a$  integral immediately, since it's simply a Gaussian integral for each  $p_a(t)$ . This gives (with discrete  $q_j$  and discrete time)

$$\begin{aligned} \langle q_{j+1} | e^{-iH\delta t} | q_j \rangle &= N e^{-iV(q_j, t_j)\delta t + i \frac{m_a}{2} \frac{(q_{j+1} - q_j)^2}{\delta t^2}} \\ &= N e^{iL(q(t), \dot{q}(t))\delta t} \end{aligned} \quad (2.29.19)$$

and the Lagrangian is

$$L = \frac{1}{2} m_a \dot{q}_a^2 - V(q_a, t) \quad (2.29.20)$$

Thus we have a Lagrangian version of the path integral

$$\langle f | i \rangle = N \int_{q(t_i)=q_i}^{q(t_f)=q_f} \mathcal{D}q(t) e^{iS[q(t)]} \quad (2.29.21)$$

where the action

$$S = \int_{t_i}^{t_f} L(q(t), \dot{q}(t), t) dt \quad (2.29.22)$$

In essentially all cases we will use this Lagrangian formulation as our starting point.

### 2.29.3 And in QFT

Since we indexed the  $q_a$  and  $p_a$  so that we can accommodate an arbitrary number of them, generalizing to QFT is formally trivial. As we recall from canonical quantization, the  $q_a$  are just the fields  $\phi(x)$ . We need only note that

$$q_a(t) \rightarrow \phi(x, t), \quad \dot{q}_a(t) \rightarrow \dot{\phi}(x, t) \quad (2.29.23)$$

where  $x$  is equivalent to the  $a$  index label, when we talk about the Lagrangian. In the case of the Hamiltonian, we have

$$p_a(t) \rightarrow \pi(x, t) \quad (2.29.24)$$

since the canonical momentum is an independent variable in the Hamiltonian formalism.

Note that this means that we are working in the  $|\phi(x)\rangle$  basis of definite values of

$$\hat{\phi}(y)|\phi_c(x)\rangle = \phi_c(y)|\phi_c(x)\rangle \quad (2.29.25)$$

labeled by classical configurations  $\phi_c(x)$  for the field, where  $\hat{\phi}(y)$  is the Heisenberg picture operator. This is just one basis among many.

It's worth emphasizing, though, that the result

$$\langle \phi_f(x), t_f | \phi_i(x), t_i \rangle = \int_{\phi(x, t_i) = \phi_i(x)}^{\phi(x, t_f) = \phi_f(x)} \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \quad (2.29.26)$$

isn't very closely connected to the QM version with, say, multi-particle QM configurations. Here we are integrating over all classical field configurations  $\phi(x, t)$  at all times subject to some boundary conditions on the classical fields at initial and final times. The states of our QFT are being defined by classical  $\phi(x, t)$  configurations, not by some number of creation operators acting on the vacuum. This is a basis for QFT that is very different from the basis formed by  $n$ -particle states.

Note that if we take  $\phi(x) = 0$  at the initial and final times, and take  $t_f \rightarrow \infty$  and  $t_i \rightarrow -\infty$ , it looks like we have

$$\langle 0; \infty | 0; -\infty \rangle = \int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \quad (2.29.27)$$

where we integrate over all times, in a manifestly Lorentz-invariant way.

#### 2.29.4 Classical Limit

The classical limit obtains from (putting  $\hbar$  back in)

$$\langle f | i \rangle = N \int_{q(t_i) = q_i}^{q(t_f) = q_f} \mathcal{D}q(t) e^{\frac{i}{\hbar} S[q(t)]} \quad (2.29.28)$$

which follows by dimensional analysis, since  $S$  and  $\hbar$  both have units of action. When we take  $\hbar \rightarrow 0$ , this is dominated by those field configurations where

$$\frac{\delta S}{\delta q_a(t)} = 0 \quad (2.29.29)$$

which gives the classical equations of motion or Euler-Lagrange equations. This is because we are using the method of stationary phase approximation for the path integral (it's a very important kind of approximation; look it up if you don't know about it).

It's worth pausing and thinking about this some more. What are we going to do when we evaluate path integrals in interacting theories? We will start with a classical solution to the field equations, and write

$$\phi(x) = \phi_c(x) + \delta\phi(x) \quad (2.29.30)$$

Then we will compute small quantum fluctuations about the classical solution, which is a stationary point of the path integral. The simplest stationary point in the theories that we have studied is

$$\phi_c(x) = 0 \quad (2.29.31)$$

but there may be other, non-trivial stationary points, and one could expand about those as well. This is how one observes non-perturbative phenomenon in the PI formalism, including ‘instanton’ effects and *tunneling between different vacua in QFT*.

This also makes it easy to see why perturbation theory in QM and QFT is an expansion in  $\hbar$ . Let's write out the path integral for one of our favorite theories

$$\langle \phi_f(x), t_f | \phi_i(x), t_i \rangle = \int_{\phi(x, t_i) = \phi_i(x)}^{\phi(x, t_f) = \phi_f(x)} \mathcal{D}\phi(x, t) e^{\frac{i}{\hbar} \int d^4x \frac{1}{2} (\partial\phi)^2 - \frac{\lambda}{4!} \phi^4} \quad (2.29.32)$$

Now  $\phi$  is nothing but a dummy integration variable (!), so we can rescale it by  $\phi \rightarrow \phi/\sqrt{\lambda}$ . Now up to an overall normalization constant, we have

$$\langle \phi_f(x), t_f | \phi_i(x), t_i \rangle = \int_{\phi(x, t_i) = \phi_i(x)}^{\phi(x, t_f) = \phi_f(x)} \mathcal{D}\phi(x, t) e^{\frac{i}{\lambda\hbar} \int d^4x \frac{1}{2} (\partial\phi)^2 - \frac{1}{4!} \phi^4} \quad (2.29.33)$$

So we see that expanding in small  $\lambda$  is exactly the same thing as expanding in small  $\hbar$ . This is true for all perturbative couplings involving polynomial interactions.

### 2.29.5 Time Ordered Correlators

Now let us see what happens if we insert fields into the path integral. If we evaluate

$$\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \phi(y, t_y) \quad (2.29.34)$$

then we can interpret this by using the definition of the PI as a limit of many insertions of a sum over all states. We summed over states in the  $|\phi(x)\rangle$  basis of definite values for the field  $\phi(x)$ . Thus we get

$$\langle 0; \infty | \hat{\phi}(y, t_y) | 0; -\infty \rangle \quad (2.29.35)$$

because at the time  $t_y$ , this operator acts to give us the value of  $\phi(y, t_y)$ .

If we insert two fields we get

$$\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \phi(y_1, t_1) \phi(y_2, t_2) = \langle T\{\phi(y_1, t_1) \phi(y_2, t_2)\} \rangle \quad (2.29.36)$$

The reason for the time ordering is that no matter which order we use in the path integral, ie

$$\phi(y_1, t_1)\phi(y_2, t_2), \quad \phi(y_2, t_2)\phi(y_1, t_1) \quad (2.29.37)$$

we get the later time on the left, since the time ordering follows from the ordering of states  $|\phi, t\rangle\langle\phi, t|$  inserted to define the path integral. Thus we get time ordering for free in the PI formalism!

To get normalized interacting time ordered correlators, eliminating vacuum bubbles, we need only compute

$$\langle\Omega|T\{\phi(y_1)\cdots\phi(y_n)\}|\Omega\rangle = \frac{\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \phi(y_1)\cdots\phi(y_n)}{\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]}} \quad (2.29.38)$$

This formula makes it possible to use the path integral and LSZ to obtain S-Matrix elements for the scattering of particles. Generally speaking, we do not include particles via boundary conditions on the path integral (which might have been an intuitive thing to try to do), instead we just compute time ordered correlators and apply LSZ to get scattering amplitudes.

## 2.30 Generating Function(al)s and Feynman Rules

### 2.30.1 Generating Functions

We already saw that in the case of Gaussian integrals, by defining

$$Z_0(J) = \int dp e^{-\frac{1}{2}ap^2 + Jp} \quad (2.30.1)$$

and differentiating with respect to  $J$ , we can get an arbitrary ‘correlation function’

$$\left(\frac{d}{dJ}\right)^k Z_0(J) \Big|_{J=0} = \int dp p^k e^{-\frac{1}{2}ap^2} \quad (2.30.2)$$

This will generalize immediately to PIs in QFT.

Note that we could have defined a generating function in a theory with a more complicated ‘action’, such as

$$Z(J) = \int dp e^{-\frac{1}{2}ap^2 - \frac{1}{4!}\lambda p^4} e^{Jp} \quad (2.30.3)$$

Even though we cannot evaluate this integral in closed form, we can still obtain the ‘correlator’

$$\left(\frac{d}{dJ}\right)^k Z(J) \Big|_{J=0} = \int dp p^k e^{-\frac{1}{2}ap^2 - \frac{1}{4!}\lambda p^4} \quad (2.30.4)$$

by taking derivatives of  $Z(J)$ .

But we can go further. As you saw in an early problem set, we can evaluate such an integral in perturbation theory, assuming  $\lambda \ll 1$ , by expanding the exponential

$$Z(J) = \int dp e^{-\frac{1}{2}ap^2} e^{Jp} \left[ \sum_n \frac{1}{n!} \left( -\frac{1}{4!} \lambda p^4 \right)^n \right] \quad (2.30.5)$$

To work to a given order in  $\lambda$  we need only expand to that order.

But now we are just evaluating gaussian integrals multiplied by  $p^{4n}$ . We can therefore re-write this in a clever way as

$$\begin{aligned} Z(J) &= \sum_n \frac{1}{n!} \left( -\frac{1}{4!} \lambda \left( \frac{\partial}{\partial J} \right)^4 \right)^n \int dp e^{-\frac{1}{2}ap^2} e^{Jp} \\ &= \exp \left[ -\frac{1}{4!} \lambda \left( \frac{\partial}{\partial J} \right)^4 \right] Z_0[J] \end{aligned} \quad (2.30.6)$$

This shows how correlators in the ‘interacting’ theory can be written as a sum over correlators in the free theory. Of course we know the answer that

$$Z_0[J] = e^{\frac{J^2}{2a}} \quad (2.30.7)$$

and so we can write out the expansion for  $Z[J]$  using ‘Feynman diagrams’, where each vertex gets a  $\lambda$  and the vertices are connected by propagators associated with factors of  $\frac{1}{a}$ .

There are connected and disconnected diagrams, as usual. However, something nice happens when we think about them in terms of generating functions. Let us imagine that

$$W[J] \quad (2.30.8)$$

is the generating function purely made up of a sum of connected diagrams. Then we can get  $Z[J]$  as

$$Z[J] = 1 + W[J] + \frac{W[J]^2}{2} + \frac{W[J]^3}{6} + \dots \quad (2.30.9)$$

because this includes any number of disconnected diagrams from products of connected diagrams. But this just means that

$$W[J] = \log Z[J] \quad (2.30.10)$$

so connected diagrams are generated by  $\log Z$ . Thus we can get only connected Feynman diagrams by using  $W[J]$  instead of  $Z[J]$ .

### 2.30.2 Generating Functionals and the Feynman Propagator

A functional is a function of a function. So for example, the wavefunction in a QFT in the  $\phi$ -basis is a functional

$$\Psi[\phi(x)] \quad (2.30.11)$$

that gives the quantum mechanical amplitude for some given classical configuration  $\phi(x)$ . This generalizes the wavefunction insofar as there are an infinite number of independent variables in  $\phi(x)$ . The best way to think about this is by imagining that we have discretized spacetime and put it into a finite box, so that there are only a finite number of spacetime lattice points  $x_i$ . We evaluate the field  $\phi(x_i)$  at each of these points, and so our wavefunction depends on all of the  $\phi(x_i)$  at all the lattice points. Taking the continuum limit of the lattice gives the wave functional.

So there is a very general and natural way to compute time ordered correlators, by using a generating functional

$$Z[J(x)] = \int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)] + i \int d^4x J(x)\phi(x)} \quad (2.30.12)$$

We define functional derivatives via

$$\frac{\delta}{\delta\phi(x)}\phi(y) \equiv \delta^4(x - y) \quad (2.30.13)$$

so that, for example

$$\begin{aligned} \frac{\delta}{\delta J(x)} e^{i \int d^4y J(y)\phi(y)} &= e^{i \int d^4y J(y)\phi(y)} \int d^4y \phi(y) \delta^4(x - y) \\ &= i\phi(x) e^{i \int d^4y J(y)\phi(y)} \end{aligned} \quad (2.30.14)$$

This is just the limit of what would happen if we discretized  $x$  and viewed  $\phi(x)$  at different values of  $x$  as a large but finite number of different (independent) variables.

We can therefore write

$$\langle T\{\phi(x_1)\phi(x_2)\} \rangle = (-i) \frac{\delta}{\delta J(x_1)} (-i) \frac{\delta}{\delta J(x_2)} Z[J(y)] \Big|_{J=0} \quad (2.30.15)$$

in order to compute the time ordered 2-point correlator in a QFT. Usually I will just write  $Z[J]$  for the generating functional.

This gives us a major hint as to what the generating functional  $Z[J(y)]$  must be for a free QFT. Now let us compute it.

We begin with

$$Z[J(y)] = \int \mathcal{D}\phi(x) e^{i \int d^4x \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - J\phi} \quad (2.30.16)$$

This is just an infinite number of Gaussian integrals, with a ‘vector’  $\phi(x)$  and a ‘matrix’

$$A = \partial^2 + m^2 \quad (2.30.17)$$

So it’s easy. The answer is, roughly speaking

$$Z[J(y)] \propto e^{\frac{1}{2}JA^{-1}J} \quad (2.30.18)$$

from our formula above. We just need to invert the matrix  $\partial^2 + m^2$ . How?

As usual, this becomes trivial in momentum space. We write

$$Z[\tilde{J}(p)] = \int \mathcal{D}\tilde{\phi}(k) e^{\frac{i}{2} \int \frac{d^4 k}{(2\pi)^4} \tilde{\phi}(k)(k^2 + m^2)\tilde{\phi}(-k) - \tilde{J}(k)\tilde{\phi}(-k) + \tilde{J}(-k)\tilde{\phi}(k)} \quad (2.30.19)$$

Now we change variables to

$$\tilde{\chi}(k) = \tilde{\phi}(k) - \frac{\tilde{J}(k)}{k^2 + m^2} \quad (2.30.20)$$

in order to complete the square. The result is that

$$Z_0[\tilde{J}(p)] = \exp \left[ \frac{i}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2 + m^2 + i\epsilon} \right] \quad (2.30.21)$$

or

$$Z_0[J(y)] = \exp \left[ \frac{i}{2} \int d^4 x d^4 x' J(x) D_F(x - x') J(x') \right] \quad (2.30.22)$$

where  $D_F(x, x')$  is the usual Feynman propagator. Note that this reproduces the result above for the second functional derivative with respect to  $J(x)$ , as claimed.

### 2.30.3 What about the $i\epsilon$ ?

The  $i\epsilon$  prescription actually did not appear in our computation, so where does it come from?

- One can essentially view it as a term that guarantees the convergence of the path integral, by inserting

$$e^{-\epsilon \int d^4 x \phi^2(x)} = e^{i \int d^4 x i\epsilon \phi^2(x)} \quad (2.30.23)$$

so that the PI is well defined. Alternatively, we can view this as selecting out the vacuum state.

- One can also obtain it by starting out in Euclidean space (as we'll discuss in the context of thermodynamics) and then analytically continuing back to Lorentzian signature.

### 2.30.4 Feynman Rules With Interactions

We can apply our trick from generating functions to generating functionals in order to derive the Feynman rules for interacting theories. We saw above that

$$\begin{aligned} \langle \Omega | T \{ \phi(y_1) \cdots \phi(y_n) \} | \Omega \rangle &= \frac{\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]} \phi(y_1) \cdots \phi(y_n)}{\int \mathcal{D}\phi(x, t) e^{iS[\phi(x, t)]}} \\ &= \frac{1}{Z[0]} (-i) \frac{\delta}{\delta J(y_1)} \cdots (-i) \frac{\delta}{\delta J(y_n)} Z[J(y)] \Big|_{J=0} \end{aligned} \quad (2.30.24)$$



But we would like to compute  $Z[J(x)]$ .

Just as we saw for generating functions, in perturbation theory we can write

$$\begin{aligned}
Z[J] &= \int \mathcal{D}\phi(x, t) \exp \left[ i \int d^4x \frac{1}{2} (\partial\phi)^2 - \frac{1}{2} m^2 \phi^2 - J\phi - \frac{\lambda}{4!} \phi^4 \right] \\
&= \int \mathcal{D}\phi(x, t) \exp \left[ i \int d^4x \frac{1}{2} (\partial\phi)^2 - \frac{1}{2} m^2 \phi^2 - J\phi \right] \sum_{n=0}^{\infty} \frac{1}{n!} \left( -i \int d^4y \frac{\lambda}{4!} \phi^4(y) \right)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \left( -i \int d^4y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 \right)^n \int \mathcal{D}\phi(x, t) \exp \left[ i \int d^4x \frac{1}{2} (\partial\phi)^2 - \frac{1}{2} m^2 \phi^2 - J\phi \right] \\
&= \exp \left[ -i \int d^4y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 \right] \exp \left[ \frac{i}{2} \int d^4x d^4x' J(x) D_F(x - x') J(x') \right] \\
&= \exp \left[ -i \int d^4y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 \right] Z_0[J(x)]
\end{aligned} \tag{2.30.25}$$

This expression gives us the full  $Z[J]$ , at least in a perturbative expansion.

For example, to compute the 2-pt function, we want

$$\langle \Omega | T \{ \phi(x_1) \phi(x_2) \} | \Omega \rangle = - \frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \exp \left[ -i \int d^4y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 \right] Z_0[J(x)] \Big|_{J=0} \tag{2.30.26}$$

If we work to zeroth order in  $\lambda$ , we just get the Feynman propagator itself. If we work to first order in  $\lambda$ , we get the one connected bubble diagram that we are used to, plus several other disconnected Feynman diagrams. Obviously if we have a scalar field theory with other types of vertices, then we can derive the Feynman rules in the same way.

### 2.30.5 Fermionic Path Integrals

As the book says, path integrals for fermions would like to be essentially equivalent to those for bosons, but fermions are anti-commuting ‘numbers’, which gets confusing. At the end of the day they basically are the same, except for the anti-commutation, but here’s how to justify that.

A Grassman algebra is a set of objects generated by a basis  $\{\theta_i\}$ , where the  $\theta_i$  are called Grassman numbers. They anticommute, so

$$\theta_i \theta_j = -\theta_j \theta_i \tag{2.30.27}$$

but they add commutatively, and can be multiplied by complex numbers (as though they are basis vectors in a vector space). Note that

$$\theta_i^2 = 0 \tag{2.30.28}$$

a crucial property that means series expansions in  $\theta$  truncate. So if there’s only one  $\theta$  then

$$g = a + b\theta \tag{2.30.29}$$

is the most general object in the algebra. With two thetas it's

$$g = A + B\theta_1 + C\theta_2 + F\theta_1\theta_2 \quad (2.30.30)$$

Objects with an even-grade commute with other objects, whereas objects with an odd grade anti-commute.

To define a path integral, we need to know how to integrate

$$\int \mathcal{D}\psi \quad (2.30.31)$$

where  $\psi$  is Grassman. How do we integrate anti-commuting numbers!?

Integrating over functions is hard because there are a lot of possible functions, but since the number of objects in a Grassman algebra is rather small, we can just work directly. The most general integral in one variable is

$$\int d\theta(a + b\theta) = a \int d\theta + b \int d\theta \theta \quad (2.30.32)$$

If we want to map non-trivial elements of the Grassman algebra to  $\mathbf{C}$ , then we choose the first term to vanish, and the second integral to give 1, meaning

$$\int d\theta(a + b\theta) \equiv b \quad (2.30.33)$$

Note that this is exactly the same thing as the naive definition of a derivative

$$\frac{d}{d\theta}(a + b\theta) = b \quad (2.30.34)$$

so integrals and derivatives are identical! With more variables, we need to be careful and set

$$\int d\theta_1 d\theta_2 \cdots d\theta_n \theta_n \theta_{n-1} \cdots \theta_1 = 1 \quad (2.30.35)$$

but these variables anti-commute, so there's a sign to keep track of, and we evaluate these from the inside out.

Why do we call this 'an integral'? One reason is that it has the property

$$\int d\theta f(\theta) = \int d\theta f(\theta + X) \quad (2.30.36)$$

as one can compute directly, for any  $X$  that doesn't depend on  $\theta$ . This is equivalent to the usual shift symmetry of integrals from  $-\infty$  to  $\infty$ .

When we study path integrals, we have Gaussian integrals like

$$\int d\bar{\theta} d\theta e^{-\bar{\theta} A \theta} = \int d\bar{\theta} d\theta (1 - \bar{\theta} A \theta) = A \quad (2.30.37)$$

That's the full Taylor expansion of the integrand, because  $\theta_i^2 = 0$ . We should compare that to the usual 2-variable Gaussian integral, which gives  $1/\sqrt{A}$ . We see that in general

$$\begin{aligned} \int d\bar{\theta}_1 \cdots d\bar{\theta}_n d\theta_1 \cdots d\theta_n e^{-\bar{\theta}_i A_{ij} \theta_j} &= \frac{1}{n!} \sum_{perms} \pm A_{i_1 i_2} \cdots A_{i_{n-1} i_n} \\ &= \det(\mathbf{A}) \end{aligned} \quad (2.30.38)$$

This is crucial for QFT, since most of what we compute will be built from Gaussian integrals, and we see that bosonic vs fermionic integrals just differ as to whether the determinant is in the numerator or denominator. In the presence of external currents we get

$$\int d\bar{\theta}_1 \cdots d\bar{\theta}_n d\theta_1 \cdots d\theta_n e^{-\bar{\theta}_i A_{ij} \theta_j + \bar{\eta}_i \theta_i + \bar{\theta}_i \eta_i} = \det(\mathbf{A}) e^{\bar{\eta} \mathbf{A}^{-1} \eta} \quad (2.30.39)$$

which is what we need to evaluate correlators.

Now let's take the continuum limit and compute the Dirac propagator. The path integral is

$$Z[\bar{\eta}, \eta] = \int \mathcal{D}\bar{\psi}(x) \mathcal{D}\psi(x) e^{i \int d^4x [\bar{\psi}(i\cancel{\partial} - m)\psi + \bar{\psi}\eta + \bar{\eta}\psi]} \quad (2.30.40)$$

We can immediately evaluate this as a Gaussian integral

$$Z[\bar{\eta}, \eta] = \mathcal{N} e^{i \int d^4x \int d^4y \bar{\eta}(y) (i\cancel{\partial} - m + i\epsilon)^{-1} \eta(x)} \quad (2.30.41)$$

where  $\mathcal{N}$  is an infinite constant from the determinant of the Dirac operator. We get the 2-pt function of this free theory from

$$\begin{aligned} \langle T\{\psi(x) \bar{\psi}(y)\} \rangle &= \frac{1}{Z[0]} \frac{\delta^2}{\delta \bar{\eta}(x) \delta \eta(y)} Z[\bar{\eta}, \eta] \Big|_{\bar{\eta}, \eta=0} \\ &= \int \frac{d^4p}{(2\pi)^4} \frac{i}{\not{p} - m + i\epsilon} e^{-ip(x-y)} \\ &= \int \frac{d^4p}{(2\pi)^4} \frac{i(\not{p} + m)}{p^2 - m^2 + i\epsilon} e^{-ip(x-y)} \end{aligned} \quad (2.30.42)$$

which is the Dirac propagator.

## 2.31 Path Integrals and Statistical Physics

We have been using the notation  $Z[J]$  for generating functionals, as though they are the same thing as the *Partition Function* of a statistical system. In fact, they almost are!

This surprising connection is deep and very important; it will not be a primary topic for us but it would be silly not to mention it briefly. It implies an intimate connection between thermodynamics and quantum mechanics, and the methods it suggests are used to study the statistical mechanics of relativistic (and non-relativistic) systems.

Let us derive a shocking statement: *the canonical partition function at temperature  $T$  is given by a Euclidean path integral, where we impose periodic boundary conditions on the Euclidean time  $t \rightarrow -i\tau$ , with a period  $\beta = 1/T$ .*

First of all, note that if we have any QM theory with a conventional kinetic term, then

$$iS = i \int dt \left[ \frac{1}{2} \dot{x}^2 - V(x) \right] \quad (2.31.1)$$

When we move to Euclidean space, we take  $t \rightarrow -i\tau$ , and find

$$\begin{aligned} iS &\rightarrow i \int (-i) d\tau \left[ -\frac{1}{2} \dot{x}^2 - V(x) \right] \\ &= - \int d\tau H(x, \partial_\tau x) \end{aligned} \quad (2.31.2)$$

This also works in QFT of course, with kinetic term  $(\partial_t \phi)^2 - (\partial_i \phi)^2$ . This means that under  $t \rightarrow -i\tau$ , we have

$$\int \mathcal{D}\phi(x) e^{i \int d^4 x \mathcal{L}[\phi]} \rightarrow \int \mathcal{D}\phi(x) e^{- \int d\tau d^3 x \mathcal{H}[\phi]} \quad (2.31.3)$$

So we see that the Lagrangian has turned into the Hamiltonian, and the phase has turned into an exponential suppression.

But what should be the measure of integration over  $\tau$ , and what does it mean? For that, consider the thermal expectation value

$$\langle \psi_f | e^{-\beta H} | \psi_i \rangle \quad (2.31.4)$$

Here the ‘initial’ and ‘final states are different. But when we compute the partition function we take the expectation value of  $e^{-\beta H}$  in a single state, setting  $\psi_f = \psi_i$ .

So the partition function is just

$$Z[\beta] = \sum_{\psi} \langle \psi | e^{-\beta H} | \psi \rangle \quad (2.31.5)$$

as long as the states of the theory  $|\psi\rangle$  are normalized. Just as we broke up the conventional time evolution into tiny slices, we can also imagine breaking up the Euclidean time evolution. The fact that the initial and final states are the same tells us that we must impose *periodic boundary conditions in Euclidean time*. Since

$$\beta H = \int_0^\beta d\tau \int d^3 x \mathcal{H}(\tau, \vec{x}) \quad (2.31.6)$$

we see that

$$Z[\beta] = \int_{\phi(0, \vec{x}) = \phi(\beta, \vec{x})} \mathcal{D}\phi e^{- \int_0^\beta d\tau \int d^3 x \mathcal{H}(\tau, \vec{x})} \quad (2.31.7)$$

is the partition function for our QFT at temperature  $\beta = 1/T$ !

One can take this further to get all sorts of thermodynamic information. Specifically, to get the density matrix at finite temperature we just compute

$$\begin{aligned}\rho_{\Psi_1, \Psi_2} &= \langle \Psi_1 | e^{-H\beta} | \Psi_2 \rangle \\ &= \int_{\Psi_1[\phi(0, x)]}^{\Psi_2[\phi(\beta, x)]} \mathcal{D}\phi e^{-\int_0^\beta d\tau \int d^3x \mathcal{H}(\tau, \vec{x})}\end{aligned}\tag{2.31.8}$$

The partition function is the trace of this object.

And we can use other boundary conditions to compute more exotic properties of theory. For example, what if we want to study the reduced density matrix after integrating out all the degrees of freedom in some region  $B$ , leaving only a density matrix  $\rho_A$  for what's left over? All we need to do is integrate over the fields in  $B$ , with boundary conditions imposed in the region  $A$ . Entanglement entropy has become a popular computable (it isn't observable) in condensed matter theory, in order to study exotic topological phases of matter, and also in high-energy physics, as a probe of spacetime.

## 2.32 Soft Limits and Weinberg's Theorem

Using Lorentz invariance we can prove that

- Massless spin 1 particles can only produce long-range forces by coupling to a conserved charge.
- Massless spin 2 particles can only produce long-range forces by coupling to energy-momentum.
- Massless higher spin particles can never produce long-range forces.

The key element is the Lorentz transformation of polarization vectors, as we emphasized earlier.

Consider some scattering amplitude, and let us add to it a *soft photon* with momentum  $q$ . In the soft limit, the amplitude will be dominated by soft photons that are attached to external legs. On the  $i$ th leg, this gives

$$M_i(p_i, q) = (-ieQ_i) \frac{i(p_i^\mu + (p_i^\mu - q))}{(p_i - q)^2 - m^2} \epsilon_\mu M_0(p_i - q)\tag{2.32.1}$$

where  $Q_i$  is the charge of the  $i$ th particle. Note that since  $p_i^2 = m^2$  and  $q^2 = 0$ , this is

$$M_i(p_i, q) \approx eQ_i \frac{p_i \cdot \epsilon}{p_i \cdot q} M_0(p_i)\tag{2.32.2}$$

to leading order at small  $q$ .

Now we see the reason that *external lines* dominate – it's because only there can we get poles in  $q$  as  $q \rightarrow 0$ . Another way of saying the same thing is that soft photons have very large wavelength, and so they do not care about short-distance processes. They act like a classical background against which the rest of the scattering process unfolds.

In fact, one can obtain the soft amplitude factor by viewing the external particles as classical sources for the photon field  $A_\mu(x)$ . The source, or classical current for the  $i$ th particle, is

$$J_i(x) = Q_i \delta^3(\vec{x} - \hat{q}_i t) \quad (2.32.3)$$

and so we can obtain the soft amplitude factor from

$$\begin{aligned} \frac{p_i \cdot \epsilon}{p_i \cdot q} &= \int_0^\infty d\tau e^{i\hat{n}_i \cdot A(x(\tau))} \\ &\approx \int_0^\infty d\tau i(\hat{n}_i \cdot \epsilon) e^{i\hat{n}_i \cdot q\tau} \end{aligned} \quad (2.32.4)$$

where we note that for a mode with fixed momentum  $q$  we have

$$A_\mu(x) = \epsilon_\mu e^{iq \cdot x} \quad (2.32.5)$$

We have obtained the soft factor from integrating the electromagnetic field  $A(x(\tau))$  along the classical worldline of each charged particle in the scattering process.

So we have the soft factor for emission from one external leg. This means that the total soft photon emission amplitude will be

$$M \approx eM_0(p_j) \left[ \sum_{incoming} Q_i \frac{p_i \cdot \epsilon}{p_i \cdot q} - \sum_{outgoing} Q_i \frac{p_i \cdot \epsilon}{p_i \cdot q} \right] \quad (2.32.6)$$

where the approximation means that we are only considering the leading term when  $q$  is small. Now we ask the crucial question: *is this Lorentz invariant?*

Because of the Lorentz transformation properties of  $\epsilon$ , we have

$$\epsilon_\mu \rightarrow (\Lambda\epsilon)_\mu + c(\Lambda)q_\mu \quad (2.32.7)$$

as we saw before. So in order for the emission amplitude to be Lorentz invariant, we must have that when we replace  $\epsilon \rightarrow q$ , the result vanishes. This means that

$$\sum_{incoming} Q_i \frac{p_i \cdot q}{p_i \cdot q} - \sum_{outgoing} Q_j \frac{p_j \cdot q}{p_j \cdot q} = 0 \quad (2.32.8)$$

or

$$\sum_{incoming} Q_i = \sum_{outgoing} Q_j \quad (2.32.9)$$

Thus we have derived charge conservation for every scattering process! In other words, a massless spin 1 particle that creates a long-range force must couple to a conserved charge. The book has a nice discussion of form factors following up on this result.

Now let us consider massless spin 2 particles. Lorentz transformations act on the polarization tensor to produce

$$\epsilon_{\mu\nu} \rightarrow \epsilon_{\mu\nu} + v_\mu q_\nu + q_\mu v_\nu + c q_\mu q_\nu \quad (2.32.10)$$

Thus theories involving massless spin 2 particles must satisfy a Ward identity to maintain Lorentz invariance without propagating unphysical polarization states.

In the spin 2 case the soft factor for the  $i$ th line in the limit of small  $q$  is

$$\kappa_i \frac{p_i^\mu p_i^\nu \epsilon_{\mu\nu}}{p_i \cdot q} \quad (2.32.11)$$

where we included a spin 2 coupling  $\kappa_i$  for the  $i$ th particle. Thus we find a soft amplitude

$$M \approx M_0(p_j) \left[ \sum_{incoming} \kappa_i \frac{p_i^\mu p_i^\nu \epsilon_{\mu\nu}}{p_i \cdot q} - \sum_{outgoing} \kappa_i \frac{p_i^\mu p_i^\nu \epsilon_{\mu\nu}}{p_i \cdot q} \right] \quad (2.32.12)$$

Now Lorentz invariance demands that

$$\sum_{incoming} \kappa_i p_i = \sum_{outgoing} \kappa_j p_j \quad (2.32.13)$$

so that  $\kappa_i p_i$  is conserved. But this equation would place an extra algebraic constraint on the allowed momenta unless

$$\kappa_i = \kappa \quad (2.32.14)$$

is universal for all particles. Note that since every particle carries a momentum, this means that *every species of particles, elementary or composite, must couple in the same universal way to massless spin 2 particles.*

In the case of higher spin particles, one finds the constraint

$$\sum_{incoming} \kappa_i p_i^\mu p_i^\nu = \sum_{outgoing} \kappa_j p_j^\mu p_j^\nu \quad (2.32.15)$$

which can never be satisfied by any choice of  $\kappa_i$  for generic momenta. So higher spin particles cannot couple in a way that produces long-range forces, and in fact we have never encountered any massless higher spin particles.

*These arguments go a long way towards explaining the spectrum of particles that we have observed in our universe, and the way that forces act.*

## 3 Spring Semester

### 3.1 Effective Actions, Renormalization, and Symmetry

#### 3.1.1 Connected Generating Functional

We can get connected diagrams immediately, from a generating functional that knows only about them. The connected diagrams are computed from the functional

$$W[J] = -i \log Z[J] \quad (3.1.1)$$

as we saw in the case of generating functions for simple integrals. Let's see how this works at first order in  $\lambda$  for our 2-pt function:

$$\begin{aligned} & -\frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} W[J] \Big|_{J=0} \\ = & -\frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \log \left( \exp \left[ -i \int d^4 y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 Z_0[J(x)] \right] \right) \Big|_{J=0} \\ \approx & -\frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \log \left( Z_0[J(x)] - i \int d^4 y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 Z_0[J(x)] \right) \Big|_{J=0} \\ \approx & -\frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \frac{-i}{Z_0[J(x)]} \int d^4 y \frac{\lambda}{4!} \left( -i \frac{\delta}{\delta J(y)} \right)^4 Z_0[J(x)] \Big|_{J=0} \end{aligned} \quad (3.1.2)$$

Now the derivatives that produce the external fields at  $x_1$  and  $x_2$  can act on either the  $Z_0[J]$  in the numerator, which has a vertex, or on the  $1/Z_0[J(x)]$  factor. Acting only on the first produces all possible diagrams, connected and disconnected, while acting on the latter produces a disconnected diagram, canceling them out.

#### 3.1.2 1-PI Effective (or Quantum Effective) Action

We just saw that we can write down a generating functional for only the connected Feynman diagrams. In fact, we can go further and write down a generating function purely for 1-Particle Irreducible (1-PI) vertices.

This is sometimes called the *quantum effective action* because formally, its propagator is the exact 2-pt function, and its vertices are the exact  $k$ -pt functions of the theory. The exact correlation functions would be made exclusively from tree diagrams with vertices drawn from the 1-PI effective action

$$\Gamma[\Phi(x)] \quad (3.1.3)$$

How do we define it? Well, what if we could turn on the current  $J_\Phi(x)$  (by turn on I mean give it some finite value, not just an infinitesimal value around 0) so that

$$\Phi(x) = \langle \phi(x) \rangle = -i \frac{\delta}{\delta J(x)} W[J] \Big|_{J=J_\Phi} \quad (3.1.4)$$



Now we define the quantum effective action as the Legendre transform

$$\Gamma[\Phi(x)] = W[J_\Phi] - \int d^4x J_\Phi(x) \Phi(x) \quad (3.1.5)$$

Note that taking some functional derivatives and using the definition of  $\Phi$ , we see that

$$\frac{\delta \Gamma}{\delta \Phi(x)} = -J_\Phi(x) \quad (3.1.6)$$

so this means that when  $J_\Phi = 0$ , the possible values of the classical field (expectation value)  $\Phi(x)$  are given by stationary points of  $\Gamma$ . This is the sense in which  $\Gamma$  is really the ‘effective action’. It is the improvement of the classical action, taking quantum corrections into account.

Because  $\Gamma$  is the effective action, it also must sum up all 1-PI contributions to any given vertex. That is, if we look at

$$\frac{\delta}{\delta \Phi(x_1)} \cdots \frac{\delta}{\delta \Phi(x_k)} \Gamma[\Phi(y)] \quad (3.1.7)$$

then this is the exact 1-PI contribution to a  $k$ -point function. The reason is that if we path integrated over  $\Phi$  using  $\Gamma[\Phi]$  as the action, and we only looked at the classical limit, we would get exactly  $W[J_\Phi]$ , now viewing  $J_\Phi \rightarrow J$  as a variable. For more details see Weinberg, chapter 16.

### 3.1.3 Symmetries of the Effective Action

Imagine we have a theory with some symmetry. For example, the simplest possible case would be

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 - \cdots \quad (3.1.8)$$

where there is a symmetry under

$$\phi \rightarrow -\phi \quad (3.1.9)$$

Will this symmetry be preserved by quantum corrections? If not, we have a major problem with renormalization, since once we break the symmetry with quantum corrections, we will probably find that these corrections have short-distance sensitivity, and so we will need new counter-terms in the Lagrangian to absorb that sensitivity. But now we have a Lagrangian that doesn’t have the symmetry at all!

The solution, of course, is that quantum corrections do respect symmetries<sup>4</sup>, so the question is how to see that this is the case. A natural formalism we can use is that of the effective action.

So let us consider some transformation of the fields generated by an infinitesimal transformation

$$\phi_i \rightarrow \phi_i + \epsilon F_i[\phi] \quad (3.1.10)$$

---

<sup>4</sup>Except in very special cases, where there are so-called anomalies.

which leaves both the measure of the path integral and the action invariant. This means that if we act with this transformation on  $Z[J]$ , we simply get

$$Z[J] \rightarrow Z[J] + i\epsilon \int D\phi_i \int d^4y F_i[\phi_i(y)] J^i(y) e^{iS[\phi_i] + \int d^4z \phi_i(z) J^i(z)} \quad (3.1.11)$$

which means that

$$\langle F_i[\phi_i(y)] J^i(y) \rangle_{J_j} = 0 \quad (3.1.12)$$

where this denotes the expectation value with the currents  $J_i$  turned on. But since by definition

$$J_i(y) = -\frac{\delta\Gamma[\phi]}{\delta\phi_i} \quad (3.1.13)$$

we can write this as

$$\left\langle F_i[\phi_i(y)] \frac{\delta\Gamma[\phi]}{\delta\phi_i} \right\rangle_{J_j} = 0 \quad (3.1.14)$$

But this is precisely the statement that the quantum effective action  $\Gamma[\phi]$  is invariant under

$$\phi_i \rightarrow \phi_i + \epsilon \langle F_i[\phi] \rangle_J \quad (3.1.15)$$

Is the quantum expectation value of  $F_i$  the same thing as  $F_i$  itself? It is in the case where

$$F_i[\phi(x)] = s_i(x) + \int d^4y T_i^j(x, y) \phi_j(y) \quad (3.1.16)$$

so that  $F_i$  is *linear*, because we can take expectation values of both sides. We only have problems because the expectation value of e.g. the square of a field isn't the same thing as the square of the expectation value.

## 3.2 Schwinger-Dyson from the PI and Ward Identities

### 3.2.1 Contact Terms and S-D from PI

Let us see how we can obtain the Schwinger-Dyson equations from the PI. The usual classical equations come from a variation

$$\phi(x) \rightarrow \phi(x) + \epsilon(x) \quad (3.2.1)$$

so let us see what happens with that in the PI. If we study the 1-pt function in a free theory, we see

$$\langle \phi(x) \rangle = \frac{1}{Z[0]} \int \mathcal{D}\phi e^{-i \int d^4y \frac{1}{2} \phi \square_y \phi} \phi(x) \quad (3.2.2)$$

Now what happens when we perform the field re-definition inside the path integral? The measure is invariant under such a shift, so we get

$$\begin{aligned}\langle \phi(x) \rangle &= \frac{1}{Z[0]} \int \mathcal{D}\phi e^{-i \int d^4 y \frac{1}{2} (\phi + \epsilon) \square_y (\phi + \epsilon)} (\phi(x) + \epsilon(x)) \\ &\approx \frac{1}{Z[0]} \int \mathcal{D}\phi e^{-i \int d^4 y \frac{1}{2} \phi \square_y \phi} \left[ \phi(x) + \epsilon(x) - i \phi(x) \int d^4 z \epsilon(z) \square_z \phi(z) \right]\end{aligned}\quad (3.2.3)$$

The variation terms must vanish, so we find that

$$\int \mathcal{D}\phi e^{-i \int d^4 y \frac{1}{2} \phi \square_y \phi} [\phi(x) \square_z \phi(z) + i \delta^4(x - z)] = 0 \quad (3.2.4)$$

since  $\epsilon(z)$  can be any function. This is just the Green's function equation for the Feynman propagator. Repeating the derivation for an interacting theory produces the full Schwinger-Dyson equation for the correlators.

One can proceed and write the S-D equations in terms of the generating function, yielding

$$-i \square_x \frac{\delta Z[J]}{\delta J(x)} = \left( \mathcal{L}'_{int} \left[ -i \frac{\delta}{\delta J(x)} \right] + J(x) \right) Z[J] \quad (3.2.5)$$

which the book calls the S-D differential equation. Note that this is formally true *for any*  $J(x)$ , not just for  $J$  near zero, so it gives a lot of non-perturbative information. One immediate way to identify the PI with the ordinary Canonical approach to QFT is to note that both versions yield this same equation.

### 3.2.2 Gauge Invariance

We can prove that when we evaluate *gauge invariant* correlation functions, the result will always be independent of the parameter  $\xi$  in the photon propagator. This is just a first step though, and is not enough to prove  $\xi$  independence of the S-Matrix.

Consider the function

$$f(\xi) = \int \mathcal{D}\pi e^{-i \int d^4 x \frac{1}{2\xi} (\square\pi)^2} \quad (3.2.6)$$

The idea is that by multiplying and dividing by  $f(\xi)/f(\xi) = 1$ , we can obtain the Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 - \frac{1}{2\xi} (\partial_\mu A^\mu)^2 \quad (3.2.7)$$

All we have to do is perform some field redefinitions.

First, let us take  $f(\xi)$  and send

$$\pi \rightarrow \pi - \frac{1}{\square} \partial_\mu A^\mu \quad (3.2.8)$$

where by  $1/\square$  we mean the usual thing. This is just a shift of  $\pi$  so it doesn't change the measure, and so  $f$  doesn't change. We have

$$f(\xi) = \int \mathcal{D}\pi e^{-i \int d^4x \frac{1}{2\xi} (\square\pi - \partial_\mu A^\mu)^2} \quad (3.2.9)$$

which is actually independent of  $A_\mu$ .

Now let us multiply and divide by  $f$  in the path integral for some gauge invariant correlator

$$\begin{aligned} \langle \mathcal{O}_1 \cdots \mathcal{O}_k \rangle &= \frac{1}{f(\xi)} \int \mathcal{D}\pi \mathcal{D}A_\mu \mathcal{D}\phi_i \mathcal{D}\phi_i^* \\ &\times e^{i \int d^4x (\mathcal{L}[A_\mu, \phi_i] - \frac{1}{2\xi} (\square\pi - \partial_\mu A^\mu)^2)} \mathcal{O}_1 \cdots \mathcal{O}_k \end{aligned} \quad (3.2.10)$$

Now let us perform a gauge transformation

$$A_\mu \rightarrow A_\mu + \partial_\mu \pi, \quad \phi_i \rightarrow e^{i\pi} \phi_i \quad (3.2.11)$$

Since this is a gauge transformation it leaves  $\mathcal{L}$  invariant, and because we insisted on computing gauge invariant correlators, it does not affect the  $\mathcal{O}_i$ . Thus we obtain

$$\begin{aligned} \langle \mathcal{O}_1 \cdots \mathcal{O}_k \rangle &= \frac{1}{Z[0]f(\xi)} \int \mathcal{D}\pi \mathcal{D}A_\mu \mathcal{D}\phi_i \mathcal{D}\phi_i^* \\ &\times e^{i \int d^4x (\mathcal{L}[A_\mu, \phi_i] - \frac{1}{2\xi} (\partial_\mu A^\mu)^2)} \mathcal{O}_1 \cdots \mathcal{O}_k \end{aligned} \quad (3.2.12)$$

So we get the gauge fixed Lagrangian. Clearly the normalization drops out when we divide by  $Z[0]$ , so we obtain

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_k \rangle = \frac{\int \mathcal{D}\pi \mathcal{D}A_\mu \mathcal{D}\phi_i \mathcal{D}\phi_i^* e^{i \int d^4x (\mathcal{L}[A_\mu, \phi_i] - \frac{1}{2\xi} (\partial_\mu A^\mu)^2)} \mathcal{O}_1 \cdots \mathcal{O}_k}{\int \mathcal{D}\pi \mathcal{D}A_\mu \mathcal{D}\phi_i \mathcal{D}\phi_i^* e^{i \int d^4x (\mathcal{L}[A_\mu, \phi_i] - \frac{1}{2\xi} (\partial_\mu A^\mu)^2)}} \quad (3.2.13)$$

Thus we can take  $\xi$  to have any value we like. This provides a partial proof of the photon propagator formula.

Unfortunately, this does not apply to correlators of fields that are gauge covariant, but not gauge invariant, because then the  $\mathcal{O}_i$  in the path integral would shift under the gauge transformation we used. But we will prove  $\xi$  independence in general in the next section.

### 3.2.3 Noether's Theorem and Ward Identities

Let us begin by considering what happens to global symmetries and Noether's Theorem in the PI formalism. We saw how linear symmetries remain symmetries of the effective action, let us see what happens to current conservation.

If we have some lagrangian invariant under a global symmetry, we can try performing a field redefinition

$$\phi(x) \rightarrow e^{-i\alpha(x)} \phi(x), \quad \phi^\dagger(x) \rightarrow e^{i\alpha(x)} \phi^\dagger(x) \quad (3.2.14)$$

The measure of the path integral will be invariant, but since  $\alpha(x)$  depends on  $x$  the Lagrangian will not be invariant. The Lagrangian will shift by

$$J_\mu(x)\partial^\mu\alpha(x) \quad (3.2.15)$$

by definition of the current  $J_\mu(x)$ . If we are studying a correlator

$$\langle\phi(x_1)\phi^\dagger(x_2)\rangle \quad (3.2.16)$$

then the fields in the correlator will also transform.

Finally, since  $\phi$  and  $\phi^\dagger$  are just dummy integration variables in the PI, the whole path integral must be invariant. Thus

$$0 = \int \mathcal{D}\phi \mathcal{D}\phi^\dagger e^{iS} \left[ i \int d^4x J_\mu(x) \partial^\mu \alpha(x) \phi(x_1) \phi^\dagger(x_2) + i\alpha(x_2) \phi(x_1) \phi^\dagger(x_2) - i\alpha(x_1) \phi(x_1) \phi^\dagger(x_2) \right]$$

But since this must hold for any  $\alpha(x)$ , it implies

$$\partial_\mu \langle J^\mu(x) \phi(x_1) \phi^\dagger(x_2) \rangle = (\delta^4(x - x_2) - \delta^4(x - x_1)) \langle \phi(x_1) \phi^\dagger(x_2) \rangle \quad (3.2.17)$$

This is much like the Schwinger-Dyson equation – it says that in correlation functions, current conservation holds up to contact terms. The contact terms generate the global symmetry transformations on the fields. Here the signs on the right hand side are just the charge of the fields  $\phi$  and  $\phi^\dagger$ .

To understand this relation in a different way, we can perform a Fourier transform, defining

$$M^\mu(p, q_1, q_2) = \int d^4x d^4x_1 d^4x_2 e^{ipx + iq_1x_1 - iq_2x_2} \langle J^\mu(x) \phi(x_1) \phi^\dagger(x_2) \rangle \quad (3.2.18)$$

and similarly for the other terms. Then the identity becomes

$$ip_\mu M^\mu(p, q_1, q_2) = M_0(q_1 + p, q_2) - M_0(q_1, q_2 + p) \quad (3.2.19)$$

This is often called the Ward-Takahashi identity. It is an exact, non-perturbative statement that implies that charge conservation survives renormalization.

This identity can be written diagrammatically, representing off-shell correlation functions, which can appear inside other Feynman diagrams. It can be generalized easily to correlators with many currents and many charged fields.

### 3.2.4 Conventional Ward Identities

Now we can use the S-D equations to relate the W-T identities to the standard Ward identities.

We can write an S-Matrix element as

$$\begin{aligned} & \langle \epsilon, \dots, \epsilon_k, \dots | S | \dots \rangle \\ &= \epsilon_\mu \epsilon_\alpha^k \left[ i^n \int d^4x e^{ipx} \square_{\mu\nu}^x \dots \int d^4x_k e^{ip_k x_k} \square_{\alpha\beta}^k \dots \right] \langle A_\nu(x) \dots A_\beta(x_k) \dots \rangle \end{aligned} \quad (3.2.20)$$

The term  $\square_{\mu\nu}$  is shorthand for the photon kinetic term, ie

$$\square_{\mu\nu} = g_{\mu\nu}\square - \left(1 - \frac{1}{\xi}\right)\partial_\mu\partial_\nu \quad (3.2.21)$$

in the gauges we are using. Now we apply the Schwinger-Dyson equations to the S-Matrix, giving

$$\begin{aligned} \square_{\mu\nu}^x \square_{\alpha\beta}^k \langle A_\nu(x) \cdots A_\beta(x_k) \cdots \rangle &= \square_{\alpha\beta}^k [\langle J_\mu(x) \cdots A_\beta(x_k) \cdots \rangle - i\delta^4(x - x_k)g_{\mu\beta}\langle \cdots \rangle] \\ &= \langle J_\mu(x) \cdots J_\alpha(x_k) \cdots \rangle - \square_{\mu\alpha}^k \square D_F(x - x_k)\langle \cdots \rangle \end{aligned} \quad (3.2.22)$$

where we wrote the delta function as  $\square D_F(x - x_k)$  to connect with perturbation theory.

Thus we have related the S-Matrix element to a correlator of currents  $J_\mu$  plus a disconnected piece, which drops out of the non-trivial S-Matrix. In other words, S-Matrix elements involving photons are equal to time-ordered products involving currents. This is also true even if we do not force the external momenta to be on-shell.

We also have a proof of the usual Ward identity, that for an S-Matrix element

$$\epsilon_\mu M^\mu(p, q_i) \quad (3.2.23)$$

we must have

$$p_\mu M^\mu(p, q_i) = 0 \quad (3.2.24)$$

This follows because we can apply the above reasoning to write the photon matrix element in terms of current correlators, but those satisfy the Ward-Takahashi identity. When we compute the S-Matrix, the contact terms from W-T must vanish because they are not on-shell for the external states, due to the addition of  $p$  to the various  $q_i$ . Thus the conventional Ward identity is proven.

Finally, note that by writing a scattering amplitude as

$$\mathcal{M} = \epsilon_1^{\alpha_1} \cdots \epsilon_n^{\alpha_n} \int d^4k_1 \prod_{i=1}^m \cdots d^4k_i \Pi_{\mu_i\nu_i}(k_i) \mathcal{M}^{\mu_1\nu_1 \cdots \mu_m\nu_m \alpha_1 \cdots \alpha_n}(k_i, q_i) \quad (3.2.25)$$

By the Ward identity, which does not require propagators to be on-shell, we have that under

$$\Pi_{\mu\nu}(k) \rightarrow \Pi_{\mu\nu}(k) + \alpha k_\mu k_\nu \quad (3.2.26)$$

the correction must vanish. So the scattering amplitudes must be independent of  $\xi$ , as desired. This requires all external charged particles to be on-shell, though, because otherwise there would be contact terms. Technically, the external photons do not need to be on-shell.

### 3.3 Discrete Symmetries and Spinors

Recall that the most useful way to write the Lorentz algebra is using

$$J_i^+ \equiv \frac{1}{2}(J_i + iK_i), \quad J_i^- \equiv \frac{1}{2}(J_i - iK_i) \quad (3.3.1)$$

where  $J_i$  generate rotations and  $K_i$  generate boosts. In this basis the algebra is just

$$[J_i^\pm, J_j^\pm] = i\epsilon_{ijk}J_k^\pm \quad (3.3.2)$$

$$[J_i^+, J_j^-] = 0 \quad (3.3.3)$$

So we just have two copies of  $so(3) = sl(2, R) = su(2)$ . This can be written as

$$so(1, 3) = su(2) \oplus su(2) \quad (3.3.4)$$

The Pauli matrices are

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.3.5)$$

We have  $\sigma^\mu = (1, \vec{\sigma})$ , and  $\bar{\sigma}^\mu = (1, -\vec{\sigma})$ . In the Weyl basis, the gamma matrices are

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \text{i.e.} \quad \gamma^0 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad (3.3.6)$$

This means that the gamma matrices are all real, except for  $\gamma^2$ , which is pure imaginary.

### 3.3.1 Chirality, Helicity, and Spin

We have studied both Weyl and Dirac spinors. Recall that the Dirac spinor is simply a 4-component object that can be written in terms of two Weyl spinors, a  $\psi_L$  and a  $\psi_R$ , via

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad (3.3.7)$$

The handedness of a spinor is also called its *chirality*, and corresponds to whether it is a  $(\frac{1}{2}, 0)$  vs  $(0, \frac{1}{2})$ , or left-handed vs right-handed, in terms of its transformation properties under the Lorentz group. In general, we refer to a theory as *chiral* if it chooses a handedness by including an  $(A, B)$  representation of the Lorentz group without a  $(B, A)$  representation. QED is not chiral, but the full standard model is chiral (the weak interactions are chiral).

If we're living in an even dimension, we can add one more dimension by considering

$$\gamma^{2n+1} = i\epsilon_{a_1 \dots a_{2n}} \gamma^{a_1} \dots \gamma^{a_{2n}} \quad (3.3.8)$$

For example, we see that

$$\sigma^3 = i\sigma^1\sigma^2 \quad (3.3.9)$$

allows us to go from 2 dimension to 3 dimensions. This automatically works – one can check that  $\sigma^3$  is forced to satisfy the relations of the Clifford algebra. Similarly, note that

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 \quad (3.3.10)$$

is the 4-by-4 matrix

$$\gamma^5 = \begin{pmatrix} -\mathbf{1} & 0 \\ 0 & \mathbf{1} \end{pmatrix} \quad (3.3.11)$$

in the Weyl representation (recall that the  $\gamma$  matrices can be more formally defined in terms of their Clifford algebra, and that the Weyl representation is one explicit matrix representation of that more general algebra).

The reason we call it  $\gamma^5$  is because it satisfies

$$(\gamma^5)^2 = 1, \quad \{\gamma^5, \gamma^\mu\} = 0 \quad (3.3.12)$$

and so it extends the Clifford algebra of the  $\gamma^\mu$  to a 5th dimension. If we wanted to discuss spinors in 4 + 1 spacetime dimensions we would literally use  $\gamma^\mu$  and  $\gamma^5$ , and there would be no Weyl 2-spinor representation, but only the 4-spinor representation with  $\gamma$ .

Now we can define left-handed and right-handed projectors

$$P_R = \frac{1 + \gamma^5}{2}, \quad P_L = \frac{1 - \gamma^5}{2}, \quad (3.3.13)$$

that project onto Dirac spinors of definite chirality. It's no accident that we used the anti-symmetric  $\epsilon$  tensor (or Levi-Civita tensor) when we defined  $\gamma^{2n+1}$ , or in this case,  $\gamma^5$ . The  $\epsilon$  tensor changes sign under parity transformations, where space  $\rightarrow$   $-\text{space}$ , and this is also why  $\gamma^5$  picks out states of definite chirality, as we will see.

Recall that the Dirac equation in Fourier space is

$$\sigma^\mu p_\mu \psi_R = m \psi_L \quad (3.3.14)$$

$$\bar{\sigma}^\mu p_\mu \psi_L = m \psi_R \quad (3.3.15)$$

We discussed the possibility of Majorana spinors last semester, where we use the fact that

$$\sigma_2 \psi_R^* \quad (3.3.16)$$

transforms as a left-handed spinor in order to write a mass term with only  $\psi_R$ . But if we don't do that, then the mass term mixes left and right handed spinors.

When we study massless spinors, the left and right handedness never changes. In other words, the *helicity* operator

$$\hat{h} = \frac{\vec{\sigma} \cdot \vec{p}}{|\vec{p}|} \quad (3.3.17)$$

commutes with the Hamiltonian, and can be diagonalized. The *helicity is the spin projected on the direction of motion*. Another way of understanding why it cannot be preserved in the massive case is that massive spinors can be brought to rest, in which case the helicity clearly cannot be defined, since  $\vec{p} = 0$ ! But for massless spinors, or spinor particles that are ultra-relativistic, helicity makes is a good quantum number.

We should contrast chirality and helicity with good old *spin*, which is just the representation under spatial rotations, and is identical with the notion of spin familiar from non-relativistic mechanics.



### 3.3.2 Charge Conjugation

The operation that acts on Dirac spinors as

$$C : \quad \psi \rightarrow \psi_c \equiv -i\gamma_2\psi^* \quad (3.3.18)$$

is called charge conjugation. We can immediately see that  $C^2 = 1$  because

$$C^2 : \quad \psi \rightarrow -i\gamma_2\psi^* \rightarrow -i\gamma_2(-i\gamma_2\psi^*)^* = \psi \quad (3.3.19)$$

since  $\gamma_2^* = -\gamma_2$  and  $\gamma_2^2 = -1$ . This is why we call it ‘conjugation’.

Note that when we write Majorana fermions using Dirac spinors (and not just a single Weyl spinor  $\psi_L$ ) we have

$$\psi = \begin{pmatrix} \psi_L \\ i\sigma_2\psi_L^* \end{pmatrix} \quad (3.3.20)$$

When we apply charge conjugation to the Majorana fermion, we find

$$-i\gamma_2\psi^* = \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \begin{pmatrix} \psi_L \\ i\sigma_2\psi_L^* \end{pmatrix}^* = \psi \quad (3.3.21)$$

and so Majorana fermions are their own charge conjugate. This is as we have claimed in the past – Majorana fermions cannot carry a conserved  $U(1)$  charge, ie they cannot transform as  $\psi \rightarrow e^{i\alpha}\psi$  because it would violate  $\psi = \psi_c$  (although they can carry charges under a real representation of a non-abelian group).

For a better understanding of why we call this operation charge conjugation, take the complex conjugate of the Dirac equation

$$(i\gamma_\mu\partial^\mu - e\gamma_\mu A^\mu - m)\psi = 0 \quad (3.3.22)$$

to give

$$(-i\gamma_\mu^*\partial^\mu - e\gamma_\mu^*A^\mu - m)\psi^* = 0 \quad (3.3.23)$$

This implies that

$$\gamma_2(-i\gamma_\mu^*\partial^\mu - e\gamma_\mu^*A^\mu - m)\gamma_2\psi_c = 0 \quad (3.3.24)$$

Now note that by our choice of basis for the  $\gamma$  matrices (Weyl basis),  $\gamma_2$  is imaginary while the others are real, so if we define new  $\gamma$  matrices via

$$\gamma'_\mu = \gamma_2\gamma_\mu^*\gamma_2 \quad (3.3.25)$$

then these new  $\gamma'$  satisfy the Dirac algebra. This means that

$$(i\gamma'_\mu\partial^\mu + e\gamma'_\mu A^\mu - m)\psi_c = 0 \quad (3.3.26)$$

so that  $\psi_c$  satisfies the Dirac equation using the new basis of  $\gamma'$  matrices. But all bases are equivalent, so by comparing with the Dirac equation

$$(i\gamma_\mu\partial^\mu - e\gamma_\mu A^\mu - m)\psi = 0 \quad (3.3.27)$$

we see that  $\psi_c$  has opposite charge, ie  $e \rightarrow -e$ .

It is also worthwhile to see how  $C$  acts on solutions to the Dirac equation. We can write the four spinor solutions as

$$u_+ = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_- = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad v_+ = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \quad v_- = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \quad (3.3.28)$$

in which case we find that

$$(u_+)^c = v_-, \quad (u_-)^c = v_+ \quad (3.3.29)$$

with the other possibilities determined by  $C^2 = 1$ . This means that *charge conjugation flips the spin and takes particles to antiparticles*. So in the case of massless uncharged particles, the charge conjugate has its helicity flipped.

A given theory, or a given set of interactions, may or may not respect charge conjugation. It does not have to be a symmetry. Let us see how it acts on various Lagrangian terms.

For example

$$C : \bar{\psi}\psi \rightarrow (-i\gamma_2\psi)^T\gamma_0(-i\gamma_2\psi^*) = -\psi^T\gamma_0\psi^* = \bar{\psi}\psi \quad (3.3.30)$$

where in the last step we assumed that the spinors anticommute. Similarly

$$C : \bar{\psi}\not{\partial}\psi \rightarrow \bar{\psi}\not{\partial}\psi \quad (3.3.31)$$

so the free massive Lagrangian for a Dirac spinor is  $C$  invariant, as one might expect. However

$$C : \bar{\psi}\gamma^\mu\psi \rightarrow -\bar{\psi}\gamma^\mu\psi \quad (3.3.32)$$

as one would expect for an electric current (it better flip sign under charge conjugation). Note that it differs from the  $\not{\partial}$  case due to the necessity of an integration by parts. This means that the QED Lagrangian can only be  $C$  invariant if

$$C : A_\mu \rightarrow -A_\mu \quad (3.3.33)$$

Note that we could have ignored this sign, in which case we would have obtained a ‘ $C$ ’ transformation that isn’t a symmetry. The lesson is that we should *always try* to figure out if a  $C$  transformation can be defined such that the Lagrangian is invariant. The existence of ‘bad’ choices of  $C$  that are not symmetries is not good enough to show that  $C$  isn’t a discrete symmetry of our theory.

### 3.3.3 Parity

In addition to the continuous rotations and boosts in the Lorentz group, we can also consider the operations

$$P : (t, \vec{x}) \rightarrow (t, -\vec{x}) \quad (3.3.34)$$

$$T : (t, \vec{x}) \rightarrow (-t, \vec{x}) \quad (3.3.35)$$

corresponding to parity, or mirror reflection, and time reversal. There is no guarantee that these are good symmetries of some given theory. For example, QED is  $P$  and  $T$  symmetric, but the weak interactions are not. We always want to *define* the action of these generators on fields so that  $P^2 = T^2 = 1$ .

How can parity act? Presumably a real scalar field will have an invariant action, which means that

$$P : \phi(t, \vec{x}) \rightarrow \pm \phi(t, \pm \vec{x}) \quad (3.3.36)$$

The sign is known as the *intrinsic parity* of  $\phi$ . There are particles with even intrinsic parity, such as the Higgs boson, and particles with odd intrinsic parity (so-called pseudo-scalars) such as the  $\pi^0$  meson. Again, the point is that we identify parity by *trying* to make it a good symmetry.

For complex scalars parity can act to preserve the free Lagrangian as long as

$$P : \phi(t, \vec{x}) \rightarrow \eta \phi(t, \pm \vec{x}) \quad (3.3.37)$$

where  $\eta$  can be any complex phase. If  $\phi$  has a global symmetry under  $\phi \rightarrow e^{i\alpha} \phi$  then we can combine  $P$  with this global symmetry, meaning that  $\eta$  is not well-defined. But if many complex scalars have the same charge then once we fix  $\eta$  for one of them, the  $\eta$  for the others becomes meaningful.

In fact we can guarantee that  $\eta = \pm 1$  for all particles, *assuming* we have some continuous  $U(1)$  symmetry. If we have  $\eta^2 = e^{i\alpha Q}$  for some  $\alpha$  for a particle of charge  $Q$ , we can define a new

$$P' = P e^{-\frac{i}{2}\alpha Q} \quad (3.3.38)$$

so that  $(P')^2 : \psi \rightarrow \psi$ . We may as well call this operation parity, so that for all fields  $P : \psi \rightarrow \pm \psi$ . In the Standard Model it is conventional to use lepton number, baryon number, and electromagnetic charge to set the parity of the proton, neutron, and electron to  $+1$ .

For vector fields  $P$  should act much as it does on 4-vectors. For the free vector theory to be invariant, we only need

$$P : V_0(t, \vec{x}) \rightarrow \pm V_0(t, -\vec{x}), \quad V_i(t, \vec{x}) \rightarrow \mp V_i(t, -\vec{x}) \quad (3.3.39)$$

By convention when  $P : V_i \rightarrow -V_i$  we say that  $V_\mu$  is a vector with parity  $-1$ , but when  $V_i \rightarrow V_i$  we say that it is a pseudovector with parity  $+1$ . The electric field is a vector while the magnetic field is a pseudovector.

The photon has parity  $-1$ , because it must transform in the same way as  $\partial_\mu$  in order to have parity preserving couplings to electromagnetic currents. So it's a vector.

Now let us discuss spinors. Since  $P$  commutes with rotations, it should not change the spin of a state. Massless spinors can be taken to be eigenstates of the helicity operator

$$\frac{\vec{\sigma} \cdot \vec{p}}{|\vec{p}|} \psi_R = \psi_R, \quad \frac{\vec{\sigma} \cdot \vec{p}}{|\vec{p}|} \psi_L = -\psi_L \quad (3.3.40)$$

Since parity commutes with the spin but flips the momentum, it must exchange left and right-handed spinors, mapping  $(A, B)$  representations of the Lorentz group to  $(B, A)$  representations. For Dirac spinors in the Weyl basis, this swapping of left and right can be written

$$P : \psi \rightarrow \gamma_0 \psi \quad (3.3.41)$$

In principle there could be a phase ambiguity, but we can eliminate it using a  $U(1)$  charge rotation, if such a  $U(1)$  exists.

We see that

$$P : \bar{\psi} \gamma_0 \psi(t, \vec{x}) \rightarrow \bar{\psi} \gamma_0 \psi(t, -\vec{x}), \quad \bar{\psi} \gamma_i \psi(t, \vec{x}) \rightarrow -\bar{\psi} \gamma_i \psi(t, -\vec{x}) \quad (3.3.42)$$

so these bilinears transform as vectors, not pseudovectors. In contrast, the bilinear  $\bar{\psi} \gamma_\mu \gamma^5 \psi$  transforms as a pseudovector. The corresponding currents are called vector currents and axial vector currents for this reason. This also shows the relation between  $\gamma^5$  and parity, as claimed above.

### 3.3.4 Time Reversal

Time reversal appears confusing at first sight. The problem is that we need to make  $i\bar{\psi}\not{\partial}\psi$  invariant, but this means that  $T$  since  $T : \partial_t \rightarrow -\partial_t$ , we must take  $\bar{\psi}\gamma^0\psi = \psi^\dagger\psi$  to  $-\psi^\dagger\psi$ , or in other words, it must negate a positive definite quantity. But this is impossible for a linear transformation. One workable solution effectively defines  $T \sim (CP)^{-1}$ , but this is somewhat trivial.

A better resolution is to make  $T$  a so-called anti-linear transformation, so that

$$T : i \rightarrow -i \quad (3.3.43)$$

This means that  $T(a + ib) = T(a) - iT(b)$  for real  $a$  and  $b$ . This fixes our problem above because of the  $i$  in the fermion kinetic terms. Note that this also means that  $T$  acts on the  $\gamma$  matrices; in the Weyl basis

$$T : \gamma_{0,1,3} \rightarrow \gamma_{0,1,3}, \quad \gamma_2 \rightarrow -\gamma_2 \quad (3.3.44)$$

since only  $\gamma_2$  is imaginary. By defining

$$T : \psi(t, \vec{x}) \rightarrow \tilde{\Gamma} \psi(-t, \vec{x}) \quad (3.3.45)$$

demanding invariance of the kinetic term, and following our nose, we find that  $\tilde{\Gamma} = \gamma_1 \gamma_3$  up to a constant phase. This means that  $T$  *flips the spin of particles, but does not turn particles into antiparticles*.  $T$  does not have a well-defined action on Weyl spinors, because they have only one spin state.

Note that  $T$  reverses the momentum, because  $\vec{p} = i\vec{\nabla}$ , and  $T$  flips the sign of  $i$ . So  $T$  makes it look like things are going forward in time, but with their momenta and spins flipped.

Similarly, for the QED Lagrangian to be invariant, we must have

$$T : A_0(t, \vec{x}) \rightarrow A_0(-t, \vec{x}), \quad A_i(t, \vec{x}) \rightarrow -A_i(-t, \vec{x}) \quad (3.3.46)$$

so that  $A_\mu$  transforms like  $i\partial_\mu$ , which may not be what one would guess.

Note that  $CPT$  acts as

$$CPT : \psi(x) \rightarrow -\gamma_5 \psi^*(-x) \quad (3.3.47)$$

sending particles into antiparticles moving in reverse, in a mirror. One can check that all terms you can write down are CPT symmetric, which is a version of the CPT theorem.

You can find all sorts of examples of the use of C, P, and T in particle physics in Weinberg's book, chapter 3.3.

### 3.4 More on Spin and Statistics

We already saw, albeit briefly, that one can obtain the spin statistics relation that integer spin particles are bosons, while half-integer spin particles are fermions, by demanding *stability* – that the energy is bounded from below. Let us now see how Spin-Statistics follows from Lorentz invariance of the S-Matrix, or causality. First we will lay the groundwork and discuss some more intuitive arguments.

Particles can be classified as belonging to species, and within a given species they are indistinguishable. This is true of both fundamental and composite particles. If particles were distinguishable, physics would be very different. For example, we would have the Gibbs paradox, and entropy would not be an extensive quantity. The ‘statistics’ of the spin-statistics relation tells us how the wavefunction changes when we exchange particles. It's very easy to get confused about this subject because the notation we use has built in assumptions.

We have been using creation and annihilation operators to create multi-particle states. So for example we have

$$|p_1, n_1; p_2, n_2\rangle = a_{p_1, n_1}^\dagger a_{p_2, n_2}^\dagger |0\rangle \quad (3.4.1)$$

If the particles are identical, then acting in the opposite order produces the same physical state, so

$$|p_1, n; p_2, n\rangle = e^{i\phi} |p_2, n; p_1, n\rangle \quad (3.4.2)$$

for some phase  $\phi$ . If we swap the particles back then we get the same state (or do we? what does it mean to ‘swap them back’?) and so we find

$$|p_1, n; p_2, n\rangle = e^{2i\phi} |p_1, n; p_2, n\rangle \quad (3.4.3)$$

and so the phase must have been  $\pm 1$ . This implies that either

$$[a_{p, n}, a_{q, n}] = 0 \quad \text{or} \quad \{a_{p, n}, a_{q, n}\} = 0 \quad (3.4.4)$$

for bosons or fermions, respectively. The latter relation implies the Pauli exclusion principle, since the wavefunction vanishes if we create two identical fermions. We already saw from stability that half-integer spin particles must be fermions, and integer spin particles must be bosons.

### 3.4.1 A More Careful Look at Configuration Space

You have probably heard that this isn't the whole story, and that in  $2 + 1$  spacetime dimensions we can have 'anyons', or particles where the phase  $\phi$  can take any real value. But the argument above seems like it was independent of spacetime dimension, so what's going on? These issues were first carefully examined (as far as I know) by Leinaas and Myrheim, who discovered the idea of anyons in 1972. As often happens, the solution was to think very physically, in this case about what it means to 'swap' identical particles. As we discussed it above, 'swapping' particles has no physical meaning.

If particles are identical, then the wavefunction

$$\psi(t, x_1, x_2, \dots, x_n) \quad (3.4.5)$$

for an  $n$ -particle state is not a function of each of the  $x_i$  separately. More formally, it is not simply a function on the space  $(R^3)^n$  in 3 spatial dimensions. Rather it is a function on

$$(R^{d-1})^n / S_n \quad (3.4.6)$$

where  $d - 1$  is the spatial dimension, and the symmetric group  $S_n$  acts by permuting the  $n$  different  $x_i$  in an arbitrary way.

In fact, since wavefunctions are only defined up to a phase, *the wavefunction can be a function that lives on the covering space of the configuration space, as long as it only differs by a phase at equivalent points in configuration space.*

Let us specialize to the case of 2 particles. This means that we identify

$$(x_1, x_2) \sim (x_2, x_1) \quad (3.4.7)$$

Locally the configuration space is what we would expect, but in terms of global configurations, it is quite different. The center of mass coordinate is  $S_n$  invariant, so we can simply discuss the relative coordinate  $y = x_1 - x_2$ , in which case the identification is just

$$y \sim -y \quad (3.4.8)$$

This means that  $(x, x)$  or  $y = 0$  is a singular point in the configuration space, where the geometry degenerates. Either the wavefunction can have support there or not. If it can have support there, then the configuration space is simply connected, and our particles must be bosons (as we will see below). If not, then we can eliminate the point from configuration space and proceed.

If we eliminate this singular point from configuration space then it has the topology

$$(0, \infty) \times P^{d-2} \quad (3.4.9)$$

where the first factor is  $|y|$ , and the second is the real projective space of dimension  $d - 2$ . Basically this is just a circle for  $d = 3$ , and a sphere modulo the  $Z_2$  identification  $y \sim -y$  for  $d = 4$ .

The wavefunction must always be a continuous function on the configuration space, but when we go around a circle, it need only return to itself up to a phase. Thus we see that  $d \geq 4$  is very different from  $d = 3$ , because the latter is not simply connected. In  $2 + 1$  spacetime dimensions

the wavefunction need not be single-valued, and instead it can pick up an invariant phase when  $y$  encircles the singular point. If the phase isn't  $\pm 1$  then we have a pair of anyons. States with more than two particles can have an even more complicated behavior, called 'nonabelian anyons'.

But now let's just go back to  $d \geq 4$ . In that case traversing configuration space twice gives us a path that is topologically equivalent to no winding at all. This is what tells us that particles must be either bosons or fermions – the phase must be  $\pm 1$  to have a continuous wavefunction.

We can use this setup to give a rough derivation of spin-statistics. Taking account of the spin of the particles, we note that when we parallel transport in configuration space using the rotation operator of the theory,  $J_z$ , we act with

$$\begin{pmatrix} e^{\frac{i}{2}\theta_z} & 0 \\ 0 & e^{-\frac{i}{2}\theta_z} \end{pmatrix} \quad (3.4.10)$$

on e.g. Weyl spinors. This means that when we rotate by  $\theta_z = \pi$ , we pick up a minus sign for the 2-particle state. This strongly suggests that spin 1/2 particles (and the same derivation works for any half-integer spin) must be fermions, while integer spin particles should be bosons. Pairs of particles with any spin  $s$  pick up a phase  $e^{is\theta_z}$ .

One might not be entirely convinced by this derivation. Do we really need to use the rotation generator  $J_z$  to move in configuration space (or more precisely, to study the properties of the wavefunction as a function on that space)? We see that this is connected with Lorentz invariance – if we were willing to break Lorentz symmetry, we might just move the particles around without acting with  $J_z$ .

### 3.4.2 Weinbergian Philosophy, Lorentz Invariance, and Causality

Quantum Field Theory sure is a bore. Remind me why we bother with all these quantum fields in the first place?

Something that we saw a long time ago is that the S-Matrix can be written in terms of the  $S$  operator, which itself is

$$S = 1 + \sum_{n=1}^{\infty} \frac{(i)^n}{n!} \int d^4x_1 \cdots d^4x_n T \{ \mathcal{L}_I(x_1) \cdots \mathcal{L}_I(x_n) \} \quad (3.4.11)$$

where  $\mathcal{L}_I$  is the interaction part of the Lagrangian, or minus the interaction Hamiltonian density.

This is almost, but not quite, Lorentz invariant. The problem is the time ordering. When  $x_i - x_j$  is time-like, the time ordering is, in fact, Lorentz invariant. But when  $x_i - x_j$  is spacelike time ordering is most definitely not Lorentz invariant, and instead is frame dependent. We made use of this 'problem' at the beginning of the first semester in order to argue that antiparticles have to exist, that classical particle sources must also absorb particles (and vice versa), and that objects that scatter particles must also produce pairs of them.

Now let us make a more formal comment. Pretty much the only way to make  $S$  Lorentz invariant

is to insist that<sup>5</sup>

$$[\mathcal{L}_I(x), \mathcal{L}_I(y)] = 0, \quad (x - y)^2 \leq 0 \quad (3.4.12)$$

or when  $x - y$  are spacelike or light-like separated. One can give a formal proof of the Lorentz invariance of the S-Matrix (and not just the  $S$  operator) using this statement; see section 3.3 and 3.5 of Weinberg's QTF book.

So now we ask how we can construct an interaction satisfying equation (3.4.12). The answer is... we introduce quantum fields. They are nothing new, but the point that we are now emphasizing is that the reason we use them (instead of, say, the creation and annihilation operators all by themselves) is that it makes it very easy to construct  $\mathcal{L}_I$  as a polynomial in the fields. This is, of course, what the Lagrangian formalism does for us, when it is expressed in terms of quantum fields. But this method only works if the quantum fields themselves satisfy simple commutation or anti-commutation relations at spacelike separations. Thus the *Lorentz invariance of the S-Matrix can be used to prove the spin statistics theorem*.

The relation of equation (3.4.12) also has a more basic interpretation when we replace  $\mathcal{L}_I \rightarrow \mathcal{O}$ , some operator observable. As you know from linear algebra, two matrices can only be simultaneously diagonalized if they commute. And spacelike separated observables can always be viewed as occurring 'at the same time'. Thus if we want it to be possible to simultaneously diagonalize observables at the same time, but in different spatial positions, then they must commute.

### 3.4.3 Lorentz Invariance of the S-Matrix

We can definitively establish the spin-statistics relation by demanding a Lorentz invariant S-Matrix. We will do a very simple calculation and study time-ordered correlators.

But first, note that to even make sense of time ordering for fermionic fields, we must take

$$T\{\psi(x)\chi(y)\} \equiv \psi(x)\chi(y)\theta(x_0 - y_0) - \chi(y)\psi(x)\theta(y_0 - x_0) \quad (3.4.13)$$

This follows because

$$T\{\psi(x)\chi(y)\} = -T\{\chi(y)\psi(x)\} \quad (3.4.14)$$

by fermi statistics, and so this relation would vanish if we did not include the sign in its definition.

Recall that we obtained a sensible, Lorentz invariant Feynman propagator for scalar fields using commutation relations, and therefore time ordering without a minus sign. If instead we had used anti-commutation relations, we would have found

$$\begin{aligned} \langle T\{\phi(x)\phi(0)\} \rangle &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_p} e^{-i\vec{p}\cdot\vec{x}} [-e^{iE_p t} \theta(-t) + e^{-iE_p t} \theta(t)] \\ &= - \int \frac{d^4p}{(2\pi)^4} \frac{i}{2E_p} e^{-ip\cdot x} \left( \frac{1}{E - (E_p - i\epsilon)} - \frac{1}{E - (-E_p + i\epsilon)} \right) \\ &= \int \frac{d^4p}{(2\pi)^4} \frac{E}{\sqrt{p^2 + m^2}} \frac{-i}{p^2 - m^2 + i\epsilon} e^{ip\cdot x} \end{aligned} \quad (3.4.15)$$

---

<sup>5</sup>It's worth pointing out that this relation isn't very analytic, since any analytic function that vanishes in a region vanishes everywhere.



which is not Lorentz invariant. Thus the S-Matrix for spin 0 particles will only be Lorentz invariant if we assume that they are bosons. Now let us move on to spin 1/2 particles.

Recall that we quantized a Dirac spinor field as

$$\psi(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^s u_p^s e^{-ipx} + b_p^{s\dagger} v_p^s e^{ipx}) \quad (3.4.16)$$

and

$$\bar{\psi}(x) = \sum_s \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} (a_p^{s\dagger} \bar{u}_p^s e^{ipx} + b_p^s \bar{v}_p^s e^{-ipx}) \quad (3.4.17)$$

In this language  $\psi(x)$  annihilates incoming electrons and  $\bar{\psi}(x)$  annihilates incoming positrons. Now let us compute

$$\begin{aligned} \langle \psi(0) \bar{\psi}(x) \rangle &= \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \frac{1}{\sqrt{2E_q}} \\ &\quad \times \sum_{s,s'} \langle 0 | (a_p^s u_p^s + b_p^{s\dagger} v_p^s) (a_q^{s'\dagger} \bar{u}_q^{s'} e^{iq \cdot x} + b_q^{s'} \bar{v}_q^{s'} e^{-iq \cdot x}) | 0 \rangle \\ &= \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \frac{1}{\sqrt{2E_q}} \sum_{s,s'} u_p^s \bar{u}_q^{s'} \langle 0 | a_p^s a_q^{s'\dagger} | 0 \rangle e^{iqx} \\ &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_p} \sum_s u_p^s \bar{u}_p^s e^{ipx} \end{aligned} \quad (3.4.18)$$

Note that this is a matrix in spinor space. To sum over the polarizations we use

$$\sum_{s=1}^2 u_s(p) \bar{u}_s(p) = \not{p} + m, \quad \sum_{s=1}^2 v_s(p) \bar{v}_s(p) = \not{p} - m \quad (3.4.19)$$

which gives the final result

$$\langle \psi(0) \bar{\psi}(x) \rangle = (-i\not{\partial} + m) \int \frac{d^3p}{(2\pi)^3} \frac{e^{ipx}}{2E_p} \quad (3.4.20)$$

Similarly we find that

$$\langle \bar{\psi}(x) \psi(0) \rangle = -(-i\not{\partial} + m) \int \frac{d^3p}{(2\pi)^3} \frac{e^{-ipx}}{2E_p} \quad (3.4.21)$$

These equations are independent of whether we choose commutators or anti-commutators, since we didn't move any of the creation or annihilation operators around in their derivation.

Now, to play Devil's advocate, let us first assume that we have commutation relations, so we define the time ordered product via a sum. Then we get

$$T\{\psi(0) \bar{\psi}(x)\}_{comm} = \psi(0) \bar{\psi}(x) \theta(-t) + \bar{\psi}(x) \psi(0) \theta(t) \quad (3.4.22)$$

and so

$$\begin{aligned}
\langle T\{\psi(0)\bar{\psi}(x)\}\rangle_{comm} &= -(-i\not{\partial} - m) \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_p} (e^{ipx}\theta(t) - e^{-ipx}\theta(-t)) \\
&= (-i\not{\partial} + m) \int \frac{d^4p}{(2\pi)^4} \frac{p_0}{\sqrt{p^2 + m^2}} \frac{i}{p^2 - m^2 + i\epsilon} e^{ipx}
\end{aligned} \tag{3.4.23}$$

which is not Lorentz invariant! If instead we assume anti-commutation, so that

$$T\{\psi(0)\bar{\psi}(x)\}_{anti-comm} = \psi(0)\bar{\psi}(x)\theta(-t) - \bar{\psi}(x)\psi(0)\theta(t) \tag{3.4.24}$$

then we get the correct result

$$T\{\psi(0)\bar{\psi}(x)\}_{anti-comm} = (-i\not{\partial} + m) \int \frac{d^4p}{(2\pi)^4} \frac{i}{p^2 - m^2 + i\epsilon} e^{ipx} \tag{3.4.25}$$

which is Lorentz invariant. This is, of course, the correct Dirac propagator. So we need to assume that spin 1/2 particles are fermions.

Let's go back and see why this happened. In essence, for scalars we have that

$$\begin{aligned}
\langle \phi^\dagger(x)\phi(0) \rangle &= \int \frac{d^3p}{(2\pi)^3} \frac{e^{-ipx}}{2E_p} \\
\langle \phi(0)\phi^\dagger(x) \rangle &= \int \frac{d^3p}{(2\pi)^3} \frac{e^{ipx}}{2E_p}
\end{aligned} \tag{3.4.26}$$

while for massless fermions, we have

$$\begin{aligned}
\langle \psi(0)\bar{\psi}(x) \rangle &= \int \frac{d^3p}{(2\pi)^3} \frac{\not{p}}{2E_p} e^{ipx} \\
\langle \bar{\psi}(x)\psi(0) \rangle &= \int \frac{d^3p}{(2\pi)^3} \frac{\not{p}}{2E_p} e^{-ipx}
\end{aligned} \tag{3.4.27}$$

Thus the difference is that  $\not{p}$  is odd under the rotation that takes  $p \rightarrow -p$ , and so we generate an extra  $-1$  when we combine the fermions to make a time-ordered sum. This always happens for half-integer spins, because we always have a  $\not{p}$ , perhaps accompanied by various powers of  $p^2$  or  $p_\mu p_\nu$ .

### 3.5 QED Vacuum Polarization and Anomalous Magnetic Moment

Let's start to consider loop effects in QED. This isn't very different at all from what we did last semester, but some differences are

- You've heard of QED, so the effects we'll discover have more meaning to you. In particular, we'll quickly uncover the change of the electromagnetic force with distance at very short distances.

- Gauge invariance (masslessness of the photon and Lorentz invariance) puts major restrictions on the loop effects. In particular, you might naively think that the photon propagator and the vertex for photons to interact with currents can get renormalized separately, but they are linked by gauge invariance. Also, the photon cannot get a mass, because that badly violates gauge invariance!
- The electron is not a scalar, and in particular, it has a spin, and thus a magnetic moment. We can compute quantum corrections to that magnetic moment, and obtain a famous effect.

Let us first consider what corrections we can have to the photon 2-pt function. By Lorentz invariance, it must take the form

$$\Pi^{\mu\nu} = \Delta_1(p^2)p^2g^{\mu\nu} + \Delta_2(p^2)p^\mu p^\nu \quad (3.5.1)$$

We can therefore write a corrected 2-pt function as

$$iG^{\mu\nu} = -i \frac{(1 + \Delta_1)g^{\mu\nu} + \Delta_2 \frac{p^\mu p^\nu}{p^2}}{p^2 + i\epsilon} \quad (3.5.2)$$

where we have set  $\xi = 1$ , and defined

$$\langle \Omega | T \{ A^\mu(x) A^\nu(0) \} | \Omega \rangle = \int \frac{d^4p}{(2\pi)^4} e^{ipx} iG^{\mu\nu}(p) \quad (3.5.3)$$

as the dressed propagator. Note that since  $\Delta_2$  is just proportional to  $p^\mu p^\nu$ , by gauge invariance it will drop out of our physical predictions, so we can just ignore it.

In the spinor theory, there is a single diagram renormalizing the photon propagator. It is

$$i\Pi_2^{\mu\nu} = -(-ie)^2 \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2} \frac{i}{(p-k)^2 - m^2} \times \text{Tr} [\gamma^\mu (\not{k} - \not{p} + m) \gamma^\nu (\not{k} + m)] \quad (3.5.4)$$

where the overall minus sign comes from the fact that this is a fermion loop. This follows because we need to compute a contraction

$$T \{ A_\mu(x_1) A^\nu(x_2) \bar{\psi}(x) A^\alpha(x) \psi(x) \bar{\psi}(y) A^\beta(y) \psi(y) \} \quad (3.5.5)$$

and there is a sign flip when passing spinors through other spinors to get canonical  $T\{\psi\bar{\psi}\}$  Feynman propagators. We have a trace because the fermions go back to themselves.

Using trace formulas (you can read about them in the book), we find that

$$\text{Tr} [\gamma^\mu (\not{k} - \not{p} + m) \gamma^\nu (\not{k} + m)] = 4[-p^\mu k^\nu - k^\mu p^\nu + 2k^\mu k^\nu + g^{\mu\nu}(-k^2 + p \cdot k + m^2)] \quad (3.5.6)$$

The terms with a  $p^\mu$  will make irrelevant contributions, so we can simplify our task by dropping them. We get

$$i\Pi_2^{\mu\nu} = -4e^2 \int \frac{d^4k}{(2\pi)^4} \frac{2k^\mu k^\nu + g^{\mu\nu}(-k^2 + p \cdot k + m^2)}{[k^2 - m^2][(p-k)^2 - m^2]} \quad (3.5.7)$$

Now we employ the standard trick of Feynman parameterization to combine denominators, and we redefine the integration variable to

$$k \rightarrow k + p(1 - x) \quad (3.5.8)$$

This gives

$$\Pi_2^{\mu\nu} = 4ie^2 \int \frac{d^4k}{(2\pi)^4} \int_0^1 dx \frac{2k^\mu k^\nu - g^{\mu\nu}(k^2 - x(1-x)p^2 - m^2)}{[k^2 + p^2x(1-x) - m^2]^2} \quad (3.5.9)$$

Using dimensional regularization, we find

$$\Pi_2^{\mu\nu} = -\frac{e^2}{2\pi^2} p^2 g^{\mu\nu} \int_0^1 x(1-x) \left[ \frac{2}{\epsilon} + \log \left( \frac{\tilde{\mu}^2}{m^2 - p^2x(1-x)} \right) \right] \quad (3.5.10)$$

In particular, in the limit of large  $Q^2 = -p^2 \gg m^2$

$$\Pi_2^{\mu\nu} = -\frac{e^2}{12\pi^2} p^2 g^{\mu\nu} \left[ \frac{2}{\epsilon} + \log \left( \frac{\tilde{\mu}^2}{Q^2} \right) + \frac{5}{3} \right] \quad (3.5.11)$$

### 3.5.1 Vacuum Polarization and Coulomb Potential Renormalization

One can also obtain the  $p^\mu p^\nu$  pieces and see that they automatically satisfy the Ward identity. We can write the full result as

$$\Pi_2^{\mu\nu} = e^2(p^\mu p^\nu - p^2 g^{\mu\nu}) \Pi_2(p^2) \quad (3.5.12)$$

where

$$\Pi_2(p^2) = \frac{1}{2\pi^2} \int_0^1 x(1-x) \left[ \frac{2}{\epsilon} + \log \left( \frac{\tilde{\mu}^2}{m^2 - p^2x(1-x)} \right) \right] \quad (3.5.13)$$

The dressed photon propagator at 1-loop in Feynman gauge ( $\xi = 1$ ) is

$$iG^{\mu\nu} = -i \frac{(1 - e^2 \Pi_2(p^2)) g^{\mu\nu}}{p^2} \quad (3.5.14)$$

where I have ignored  $p^\mu p^\nu$  terms. This gives the corrected Coulomb potential in Fourier space

$$\tilde{V}(p) = e^2 \frac{1 - e^2 \Pi_2(p^2)}{p^2} \quad (3.5.15)$$

Since  $\Pi_2$  has a logarithmic dependence on momenta at large momentum, or at very short distances, we see that the electromagnetic force has this additional scale dependence. In fact, it grows stronger at short distances, as we can easily see by noting that  $\Pi_2 \propto -\log Q^2$ .

To see this more clearly and precisely, we need to renormalize. In this case, we literally need to *re-normalize* the strength of the Coulomb force. It would be most physical to do this by fixing

$$V(r_0) = -\frac{e_R^2}{4\pi r_0} \quad (3.5.16)$$

for some position space scale  $r_0$ , but it's easier to just keep working in momentum space. Thus we can define a renormalized charge at the scale  $p_0$  by *defining*

$$\tilde{V}(p_0^2) = \frac{e_R^2}{p_0^2} \quad (3.5.17)$$

In perturbation theory, this means that

$$e_R^2 = p_0^2 \tilde{V}(p_0) \approx e^2 - e^4 \Pi_2(p_0^2) + \dots \quad (3.5.18)$$

or alternatively we can solve for the *bare* parameter  $e$  that appears in the Lagrangian in terms of  $e_R$

$$e^2 \approx e_R^2 + e_R^4 \Pi(p_0^2) + \dots \quad (3.5.19)$$

Note that since  $\Pi(p^2)$  depends on  $\epsilon = 4 - d$ , this means that  $e$  must depend on  $\epsilon$  as well. However,  $e_R$  does not depend on  $\epsilon$ , as it is a measurable quantity associated with the strength of the Coulomb potential at the scale  $p_0$ !

The potential at another scale  $p$  (which is measurable) will be

$$\begin{aligned} \tilde{V}(p) &= \frac{e^2 - e^4 \Pi_2(p^2)}{p^2} + \dots \\ &= \frac{e_R^2 - e_R^4 (\Pi_2(p^2) - \Pi_2(p_0^2))}{p^2} + \dots \end{aligned} \quad (3.5.20)$$

This makes it very clear that the potential changes with distance. We can view the numerator as an effective charge

$$e_{eff}^2(Q) \approx e_R^2 \left( 1 + \frac{e_R^2}{12\pi^2} \log \frac{Q^2}{m^2} \right) \quad (3.5.21)$$

Near any fixed value of  $p$ , we can view the potential as Coulomb with effective coupling constant  $e_{eff}$ . Numerically, note that

$$\alpha_{eff}(Q^2) \approx \frac{1}{137} \left( 1 + 0.00077 \log \frac{Q^2}{m^2} \right) \quad (3.5.22)$$

so since the coefficient of the logarithm is small, this isn't a large effect.

Note that we can immediately sum up all of the 1-loop vacuum polarization contributions from 1-PI diagrams, giving a trivial geometric series that we saw last semester. We have the running coupling

$$e_{eff}^2(Q) = \frac{e_R^2}{1 - \frac{e_R^2}{12\pi^2} \log \frac{Q^2}{m^2}} \quad (3.5.23)$$

which satisfies the RG equation

$$\frac{de_{eff}}{d\log\mu} = \frac{e_{eff}^3}{12\pi^2} \quad (3.5.24)$$

The right-hand side is the  $\beta$  function.

### 3.5.2 Lamb Shift

What does our 1-loop correction do to the Coulomb potential at large distances?

In the case where  $|p^2| \ll m^2$  we have

$$\begin{aligned} \int_0^1 dx x(1-x) \log(m^2 - p^2 x(1-x)) &\approx \frac{1}{6} \log m^2 - \int_0^1 dx x^2(1-x)^2 \frac{p^2}{m^2} \\ &= \frac{1}{6} \log m^2 - \frac{p^2}{30m^2} \end{aligned} \quad (3.5.25)$$

If we choose  $p_0 = m$  then we find

$$\tilde{V}(p) = \frac{e_R^2}{p^2} - \frac{e_R^4}{60\pi^2 m^2} \quad (3.5.26)$$

Since the Fourier transform of 1 is a delta function, this just gives

$$V(r) = -\frac{e_R^2}{4\pi r} - \frac{e_R^4}{60\pi^2 m^2} \delta(r) \quad (3.5.27)$$

This is a very short ranged potential called the Uehling term.

Inserting this extra interaction into the Hamiltonian for the Hydrogen atom gives an effect called the Lamb shift (first measured by Wallis Lamb in 1947), which can be easily computed in first order perturbation theory. It is only non-zero for  $s$ -waves, which have support at the origin. Apparently Bethe got the answer right first using a rough method, then Feynman, Schwinger, and Tomonaga all provided a complete calculation, but only Tomonaga got the right answer.

### 3.5.3 Anomalous Magnetic Moment and Form Factors

Now let us compute the 1-loop contribution to the magnetic moments of charged particles. First, we need to understand how this appears.

Recall that we showed that

$$\not{D}^2 = D_\mu D^\mu + \frac{e}{2} F_{\mu\nu} \sigma^{\mu\nu} \quad (3.5.28)$$

where

$$\sigma^{\mu\nu} = \frac{i}{2} [\gamma^\mu, \gamma^\nu] \quad (3.5.29)$$

encodes the difference between scalar and spinor equations of motion. We wrote this as

$$\left[ D_\mu^2 + m^2 - e \begin{pmatrix} (\vec{B} + i\vec{E}) \cdot \vec{\sigma} & 0 \\ 0 & (\vec{B} - i\vec{E}) \cdot \vec{\sigma} \end{pmatrix} \right] \psi = 0 \quad (3.5.30)$$

Note that the  $i\vec{E} \cdot \sigma$  terms turn particles into anti-particles and vice versa, so if we work in the non-relativistic limit with only particles or only anti-particles then this term does not act.

If we go to momentum space this implies

$$\frac{(H - eA_0)^2}{2m} \psi = \left( \frac{m}{2} + \frac{(\vec{p} - e\vec{A})^2}{2m} - 2\frac{e}{2m} \vec{B} \cdot \vec{S} \pm i\frac{e}{m} \vec{E} \cdot \vec{S} \right) \psi \quad (3.5.31)$$

This can be compared to read off the strength of the dipole interaction

$$ge\vec{B} \cdot \vec{S} \quad (3.5.32)$$

to see that  $g = 2$  for the semi-classical theory, since  $\vec{S} = \frac{\vec{\sigma}}{2}$  for spin 1/2 particles.

This corresponds to a magnetic dipole moment, and we had obtained the prediction that

$$\mu_B = \frac{e}{2m_e} \quad (3.5.33)$$

So a clear way to obtain quantum corrections is to look for loops that generate

$$\frac{e}{2} F_{\mu\nu} \sigma^{\mu\nu} \quad (3.5.34)$$

terms.

A general way to think about this is to consider matrix elements with on-shell fermions, but off-shell photons. We take the latter to be off-shell because we want to view it as a background electromagnetic field (alternatively we can think of it as a virtual photon from some other charged particles). Thus we can write

$$i\Gamma^\mu = \bar{u}(q_2) (f_1 \gamma^\mu + f_2 p^\mu + f_3 q_1^\mu + f_4 q_2^\mu) u(q_1) \quad (3.5.35)$$

We could also have included  $\gamma_5$  in a theory that breaks parity... but in QED parity is preserved.

Of course the  $f_i$  are not independent, as they are constrained by momentum conservation and the Ward identity. The former is trivial, while the latter says that

$$\begin{aligned} 0 &= p_\mu \Gamma^\mu = p_\mu \bar{u}(f_1 \gamma^\mu + f_3 q_1^\mu + f_4 q_2^\mu) u \\ &= [(p \cdot q_1) f_3 + (p \cdot q_2) f_4] \bar{u} u \end{aligned} \quad (3.5.36)$$

because  $\bar{u} \not{p} u$  vanishes on the equations of motion, with  $p = -q_1 - q_2$ . Also, since  $p \cdot q_1 = -m^2 - q_1 \cdot q_2 = -p \cdot q_2$ , we can conclude that  $f_3 = f_4$ . Now we can use the Gordon identity

$$\bar{u}(q_2)(q_1^\mu + q_2^\mu)u(q_1) = (2m)\bar{u}(q_2)\gamma^\mu u(q_1) + i\bar{u}(q_2)\sigma^\mu{}_\nu(q_1^\nu - q_2^\nu)u(q_1) \quad (3.5.37)$$

to write the result as

$$i\Gamma^\mu = (-ie)\bar{u}(q_2) \left[ F_1 \left( \frac{p^2}{m^2} \right) \gamma^\mu + \frac{i\sigma^{\mu\nu}}{2m} p_\nu F_2 \left( \frac{p^2}{m^2} \right) \right] u(q_1) \quad (3.5.38)$$

The leading order contribution (tree level) gives

$$F_1 = 1, \quad F_2 = 0 \quad (3.5.39)$$

but loops can correct both functions. Corrections to  $F_1$  are just the running of the electromagnetic force, so only  $F_2$  will correct the magnetic moment. One can write the  $g$  factor

$$g = 2 + 2F_2(0) \quad (3.5.40)$$

so we need only compute  $F_2(0)$ .

### 3.5.4 Diagrammatics of Magnetic Moment

There are four possible graphs that could contribute to the electron-electron-photon amplitude with an off-shell photon, which we are calling  $\Gamma^\mu$ . But only one graph can produce a shift of  $F_2$ , because the propagator renormalizations only shift the vertex by  $\gamma^\mu$ , and therefore only contribute to  $F_1$ . Thus we need only compute one graph, which is

$$i\Gamma_2^\mu = (-ie)^3 \int \frac{d^4k}{(2\pi)^4} \frac{-ig^{\nu\alpha}}{(k - q_1)^2} \bar{u}(q_2) \gamma^\nu \frac{i(\not{p} + \not{k} + m)}{(p + k)^2 - m^2} \gamma^\mu \frac{i(\not{k} + m)}{k^2 - m^2} \gamma^\alpha u(q_1) \quad (3.5.41)$$

To evaluate this, one employs the usual tricks of combining denominator factors using Feynman parameterization, shifting the integration variable, and then simplifying using some algebra. You can see more of the algebraic details in the book.

One obtains three terms, because the Ward identity follows as a consequence of the computation, and isn't obvious from intermediate steps. Simplifying using the Ward identity and Gordon identity and then only keeping the  $\sigma^{\mu\nu}$  term of interest, we find that

$$F_2(p^2) = \frac{2m}{e} (4ie^3 m) \int_0^1 dx \int_0^x dz \int \frac{d^4k}{(2\pi)^4} \frac{z(1-z)}{(k^2 - \Delta + i\epsilon)^3} \quad (3.5.42)$$

where  $\Delta = -x(1-x-z)p^2 + (1-z)^2m^2$ .

The integral

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2 - \Delta + i\epsilon)^3} = \frac{-i}{32\pi^2 \Delta} \quad (3.5.43)$$

is finite and so we find

$$F_2(p^2) = \frac{\alpha}{\pi} \int_0^1 dx \int_0^x dz \frac{z(1-z)}{(1-z)^2 m^2 - x(1-x-z)p^2} \quad (3.5.44)$$



We can evaluate this explicitly at  $p^2 = 0$ , which is what we need for the magnetic moment. We find

$$F_2(0) = \frac{\alpha}{2\pi} \quad (3.5.45)$$

and so the magnetic moment is

$$g = 2 + \frac{\alpha}{\pi} = 2.00232 \quad (3.5.46)$$

with the next correction of order  $\alpha^2$ .

To summarize, we first connected the relativistic QED theory to the non-relativistic Hamiltonian, in order to show how QFT gives rise to the magnetic moment. Then we studied the relevant coupling of physical electrons to a general background electromagnetic field to understand what form factors can appear, namely  $F_1$  and  $F_2$  in our parameterizations, and we noted that only  $F_2$  affects the magnetic moment. Finally, we identified a unique diagram that contributes to  $F_2$ , and computed it using the Feynman rules, finding that it's finite (not short-distance sensitive) and that it gives a contribution to  $g - 2$ .

### 3.6 Renormalization and QED

We had an extensive discussion last semester about renormalization. To recap

- One can categorize quantum corrections as those sensitive to short-distances, those that are associated with the scale of a measurement or physical setup (e.g. the center of mass energy of a scattering process), and those associated with very long distances. The short-distance sensitivity must be absorbed into the definition of the theory, but once it's defined, long-distance predictions are unambiguous.
- Logarithmic corrections are of great importance, because they are sensitive to physics at all scales, and slowly alter the strength of forces.
- In any give QFT, there are a finite number of parameters that define the long-distance limit of the theory. These are the relevant and marginal parameters. In QED we can write

$$\mathcal{L} = -\frac{1}{4}Z_3 F_{\mu\nu}^2 + iZ_2 \bar{\psi} \not{D} \psi - Z_2 Z_m m_R \bar{\psi} \psi - e_R Z_1 \bar{\psi} \not{A} \psi + \rho_0 \quad (3.6.1)$$

Where  $Z_1, Z_2, Z_3, Z_m$  are renormalizations of the various field strengths, interactions, and masses, and  $\rho_0$  is a renormalization of the vacuum energy that we will ignore. We only need these 5 parameters to absorb all short-distance sensitivity in the theory, and in fact we will find that we must have

$$Z_1 = Z_2 \quad (3.6.2)$$

due to gauge invariance.

- It's often conventional to write

$$Z_1 = 1 + \delta_1, \quad Z_2 = 1 + \delta_2, \quad Z_3 = 1 + \delta_3, \quad Z_m = 1 + \delta_m \quad (3.6.3)$$

in recognition of the fact that these are 1 plus a ‘counterterm’, a short-distance correction used to re-normalize and re-define the theory, calibrating it with experiment. Note that the  $\delta_i$  and  $Z_i$  are just numbers, they do not depend on fields or on space and time. We can view the  $\delta_i$  as perturbative corrections, to adjust when we renormalize the theory... the book calls this procedure ‘renormalized perturbation theory’.

You should also note that we do not have a counterterm for the photon mass, because it would break gauge invariance – thus it's important that there not be any UV divergences that would require such a counterterm.

- In renormalized perturbation theory, we are performing a formal series expansion in  $e_R$ , and we legitimately have

$$|e_R| \ll 1 \quad (3.6.4)$$

since it is a renormalized quantity, insensitive to the cutoff or, in our case, the  $1/\epsilon$  of dimensional regulation.

- Much of the point of organizing things systematically in this way is to understand that one could proceed in the same way and compute effects at 2, 3, 4,  $\dots$  loops. We would not need any new counterterms, but could absorb all UV sensitivity into the ones we have, and then make very precise predictions.
- There are also irrelevant operators like

$$\epsilon_{\alpha\beta\mu\nu} \frac{F^{\alpha\beta}}{\Lambda} \bar{\psi} \sigma^{\mu\nu} \psi, \quad \frac{F_{\mu\nu}}{\Lambda} \bar{\psi} \sigma^{\mu\nu} \psi, \quad \frac{1}{\Lambda^4} F_{\mu\nu}^4, \quad \frac{(\bar{\psi}\psi)^2}{\Lambda^2}, \quad \dots \quad (3.6.5)$$

that we could add to the Lagrangian, but these do not renormalize the relevant and marginal operators, and they are irrelevant at long distances, so we can and will ignore them.

But they could be lurking in the background as a signal of new physics, especially the first term, which gives an electric dipole moment.

We have already discussed two of the three possible divergent 1-PI 1-loop diagrams in QED, the correction to the photon propagator and the vertex correction. So let us first use those effects.

### 3.6.1 Photon Self Energy and Renormalization

The photon self-energy was encapsulated in

$$\Pi_2(p^2) = \frac{1}{2\pi^2} \int_0^1 x(1-x) \left[ \frac{2}{\epsilon} + \log \left( \frac{\tilde{\mu}^2}{m^2 - p^2 x(1-x)} \right) \right] \quad (3.6.6)$$

where

$$i\Pi_2^{\mu\nu} = ie_R^2(p^\mu p^\nu - p^2 g^{\mu\nu})\Pi_2(p^2) \quad (3.6.7)$$

In renormalized perturbation theory, the counterterms contribute to this quantity as

$$i\delta_3(p^\mu p^\nu - p^2 g^{\mu\nu}) \quad (3.6.8)$$

These are the only two contributions.

If we work in Lorenz gauge where  $\xi = 0$ , things are quite simple, and we find that the photon propagator gets renormalized to (summing the geometric series)

$$iG^{\mu\nu} = -i \left( g^{\mu\nu} - \frac{p^\mu p^\nu}{p^2} \right) \frac{1}{p^2 (1 + e_R^2 \Pi_2(p) + \delta_3) + i\epsilon} \quad (3.6.9)$$

Note that, as promised, *the photon mass is still zero after renormalization*.

We need only one renormalization condition to fix  $\delta_3$ . This is up to us, as long as we do it consistently, and eliminate UV sensitivity. One possibility is to use modified minimal subtraction, so-called MS-bar, so that

$$\delta_3 = -\frac{e_R^2}{6\pi^2\epsilon} \quad (3.6.10)$$

Another choice, which the book suggests, is to impose that the *residue of the photon propagator pole remains 1*. The point is to keep the normalization of the physical photon state the same, and this normalization corresponds to the residue of the pole. This requires

$$\delta_3 = -e_R^2 \Pi_2(0) = -\frac{e_R^2}{6\pi^2\epsilon} - \frac{e_R^2}{12\pi^2} \log \frac{\tilde{\mu}^2}{m_R^2} \quad (3.6.11)$$

These are two different choices of renormalization scheme.

### 3.6.2 3-pt Vertex Renormalization

We also already studied the 1-PI contribution to the 3-pt vertex when we computed the magnetic moment. We parameterized it as

$$i\Gamma^\mu = (-ie)\bar{u}(q_2) \left[ F_1 \left( \frac{p^2}{m^2} \right) \gamma^\mu + \frac{i\sigma^{\mu\nu}}{2m} p_\nu F_2 \left( \frac{p^2}{m^2} \right) \right] u(q_1) \quad (3.6.12)$$

We saw that  $F_2$  is entirely responsible for the anomalous magnetic moment,  $g - 2$ , and that it was finite, so it doesn't require a counter-term. That's good, because we do not have a counterterm that could separately absorb a divergence; to include one we'd need the operator

$$\frac{F_{\mu\nu}}{\Lambda} \bar{\psi} \sigma^{\mu\nu} \psi \quad (3.6.13)$$

but we have not (and should not need to) explicitly included such an operator in the Lagrangian. A divergent contribution to  $F_2$  would have been a disaster.

However, we also computed (or could have)  $F_1$ , which is

$$F_1(p) = 1 - 2ie_R^2 \int \frac{d^4k}{(2\pi)^4} \int dx dy dz \delta(x+y+z-1) \frac{k^2 - 2(1-x)(1-y)p^2 - 2(1-4z+z^2)m_R^2}{[k^2 - (M_R^2(1-z)^2 - xyp^2)]^3}$$

There is a very natural renormalization scheme for  $F_1$ , which is to choose

$$\Gamma^\mu(0) = \gamma^\mu \quad \text{or} \quad F_1(0) = 1 \quad (3.6.14)$$

This implies that  $e_R$  is the physical electric charge measured by Coulomb's law at large distances (which correspond to zero momentum, of course). This condition sets

$$\begin{aligned} \delta_1 &= 1 - F_1(0) \\ &= \frac{e_R^2}{8\pi^2} \left( -\frac{1}{\epsilon} - \frac{1}{2} \log \frac{\tilde{\mu}^2}{m_R^2} - \frac{5}{2} - \log \frac{m_\gamma^2}{m_R^2} \right) \end{aligned} \quad (3.6.15)$$

where  $m_\gamma$  is an IR regulator (we'll discuss IR divergences soon). We could, of course, also use the MS-bar scheme instead, and then we would just define  $\delta_1$  to cancel the  $1/\epsilon$  pole.

### 3.6.3 Electron Propagator Corrections

Finally, we need to compute corrections to the electron propagator. There is a unique diagram, and it is

$$i\Sigma_2(\not{p}) = (-ie)^2 \int \frac{d^4k}{(2\pi)^4} \gamma^\mu \frac{i(\not{k} + m)}{k^2 - m^2} \gamma_\mu \frac{-i}{(k-p)^2} \quad (3.6.16)$$

Combining denominators and shifting gives

$$i\Sigma_2(\not{p}) = 2e^2 \int_0^1 dx \int \frac{d^4k}{(2\pi)^4} \gamma^\mu \frac{(x\not{p} - 2m)}{[k^2 - \Delta]^2} \gamma_\mu \frac{-i}{(k-p)^2} \quad (3.6.17)$$

where  $\Delta = (1-x)(m^2 - p^2x)$ . Notice that *this entire integral is merely logarithmically divergent*, and in particular there is no linear or quadratic divergence associated with a renormalization of the fermion mass. This is a crucial property – fermion masses can be *naturally small*, unlike scalar boson masses.

This is due to *chiral symmetry*. Note that the mass term is

$$m(\psi_L \psi_R + \psi_R^\dagger \psi_L^\dagger) \quad (3.6.18)$$

and so it breaks the symmetry transformation

$$\psi_L, \psi_R \rightarrow e^{i\alpha} \psi_L, e^{i\alpha} \psi_R \quad (3.6.19)$$

This chiral symmetry is preserved by the kinetic term, which involves  $\psi_L^\dagger \not{D} \psi_L$  etc. Thus in the limit that  $m \rightarrow 0$ , chiral symmetry is a good symmetry, and fermion masses cannot be generated by short-distance physics. Any corrections to  $m$  must be proportional to  $m$  itself, so  $m$  can at most receive logarithmic corrections.

In dimensional regularization this is

$$\Sigma_2(p) = -\frac{\alpha}{2\pi} \int_0^1 dx (2m - x\not{p}) \left[ \frac{2}{\epsilon} + \log \frac{\tilde{\mu}^2}{(1-x)(m^2 - p^2x)} \right] \quad (3.6.20)$$

where  $\tilde{\mu}^2 = 4\pi e^{-\gamma_E} \mu^2$  as usual. We also have a contribution from the counterterms

$$i(\delta_2 \not{p} - (\delta_m + \delta_2)m_R) \quad (3.6.21)$$

We can resum the geometric series of 1-PI graphs connected with free propagators, giving

$$iG(p) = \frac{i}{\not{p} - m + (\Sigma_2(\not{p}) + \delta_2 \not{p} - (\delta_m + \delta_2)m_R)} \quad (3.6.22)$$

If we demand that the propagator has a pole at the physical mass  $m_R$  and that the residue of that pole is 1 then we must have

$$\delta_2 = -\frac{d}{d\not{p}} \Sigma_2(\not{p}) \Big|_{\not{p}=m_R}, \quad \delta_m = \frac{1}{m_R} \Sigma(m_R) \quad (3.6.23)$$

This is a sensible physical scheme, that makes  $m_R$  the physical electron mass, with the same propagator normalization as in the free theory. In particular, we find

$$\delta_2 = \frac{\alpha}{2\pi} \left( -\frac{1}{\epsilon} - \frac{1}{2} \log \frac{\tilde{\mu}^2}{m_R^2} - \frac{5}{2} - \log \frac{m_\gamma^2}{m_R^2} \right) \quad (3.6.24)$$

$$\delta_m = \frac{\alpha}{2\pi} \left( -\frac{3}{\epsilon} - \frac{3}{2} \log \frac{\tilde{\mu}^2}{m_R^2} - \frac{5}{2} \right) \quad (3.6.25)$$

in this scheme. In this scheme, the renormalized propagator has no dependence on either  $\mu$  or  $\epsilon$ . In minimal subtraction one would only have the  $\epsilon$  pole term.

Crucially, we see that

$$\delta_1 = \delta_2 \quad (3.6.26)$$

This was important because these appear in

$$iZ_2 \bar{\psi} \not{A} \psi - e_R Z_1 \bar{\psi} \not{A} \psi \quad (3.6.27)$$

and so the gauge invariance

$$\psi \rightarrow e^{-ie_R \alpha(x)} \psi, \quad A_\mu \rightarrow A_\mu + \partial_\mu \alpha \quad (3.6.28)$$

has been preserved by renormalization. Another way of saying this is that the charge current

$$J_\mu = \bar{\psi} \gamma_\mu \psi \quad (3.6.29)$$

cannot be renormalized. One can prove this formally using the Ward-Takahashi identities. These features are special to gauge theories; there was no relation between different counterterms in  $\phi^3$  and  $\phi^4$  theory.

### 3.6.4 Summary of QED Renormalization and RG Flow

We have imposed renormalization conditions so that the renormalized mass  $m_R$  is equal to the physical electron mass, the renormalized photon and electron propagators have pole residues of 1, and the renormalized charge  $e_R$  is the charge seen at long-distances. This resulted in counterterms

$$\delta_1 = \delta_2 = \frac{e_R^2}{8\pi^2} \left( -\frac{1}{\epsilon} - \frac{1}{2} \log \frac{\tilde{\mu}^2}{m_R^2} - \frac{5}{2} - \log \frac{m_\gamma^2}{m_R^2} \right) \quad (3.6.30)$$

$$\delta_3 = -\frac{e_R^2}{6\pi^2\epsilon} - \frac{e_R^2}{12\pi^2} \log \frac{\tilde{\mu}^2}{m_R^2} \quad (3.6.31)$$

$$\delta_m = \frac{\alpha}{2\pi} \left( -\frac{3}{\epsilon} - \frac{3}{2} \log \frac{\tilde{\mu}^2}{m_R^2} - \frac{5}{2} \right) \quad (3.6.32)$$

In the MS-bar scheme, only the  $\epsilon$  dependent terms would be present.

We already saw in QED that the effective charge changes with distance/energy scale as

$$\frac{de_{eff}}{d \log \mu} = \frac{e_{eff}^3}{12\pi^2} \quad (3.6.33)$$

as a direct consequence of summing 1-PI diagrams to get an effective Coulomb potential

$$e_{eff}^2(Q) = \frac{e_R^2}{1 - \frac{e_R^2}{12\pi^2} \log \frac{Q^2}{m^2}} \quad (3.6.34)$$

We can also view this as the solution to the RG equation.

But another way to get the same result is to note that, as we discussed last semester, *the bare parameters in the Lagrangian must be  $\mu$ -independent, since they define the theory at very short distances.* The bare fields and parameters are

$$A_R = \frac{1}{\sqrt{Z_3}} A_0, \quad \psi_R = \frac{1}{\sqrt{Z_2}} \psi_0, \quad m_R = \frac{1}{Z_m} m_0, \quad e_R = \frac{Z_2}{Z_1} \sqrt{Z_3} \mu^{-\frac{\epsilon}{2}} e_0 \quad (3.6.35)$$

We have to absorb  $\mu^{\epsilon/2}$  into the coupling so that it remains dimensionless in  $d = 4 - \epsilon$  dimensions. Also,  $Z_1 = Z_2$ . This means that we can use

$$\begin{aligned} 0 &= \mu \frac{d}{d\mu} e_0 = \mu \frac{d}{d\mu} \left[ \frac{1}{\sqrt{Z_3}} \mu^{\epsilon/2} e_R \right] \\ &= e_R \frac{1}{\sqrt{Z_3}} \mu^{\frac{\epsilon}{2}} \left( \frac{\epsilon}{2} + \frac{d \log e_R}{d \log \mu} - \frac{1}{2} \frac{d \log Z_3}{d \log \mu} \right) \end{aligned} \quad (3.6.36)$$

This is the same sort of differential equation that we saw last semester. To first order in  $e_R$ , we have  $Z_3 = 1$  identically, and so

$$\frac{d \log e_R}{d \log \mu} = -\frac{\epsilon}{2} e_R \quad (3.6.37)$$

That just tells us the scale dependence of the classical coupling in  $4 - \epsilon$  dimensions. At next order we have

$$\begin{aligned}\frac{d \log Z_3}{d \log \mu} &\approx -\frac{e_R}{3\pi^2\epsilon} \frac{de_R}{d \log \mu} - \frac{e_R^2}{6\pi^2} - \frac{e_R}{6\pi^2} \frac{de_R}{d \log \mu} \log \frac{\mu^2}{m_R^2} \\ &\approx \frac{e_R^2}{6\pi^2} + \mathcal{O}(\epsilon, e_R^4)\end{aligned}\tag{3.6.38}$$

And so at next order we find

$$\frac{d \log e_R}{d \log \mu} = -\frac{\epsilon}{2} e_R + \frac{e_R^2}{12\pi^2}\tag{3.6.39}$$

So we have derived the same (renormalization) scale dependence for  $e_R$  using this more formal method! In more general theories, as discussed before, we can use the RG to resum logarithms, and to understand the strength of forces and interactions as a function of energy or distance scale.

### 3.7 IR Divergences and Long Wavelength Physics

A point that was made last semester, but upon which we never followed up, is that in order to obtain sensible results from QFT, we have to be careful to ask physically sensible questions. In particular, asking that electrons scatter without emitting any photons at all, even photons of enormous wavelength and tiny energy... is unphysical. We need to carefully define what experiment we are really doing, with finite energy resolutions, and then add up all contributions. Even the S-Matrix (as we've defined it up to this point) isn't a sufficiently good 'observable', but carefully defined cross sections still make sense.

We will see now that the emission of soft photons produces effects that are so large that they must be summed to all orders in perturbation theory, but so simple that this process is possible. This generalizes to other theories involving massless particles (there are no IR divergences in theories where all particles have significant mass), although there are complications we won't address when several types of massless particles interact with each other. Our discussion will closely follow that of Weinberg, chapter 13.

Consider some scattering amplitude, and let us add to it a *soft photon* with momentum  $q$ . In the soft limit, the amplitude will be dominated by soft photons that are attached to external legs. On the  $i$ th leg, if we assume that particle is a scalar, this gives

$$M_i(p_i, q) = (-ieQ_i) \frac{i(p_i^\mu + (p_i^\mu - q))}{(p_i - q)^2 - m^2} \epsilon_\mu M_0(p_i - q)\tag{3.7.1}$$

where  $Q_i$  is the charge of the  $i$ th particle. Note that since  $p_i^2 = m^2$  and  $q^2 = 0$ , this is

$$M_i(p_i, q) \approx eQ_i \frac{p_i \cdot \epsilon}{p_i \cdot q} M_0(p_i)\tag{3.7.2}$$

to leading order at small  $q$ . If the charged particle is a fermion of spin 1/2, then must replace

$$\bar{u}(p_i) \rightarrow \bar{u}(p_i) e \gamma^\mu \frac{i(\not{p}_i + \not{q} + m)}{(p + q)^2 - m^2}\tag{3.7.3}$$

But in the limit  $q \rightarrow 0$  we can write the numerator as

$$\not{p} + m = \sum_s u_s(p) \bar{u}_s(p) \quad (3.7.4)$$

and use the fact that

$$\bar{u}_s(p) \gamma^\mu u_{s'}(p) = 2\delta_{s,s'} p^\mu \quad (3.7.5)$$

to simplify our result to

$$\begin{aligned} \bar{u}(p_i) e \gamma^\mu \frac{i(\not{p}_i + \not{q} + m)}{(p+q)^2 - m^2} &= e \bar{u}(p_i) \gamma^\mu \frac{i \sum_s u_s(p) \bar{u}_s(p)}{2p \cdot q} \\ &= e \bar{u}(p_i) \frac{p^\mu}{p \cdot q} \end{aligned} \quad (3.7.6)$$

Thus we see that the soft factor for fermions is identical to that for scalars.

More generally, for a particle with any spin, as  $q \rightarrow 0$  the external line approaches the mass shell, and we can re-write the numerator structure so that it turns into  $2p^\mu$  times a delta function on all spin indices. In a physical renormalization scheme, none of these properties are affected, and so this result will hold to all orders in perturbation theory.

*External lines* dominate because they produce poles in  $q$  as  $q \rightarrow 0$ . Another way of saying the same thing is that soft photons have very large wavelength, and so they do not care about short-distance processes. They act like a classical background against which the rest of the scattering process unfolds.

In fact, one can obtain the soft amplitude factor by viewing the external particles as classical sources for the photon field  $A_\mu(x)$ . The source, or classical current for the  $i$ th particle, is

$$J_i(x) = Q_i \delta^3(\vec{x} - \hat{q}_i t) \quad (3.7.7)$$

and so we can obtain the soft amplitude factor from the first-quantized action of the hard particle, viewed as a source. The action will contain a term

$$S_{1-part} = \int d\tau (m + \hat{n}_\mu A^\mu(x(\tau))) \quad (3.7.8)$$

where  $\hat{n}^\mu$  is the four vector of the particles trajectory. This gives

$$\begin{aligned} e^{ie \int_0^\infty d\tau \hat{n}_i \cdot A(x(\tau))} &\approx 1 + ie \int_0^\infty d\tau (\hat{n}_i \cdot \epsilon) e^{i\hat{n} \cdot q\tau} + \dots \\ &\approx e^{\frac{p_i \cdot \epsilon}{p_i \cdot q}} \end{aligned} \quad (3.7.9)$$

where we note that for a mode with fixed momentum  $q$  we have

$$A_\mu(x) = \epsilon_\mu e^{iq \cdot x} \quad (3.7.10)$$



We have obtained the soft factor from integrating the electromagnetic field  $A(x(\tau))$  along the classical worldline of each charged particle in the scattering process.

So we have the soft factor for emission from one external leg. We can sum to get the same contribution from each leg, which means that the total soft photon emission amplitude will be

$$M \approx eM_0(p_j) \left[ \sum_{incoming} Q_i \frac{p_i \cdot \epsilon}{p_i \cdot q} - \sum_{outgoing} Q_i \frac{p_i \cdot \epsilon}{p_i \cdot q} \right] \quad (3.7.11)$$

where the approximation means that we are only considering the leading term when  $q$  is small.

### 3.7.1 Multiple Soft Emissions

Now consider the emission of two soft photons. Clearly if they are emitted from different external legs then we just get a product of soft factors

$$\left( \frac{p_i \cdot \epsilon_1}{p_i \cdot q_2} \right) \left( \frac{p_j \cdot \epsilon_2}{p_j \cdot q_2} \right) \quad (3.7.12)$$

What may be more surprising is that we get the same result if they are emitted from the same leg. This gives the sum, depending on the two possible orderings, which is

$$\begin{aligned} & \left( \frac{p \cdot \epsilon_1}{p \cdot q_2} \right) \left( \frac{p \cdot \epsilon_2}{p \cdot (q_1 + q_2)} \right) + \left( \frac{p \cdot \epsilon_2}{p \cdot q_1} \right) \left( \frac{p \cdot \epsilon_1}{p \cdot (q_1 + q_2)} \right) \\ &= \left( \frac{p \cdot \epsilon_1}{p \cdot q_2} \right) \left( \frac{p \cdot \epsilon_2}{p \cdot q_2} \right) \end{aligned} \quad (3.7.13)$$

so actually we again obtain exactly the product of the soft factors, ignoring the fact that the soft photons have been emitted by the same line.

One can prove by induction that this occurs for any number of soft emissions. Thus the amplitude for emitting  $N$  soft photons is just

$$M \approx e^N M_0(p_j) \prod_r^N \left[ \sum_{incoming} Q_i \frac{p_i \cdot \epsilon_r}{p_i \cdot q_r} - \sum_{outgoing} Q_i \frac{p_i \cdot \epsilon_r}{p_i \cdot q_r} \right] \quad (3.7.14)$$

exactly in the soft limit. So now we know how to emit any number of very long wavelength photons (or other massless particles, by equivalent reasoning).

### 3.7.2 Virtual Soft Photons

Now we know how soft photons would be emitted in an amplitude, let us study virtual corrections from soft photon loops. We will introduce two infrared scales with

$$E \gg \Lambda \gg \lambda \quad (3.7.15)$$

Here the scale  $\Lambda$  just defines what we mean by ‘soft’ photons, and all  $\Lambda$  dependence from soft photons is automatically canceled by the  $\Lambda$  dependence of the ‘hard’ photons. However, the scale  $\lambda$  is some lower limit introduced to absorb divergences, it is an infrared regulator. The  $\lambda$  dependence will only cancel *when we figure out how to ask a physically sensible question*. This involves including real soft photon emission.

For each virtual photon we must multiply by a propagator factor

$$\frac{-ig^{\mu\nu}}{q^2 + i\epsilon} \quad (3.7.16)$$

in Feynman gauge. We must multiply our multiple soft emission formula by this factor for each pair of photons, and sum over all possible ways of connecting up the propagators. We also have to divide by  $2^N N!$  because summing over the places where we can attach the photons contributes spurious combinatorial factors from photon lines and reversing photon lines.

Thus the effect of including  $N$  soft photons involves multiplying the original matrix element by

$$\frac{1}{N!2^N} \left[ \sum_{n,m} Q_n Q_m J_{nm} \right]^N \quad (3.7.17)$$

where we have

$$J_{nm} = -ip_n \cdot p_m \int_{\lambda < |\vec{q}| < \Lambda} \frac{d^4 q}{(q^2 + i\epsilon)(p_n \cdot q + i\epsilon)(-p_m \cdot q + i\epsilon)} \quad (3.7.18)$$

There’s a sign in the denominator since incoming  $q$  must be matched by outgoing  $q$ . There are also other signs that we should have been more careful about, but they can be accounted for most simply by assuming that all  $p_n$  are taken to be incoming, with negative energy and reversed momenta if they are actually outgoing.

Summing over all  $N$ , we can write our result so far as the statement that

$$M^\lambda = M^\Lambda \exp \left[ \frac{1}{2} \sum_{n,m} Q_n Q_m J_{nm} \right] \quad (3.7.19)$$

which means that the amplitude for virtual soft photons down to a scale  $\lambda \ll \Lambda$  is given by the amplitude for hard processes times a universal exponential factor.

Now we need to evaluate the integral  $J_{nm}$ , which can be done by contour integration in  $q_0$ . It turns out that the result is purely real if the photon comes from incoming and outgoing lines, but it also has an extra imaginary piece when both lines are incoming or both are outgoing. This phase is the relativistic equivalent of an infinite phase factor that also appears in the Schrodinger wave function when scattering off the Coulomb potential.

Aside from that phase, which drops out of all expectation values, we find

$$\mathcal{R}e J_{mn} = \frac{2\pi^2}{\beta_{mn}} \log \frac{1 + \beta_{mn}}{1 - \beta_{mn}} \log \frac{\Lambda}{\lambda} \quad (3.7.20)$$

where we define  $\beta_{mn}$  by

$$\beta_{mn} = \sqrt{1 - \frac{m_n^2 m_m^2}{(p_n \cdot p_m)^2}} \quad (3.7.21)$$

This means that the overall rate from  $|M|^2$  will be

$$|M^\lambda|^2 = |M^\Lambda|^2 \left( \frac{\lambda}{\Lambda} \right)^{A(p_i)} \quad (3.7.22)$$

with

$$A(p_i) = -\frac{1}{8\pi^2} \sum_{nm} \frac{Q_n Q_m \eta_n \eta_m}{\beta_{nm}} \log \frac{1 + \beta_{mn}}{1 - \beta_{mn}} \quad (3.7.23)$$

where it's important to note that the  $Q_n$  must be accompanied by a sign  $\eta_n$  associated with whether the particles are incoming or outgoing.

The exponent  $A(p_i)$  is always positive for any combination of external states. This means that including virtual corrections has the effect of making the amplitude *vanish* as we take  $\lambda \rightarrow 0$  to try to remove the infrared regulator.

Note that if we had not resummed all virtual soft photons, we would have encountered them as a logarithmic divergence in fixed order Feynman diagrams. We only obtained a power-law by exponentiating, ie summing all possible Feynman diagrams.

### 3.7.3 Real Soft Photons and Cancellation of Divergences

With only the contributions from virtual soft photons, we would seem to predict that all scattering amplitudes vanish. To see that this is not the case, we need to include real emission.

The resolution, which we have hinted at several times before, is that one cannot (in general) measure the cross section for a definite number of particles in theories with massless particles. Photons with very low energy can always escape undetected. What can be measured is the total rate  $\Gamma(E, E_T)$  for a reaction to take place with no unobserved photon with energy greater than  $E$ , and no total amount of ‘missing’ energy  $E_T$  (one can use energy conservation of the hard particles to fix  $E_T$ , and for simplicity then set  $E = E_T$ ).

The S-Matrix for emitting  $N$  soft photons can be obtained by contracting our soft factors with  $\epsilon_\mu(q_i)$  for each soft particle. The differential rate for emitting  $N$  soft photons into a volume  $\prod_r d^3 q_r$  of momentum space is given by squaring the matrix element, summing over helicities, and multiplying by  $\prod_r d^3 q_r$ . Note that

$$\sum_h \epsilon^\mu(q, h) \epsilon^{\nu*}(q, h) = \eta_{\mu\nu} + q_\mu c_\nu + q_\nu c_\mu \quad (3.7.24)$$

but the  $q_\mu$  terms drop out due to charge conservation. Thus we get a differential rate

$$d\Gamma(q_1, \dots, q_N) = \Gamma \prod_r^N \frac{d^3 \vec{q}_r}{(2\pi)^3 2\omega_{q_r}} \sum_{n,m} \frac{e_n e_m p_n \cdot p_m}{(p_n \cdot q_r)(p_m \cdot q_r)} \quad (3.7.25)$$

It's worth noting that for  $N = 1$  this corresponds to the distribution of energy emitted by a classical discontinuous current four vector

$$J^\mu(x) = \sum_n \delta^3(\vec{x} - \vec{v}_n t) \frac{e_n p_n^\mu}{E_n} \quad (3.7.26)$$

This interpretation accords with our world-line derivation of the soft amplitude itself, where we viewed the world-line of each hard particle as a classical source.

To actually calculate the differential rate for the emission of  $N$  soft photons, where we only fix their energies  $E_r = |\vec{q}_r|$ , we need to integrate over the  $\vec{q}_r$ . These are the same as those we encountered for virtual photons, and give

$$-\pi(p_n \cdot p_m) \int \frac{d^2 \hat{q}}{(E_n - \hat{q} \cdot \vec{p}_n)(E_m - \hat{q} \cdot \vec{p}_m)} = \frac{2\pi^2}{\beta_{mn}} \log \left( \frac{1 + \beta_{mn}}{1 - \beta_{mn}} \right) \quad (3.7.27)$$

Thus we obtain the differential rate

$$d\Gamma(\omega_1, \dots, \omega_n) = \Gamma_0 \frac{[A(p_i)]^N}{N!} \frac{d\omega_1}{\omega_1} \dots \frac{d\omega_N}{\omega_N} \quad (3.7.28)$$

with

$$A(p_i) = -\frac{e^2}{8\pi^2} \sum_{nm} \frac{Q_n Q_m \eta_n \eta_m}{\beta_{nm}} \log \frac{1 + \beta_{mn}}{1 - \beta_{mn}} \quad (3.7.29)$$

where we defined

$$\beta_{mn} = \sqrt{1 - \frac{m_n^2 m_m^2}{(p_n \cdot p_m)^2}} \quad (3.7.30)$$

An unrestricted integral over the  $\omega_r$  would give a divergence, but of course we must use the same IR regulators from our discussion of virtual soft photons (because of unitarity of the S-Matrix, or intuitive obviousness).

Now we need to sum over all  $N$ , while integrating over the  $\omega_r$ . One can be very careful (see Weinberg) and implement separate limits on the total soft photon energy and on the energy of each soft photon, but if we set these equal to  $E_T$ , we find that

$$\Gamma_{real}(p_i; E_T) \rightarrow \left( \frac{E_T}{\lambda} \right)^{A(p_i)} \Gamma_0(p_i) \quad (3.7.31)$$

This was reasonable because the dependence on total energy vs each soft photon energy is weak, as can be explicitly checked. Including the exact formula only changes the result to higher order in perturbation theory. The intuition is that adding more photons costs factors of the coupling constant, so if we just cap the energy of each photon we do OK, because at weak coupling the theory does not want to emit an infinite number of finite energy photons.

Because  $A(p_i) > 0$  this becomes infinite in the limit that  $\lambda \rightarrow 0$  and we remove the IR cutoff. However, we found an opposite dependence from virtual photons, so account for both we see that

$$\Gamma_{full}(p_i; E_T) \rightarrow \left( \frac{E_T}{\Lambda} \right)^{A(p_i)} \Gamma_0(p_i) \quad (3.7.32)$$

and now we can send  $\lambda \rightarrow 0$ , recovering a finite answer. Recall that  $\Lambda$  was just a convenient factor that we introduced to divide ‘hard’ photons from ‘soft’ photons.

We can use our analysis to ask a question that can be easily phrased and answered – what is the probability to produce  $n$  soft photons with energies between  $E_-$  and  $E_+$ ?

In this case we can just take  $E_+ = \Lambda$ , our division between soft and hard photons. But we want to demand that there are exactly  $n$  soft photons emitted in the  $(E_-, E_+)$  energy range, so we need to make sure that no more than  $n$  photons are produced. And we are basically indifferent to what happens below  $E_-$ .

Including the  $n$  photons means that we need to do  $n$  of the  $d\omega/\omega$  integrals in the required range, yielding

$$\frac{1}{n!} A(p_i)^n \left[ \log \frac{E_+}{E_-} \right]^n \exp \left[ -A(p_i)^n \log \frac{E_+}{E_-} \right] \quad (3.7.33)$$

This is a Poisson distribution

$$P(n) = \frac{1}{n!} \lambda^n e^{-\lambda} \quad (3.7.34)$$

with  $\lambda = A(p_i) \log \frac{E_+}{E_-}$ . You should look at chapter 6 of Peskin’s book for a purely classical guesstimate that matches this result.

### 3.7.4 General IR Divergences and Jets

QED is special – one can obtain even more general IR divergences. In fact, if you look at our analysis for QED, in the limit that  $m_n \rightarrow 0$ , so for instance if we take the electron mass to zero, then we get new logarithmic divergences, because  $\beta_{mn} \rightarrow 1$ .

In general, in any theory where multiple massless particles interact we get more divergences. It’s worth noting two different kinds of kinematical configuration that can exist, *soft* and *collinear* divergences. ‘Soft’ means that the momentum  $|\vec{q}| \rightarrow 0$ , but collinear divergences can occur even when  $|\vec{q}|$  is finite, whenever it is parallel to the initial momentum. Note that if the initial particle has a mass, then even if it emits a massless particle, there can never be a true collinear divergence, because we can always go to a frame where the initial particle is at rest. However, *a massless initial particle can ‘split’* via

$$q \rightarrow xq + (1-x)q \quad (3.7.35)$$

for any  $0 < x < 1$ . This is the collinear limit, and it can contribute new IR divergences in theories with vertices where 3 or more massless particles interact with each other.

These combined soft and collinear divergences cannot be eliminated by summing over final states. However, they can be eliminated if we sum over *both initial and final states*. The idea is that the interactions that lead to soft and collinear divergences are exactly those that turn ‘free particles’ into ‘in’ and ‘out’ states. These interactions are just some unitary transformation on the hard states plus soft radiation and collinear splittings/radiation. So if we sum over all the (nearly identical) states in these subspaces, all the soft divergences cancel. See Weinberg chapter 13.4 for a detailed formal argument.

There’s a physical point to this reasoning – namely that in theories with massless particles, *we should always only talk about the scattering of “jets”, or collimated beams of particles*. Because non-abelian gauge theories (e.g. QCD) have interacting massless particles, this means that we must talk in terms of jets when we study hadron colliders.

But why didn’t we have to sum over soft photon initial states? Weinberg gives a formal argument (which doesn’t seem very helpful to me), but I think the physical point is that collinear divergences differ from soft divergences in that, if I have a massive particle with 4-momentum  $p$  which is on-shell, it cannot absorb an on-shell photon and transition to a new on-shell state. If  $p^2 = m^2$  and  $q^2 = 0$ , then the new state must satisfy  $p \cdot q = 0$ , and in the rest frame of the massive particle, this means  $q_0 = 0$ , so the soft particle has exactly zero energy.

This means that (roughly speaking) if we take the initial state to consist of only hard particles and soft photons, the soft photons cannot interact with the hard particles before the scattering process. In contrast, a massless particle with momentum  $q$  can transition to two particles via  $q \rightarrow xq + (1-x)q$  while all particles stay exactly on-shell. This forces us to sum over these combinations of initial states, since these processes can occur into the infinite past. In summary, collinear vs soft phase spaces are fundamentally different.

### 3.7.5 The External Field Approximation

You might wonder what approximation we need to do to replace a heavy charged particle by its classical electromagnetic field. The purpose of this section is to derive that approximation.

The physical idea is that if we emit a relatively soft (virtual or real) photon with momentum  $q$  from a heavy particle, then since  $p^2 = (p+q)^2 = m^2$  we must have  $p \cdot q = 0$ . In more familiar terms, if we boost to the rest frame of the heavy particle, then the shift in energy  $q_0 = \vec{q}^2/M \sim 0$  which must be very small – a soft photon cannot change the energy of a heavy particle.

Consider some Feynman diagram where a heavy charged particle exists in the initial and final state, and it emits and absorbs some number  $N$  of soft off-shell (virtual) photons. Then the part of the Feynman diagram that connects to this heavy particle is

$$M_{\sigma, \sigma'}^{\mu_1 \cdots \mu_N}(q_1, \cdots, q_N; p) \quad (3.7.36)$$

where  $q_r$  are the 4-momenta of the soft photons. This amplitude will obviously be dominated by

$$M_{\sigma, \sigma'}^{\mu_1 \cdots \mu_N}(q_1, \cdots, q_N; p) = \frac{\delta^4(p' - p + q_1 + \cdots + q_N) G_{\sigma', \sigma_1}^{\mu_1} G_{\sigma_1, \sigma_2}^{\mu_2} \cdots G_{\sigma_{N-1}, \sigma}^{\mu_N}}{(2p \cdot q_1 - i\epsilon)(2p \cdot (q_1 + q_2) - i\epsilon) \cdots (2p \cdot (q_1 + \cdots + q_{N-1}) - i\epsilon)} + \text{perms}$$

where we sum over permtations of the order of the photons, and the vertex

$$\frac{G_{\sigma,\sigma'}^\mu}{2p_0} = \langle \vec{p}, \sigma' | J^\mu(0) | \vec{p}, \sigma \rangle \quad (3.7.37)$$

and the  $\sigma_i$  just label spin states of the heavy particle. Actually these labels are completely irrelevant, because in the soft limit

$$G_{\sigma,\sigma'}^\mu = 2Ze p^\mu \delta_{\sigma,\sigma'} \quad (3.7.38)$$

as we have already discussed. So actually the spin of the heavy particle drops out.

Now two nice things happen. First of all, to leading order in  $q_r$  we can write

$$\delta^4(p' - p + q_1 + \cdots q_N) = p^0 \delta^3(\vec{p}' - \vec{p} + \vec{q}_1 + \cdots \vec{q}_N) \delta(p \cdot (q_1 + \cdots + q_N)) \quad (3.7.39)$$

Second of all, on the support of the last delta function, we can write

$$\frac{\delta(p \cdot (q_1 + \cdots + q_N))}{(2p \cdot q_1 - i\epsilon)(2p \cdot (q_1 + q_2) - i\epsilon) \cdots (2p \cdot (q_1 + \cdots + q_{N-1}) - i\epsilon)} + \text{ perms} \quad (3.7.40)$$

$$= (2\pi i)^{N-1} \delta(p \cdot q_1) \delta(p \cdot q_2) \cdots \delta(p \cdot q_{N-1}) \quad (3.7.41)$$

This can be most easily understood as the Fourier transform of

$$\theta(t_1 - t_2) \theta(t_2 - t_3) \cdots \theta(t_{N-1} - t_N) + \text{ perms} = 1 \quad (3.7.42)$$

although for  $N = 2$  one can also see it immediately by writing  $\frac{1}{x+i\epsilon}$  in terms of a principal part and delta function piece.

Thus we find that

$$M_{\sigma,\sigma'}^{\mu_1 \cdots \mu_N} = (2\pi Ze)^N \delta_{\sigma,\sigma'} p^{\mu_1} \cdots p^{\mu_N} \delta^3(\vec{p}' - \vec{p} + \vec{q}_1 + \cdots \vec{q}_N) \delta(p \cdot q_1) \delta(p \cdot q_2) \cdots \delta(p \cdot q_{N-1})$$

This applies to both relativistic and non-relativistic heavy particles (as it's a nice Lorentz covariant answer). In the NR limit we just get  $\delta(q_r^0)$  factors, which tell us that the virtual photons carry momentum but no energy, as suggested by our rough argument above.

This derivation shows that the presence of this heavy particle (taken at rest for simplicity) is equivalent to inserting any number  $N$  of virtual photons sourced by

$$\int d^3 X |\psi(X)|^2 \prod_{r=1}^N (2\pi Ze) n^{\mu_r} \delta(q_R^0) e^{-i\vec{q}_r \cdot X} \quad (3.7.43)$$

where  $\psi$  is the heavy particle wavefunction and  $n^\mu \propto p^\mu$  of the heavy particle. But this is exactly what we'd get from a background  $\mathcal{A}_\mu$  field with

$$\begin{aligned} \mathcal{A}_\mu(x) &= \int \frac{d^4 q}{(2\pi)^4} \frac{(2\pi Ze) n^{\mu_r} \delta(q_R^0) e^{-i\vec{q}_r \cdot X}}{q^2 - i\epsilon} \\ &= \frac{Ze \delta^{0,\mu}}{4\pi |\vec{x} - \vec{X}|} \end{aligned} \quad (3.7.44)$$

which just couples like a classical source  $\mathcal{A}_\mu(x)J^\mu(x)$  to the electric current operator of the light charged particles.

Here we have summed all photon exchanges between the heavy source and the other light particles. If *both* the heavy and light particles are non-relativistic then a further simplification occurs, and one only needs ‘uncrossed ladder diagrams’.

### 3.8 Implications of Unitarity

By studying long distance physics and IR divergences, we saw that there is an intimate connection between loops corrections and multiple real emissions. This is no accident – it is a consequence of unitarity, or (roughly speaking) the statement that probabilities must add up to 1.

At a formal level, the unitarity of the S-Matrix simply says that the  $S$  operator satisfies

$$S^\dagger S = \mathbf{1} \quad (3.8.1)$$

However, we usually write  $S$  by extracting a trivial part involving no scattering, so that

$$S = 1 + i\mathcal{T} \quad (3.8.2)$$

where we write

$$\langle f|\mathcal{T}|i\rangle = (2\pi)^4\delta^4(p_f - p_i)\mathcal{M}(i \rightarrow f) \quad (3.8.3)$$

in schematic form. Then unitarity implies that

$$i(\mathcal{T}^\dagger - \mathcal{T}) = \mathcal{T}^\dagger \mathcal{T} \quad (3.8.4)$$

directly about the non-trivial part of the S-Matrix.

Now on the LHS we can sandwich this between  $\langle f|$  and  $|i\rangle$ , giving

$$\begin{aligned} i(\langle f|\mathcal{T}^\dagger|i\rangle - \langle f|\mathcal{T}|i\rangle) &= i(\langle i|\mathcal{T}^\dagger|f\rangle^* - \langle f|\mathcal{T}|i\rangle) \\ &= i(2\pi)^4\delta^4(p_f - p_i)(\mathcal{M}^*(f \rightarrow i) - \mathcal{M}(i \rightarrow f)) \end{aligned} \quad (3.8.5)$$

Similarly, on the right side we do the same thing, but we also insert a sum over all states in the form

$$\mathbf{1} = \int dX |X\rangle\langle X| \quad (3.8.6)$$

which gives

$$\langle f|\mathcal{T}^\dagger\mathcal{T}|i\rangle = (2\pi)^4\delta^4(p_f - p_X)(2\pi)^4\delta^4(p_i - p_X) \int dX \mathcal{M}(i \rightarrow X)\mathcal{M}^*(f \rightarrow X) \quad (3.8.7)$$

whwere we are summing and integrating over all possible intermediate states  $X$ , composed of any number of particles with arbitrary spins and momenta. This general statement shows how unitarity



constrains the scattering amplitudes. Note that this must hold *to each order in perturbation theory*, so it relates various tree and loop diagrams to other loop diagrams.

An important special case is when  $|f\rangle = |i\rangle = |A\rangle$ , in which case we see that

$$2\mathcal{I}m\mathcal{M}(A \rightarrow A) = \int dX (2\pi)^4 \delta^4(p_A - p_X) |\mathcal{M}(A \rightarrow X)|^2 \quad (3.8.8)$$

where  $dX$  must be lorentz invariant, so

$$dX = \prod_r \frac{d^3q_r}{(2\pi)^3 2E_r} \quad (3.8.9)$$

for each of the particles in the  $X$  state. This tells us that for one particle states, the decay rate satisfies (recall that decay rates have a factor of  $1/m_A$ )

$$\mathcal{I}m\mathcal{M}(A \rightarrow A) = m_A \sum_X \Gamma(A \rightarrow X) = m_A \Gamma_{tot} \quad (3.8.10)$$

which means that the imaginary part of the exact propagator is directly related to the decay rate of the particle.

The case with a two particle state is what's often called the *optical theorem*. It says that in the center of mass frame

$$\mathcal{I}m\mathcal{M}(A \rightarrow A) = 2E_{CoM} |\vec{p}_i| \int dX \sigma(A \rightarrow X) \quad (3.8.11)$$

so the imaginary part of the *forward* scattering amplitude is given in terms of the total scattering cross section.

All these statements of unitarity should make intuitive sense in the following respect: if the total transition probability is 1, then the square of the off-diagonal elements of the matrix must be compensated by depletions on the diagonal. What we have derived is simply the fancy QFT version of that statement about unitary matrices.

### 3.8.1 Decay Rate Example

Let's consider a theory with an interaction

$$\frac{\lambda}{2} \phi \pi^2 \quad (3.8.12)$$

Clearly if  $m_\phi > 2m_\pi$  then we can have a decay, so we should have

$$\mathcal{I}m\mathcal{M}(\phi \rightarrow \phi) = m_\phi \Gamma(\phi \rightarrow \pi\pi) \quad (3.8.13)$$

but only when the  $m_\phi$  is large enough.

The one-loop correction to the propoagator is

$$\mathcal{M}_{1-loop}(p) = -\frac{\lambda^2}{32\pi^2} \int_0^1 dx \log \left( \frac{m^2 - p^2 x(1-x) - i\epsilon}{\Lambda^2} \right) \quad (3.8.14)$$

where  $\Lambda$  is the UV cutoff. We put  $\phi$  on-shell by setting  $p^2 = M^2$  giving

$$\mathcal{M}_{1-loop}(M) = -\frac{\lambda^2}{32\pi^2} \int_0^1 dx \log \left( \frac{m_\pi^2 - m_\phi^2 x(1-x) - i\epsilon}{\Lambda^2} \right) \quad (3.8.15)$$

Note that  $x(1-x) \leq 1/4$ , so for  $m_\phi < 2m_\pi$  the result is purely real. However for larger masses we note

$$\log(-A - i\epsilon) = \log A - i\pi \quad (3.8.16)$$

so that we have

$$\begin{aligned} \mathcal{Im} \mathcal{M}_{1-loop}(M) &= \frac{\lambda^2}{32\pi} \int_0^1 dx \Theta(m_\phi^2 x(1-x) - m_\pi^2) \\ &= \frac{\lambda^2}{32\pi} \sqrt{1 - 4\frac{m^2}{M^2}} \Theta(m_\phi - 2m_\pi) \end{aligned} \quad (3.8.17)$$

The decay rate is

$$\Gamma = \frac{1}{2} \int \frac{1}{2m_\phi} |\lambda|^2 \frac{|\vec{p}_f|}{m_\phi} \frac{d\Omega}{16\pi} \quad (3.8.18)$$

where  $\vec{p}_f$  is the momentum of one of the final state  $\pi$  particles. Thus we get

$$\Gamma = \frac{\lambda^2}{32\pi m_\phi} \sqrt{1 - 4\frac{m^2}{M^2}} \Theta(m_\phi - 2m_\pi) \quad (3.8.19)$$

verifying the optical theorem.

### 3.8.2 Cutting Rules

Now let's see how the unitarity relation is manifested directly in the Feynman diagrams.

At a computational level, the main point is that the Feynman propagator

$$\frac{1}{p^2 - m^2 + i\epsilon} = \text{PP} \left( \frac{1}{p^2 - m^2} \right) - i\pi \delta(p^2 - m^2) \quad (3.8.20)$$

where the PP means ‘principal part’, it is defined to be  $1/(p^2 - m^2)$  for  $p \neq m$  and is zero otherwise. The way to see this is really to just write

$$\frac{1}{p^2 - m^2 + i\epsilon} = \frac{p^2 - m^2}{(p^2 - m^2)^2 + \epsilon^2} - \frac{i\epsilon}{(p^2 - m^2)^2 + \epsilon^2} \quad (3.8.21)$$

By looking at integrals over  $p^2$  one can extract the claimed result.

This immediately shows that for tree level diagrams, the imaginary part comes from (the isolated point where) an intermediate particle goes on-shell. It turns out that this generalizes to loops, and is equivalent to the statement of the optical theorem.

The physical point is that *imaginary parts of loop amplitudes arise when intermediate particles go on-shell*. There is a more complex-analysis friendly way of seeing this, by looking at a loop integral (or any integral) and asking when it can have a *discontinuity* as we move the kinematic invariants around. This can only happen when we force the contour of integration to wind around a pole.

Why care about discontinuities? Let us consider our amplitude  $M(s)$ . If  $s$  is real and below the threshold for any states in the theory, then  $M(s)$  will be real, giving the trivial identity

$$M(s) = M^*(s^*) \quad (3.8.22)$$

Now both of these are analytic functions of  $s$ , so we can continue both of them to  $s >$  the masses of various multi-particle states. Near the real axis we have

$$\begin{aligned} \mathcal{R}e M(s + i\epsilon) &= \mathcal{R}e M(s - i\epsilon) \\ \mathcal{I}m M(s + i\epsilon) &= -\mathcal{I}m M(s - i\epsilon) \end{aligned} \quad (3.8.23)$$

Above the multi-particle threshold, we will have a branch cut on the real  $s$ -axis, and so we see that *the discontinuity across the branch cut is*

$$\text{Disc } M(s) = 2i\mathcal{I}m M(s + i\epsilon) \quad (3.8.24)$$

Thus we can compute the imaginary parts by computing a discontinuity across the real axis. That discontinuity can only arise when we pick up little contour integrals around the poles of the integrand by analytically continuing in e.g.  $s$ .

So we can find the imaginary part by analytically continuing  $s$ . For example with  $s = 4p^2$  in an integral like

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{[(p+k)^2 - m^2 + i\epsilon][(p-k)^2 - m^2 + i\epsilon]} \quad (3.8.25)$$

we can fix  $p = e^{i\theta}(E, 0, 0, 0)$  and continue  $\theta \rightarrow \theta + 2\pi$ . Then we can write the integral as

$$\int \frac{dk_0 d^3\vec{k}}{(2\pi)^4} \frac{1}{[(k_0 + E)^2 - \vec{k}^2 - m^2][(k_0 - E)^2 - \vec{k}^2 - m^2]} \quad (3.8.26)$$

So the poles are at

$$E = \pm k_0 \pm' \sqrt{\vec{k}^2 + m^2} \quad (3.8.27)$$

As we analytically continue in  $E$ , we need to keep track of how we move the contour in  $k_0$  and  $\vec{k}$  to avoid the pole. We can always move  $\vec{k}$  so that  $k_0$  does not meet one of the poles, or vice versa.

But if  $E > m$  then we can get caught in such a way that we must move both contours. As  $E$  moves in the complex plane, we pick up a discontinuity from the region where both propagators vanish, so we have a pole from both. But this just gives a phase space integral over a product of two tree-level diagrams, confirming the relation

$$i [\mathcal{M}^*(f \rightarrow i) - \mathcal{M}(i \rightarrow f)] = (2\pi)^4 \delta^4(p_i - p_X) \int dX \mathcal{M}(i \rightarrow X) \mathcal{M}^*(f \rightarrow X) \quad (3.8.28)$$

In general, there are a set of ‘cutting rules’ (written out in the book) that tell how to compute a general discontinuity, by replacing all combinations of propagators with delta functions and summing.

But this simply amounts to the statement of the unitarity relation, namely that we get an imaginary part (or a discontinuity) by taking the product of all possible  $\mathcal{M}(i \rightarrow X) \mathcal{M}^*(f \rightarrow X)$  and summing/integrating over the intermediate states.

### 3.8.3 Comment on Spin Sums and Propagators

Loop amplitudes have propagator factors with non-trivial numerators in the case of particles with spin. This occurs on the LHS of

$$i [\mathcal{M}^*(f \rightarrow i) - \mathcal{M}(i \rightarrow f)] = (2\pi)^4 \delta^4(p_i - p_X) \int dX \mathcal{M}(i \rightarrow X) \mathcal{M}^*(f \rightarrow X) \quad (3.8.29)$$

However, on the RHS we have products of amplitudes with external particles in the state  $X$ . Thus we have spinors and/or polarization vectors. So the optical theorem implies that sums over spins have to be the same thing as propagator numerators.

This provides the bazillionth way of seeing that massless particles with spin 1 or greater must have a notion of gauge invariance. Because, in fact, their spin sums are ambiguous, but the ambiguities cancel due to the Ward identity. The end result is that we can always write

$$\Pi_s(p) = \frac{\sum_j \epsilon_j \epsilon_j^*}{p^2 - m^2 + i\epsilon} \quad (3.8.30)$$

where  $\epsilon_j$  are a basis of physical polarizations for the particle of spin  $s$ .

### 3.8.4 Bounds on the Cross Section

Unitarity gives various bounds on the total cross section. For example if we work in an angular momentum basis and we have massless scalar particles, then

$$M(E, \theta) = \sum_j a_j(E) P_j(\cos \theta) \quad (3.8.31)$$

where  $E$  is the CoM energy. But the optical theorem immediately says that

$$\text{Im}(a_j(E)) \geq \frac{1}{2} |a_j(E)|^2 \quad (3.8.32)$$

for these massless particles. Here the optical theorem is extremely simple because the S-Matrix really just is a discrete matrix  $(1 + ia_j)$ , because we diagonalized angular momentum.

Another bound is Froissart, although it only holds when all particles have a mass. The intuition behind it is that interaction ranges fall off exponentially when all particles have a mass, so the cross section is  $\sim R^2$ , where  $R \sim \frac{1}{m} \log E$ .

The main use of these bounds is to say that *when they are violated in perturbation theory, we know for certain that perturbation theory has broken down*. No theory actually violates unitarity, it's just that we learn from extrapolated violations of unitarity that the theory must change, or perturbation theory must be abandoned, before reaching some (high energy) scale.

### 3.8.5 Unitarity and the Second Law of Thermodynamics

The unitarity relations we have derived can be used for a strikingly different purpose – to derive the 2nd law of thermodynamics. In textbooks usually the Born approximation or time reversal invariance are used, but neither is exactly true. In fact, the 2nd law follows from an analysis based on the unitarity of scattering.

Let us consider a large collection of particles in a box with volume taken (for convenience) to be  $(2\pi)^3 V$ . We will study reaction rates between the particles in the box by using scattering amplitudes and rates satisfying the unitarity relation.

Note that since both  $S^\dagger S = 1$  and  $SS^\dagger = 1$ , we can use the result

$$\text{Im } M(A \rightarrow A) = -\pi \int dB \delta^4(p_A - A_B) |M(A \rightarrow B)|^2 \quad (3.8.33)$$

to write the reciprocity relation

$$\int dB \delta^4(p_A - p_B) |M(A \rightarrow B)|^2 = \int dB \delta^4(p_A - p_B) |M(B \rightarrow A)|^2 \quad (3.8.34)$$

Writing differential rates in a very formal way, we can say therefore that

$$V^{N_A} \int dB \frac{d\Gamma(A \rightarrow B)}{dB} = V^{N_B} \int dB \frac{d\Gamma(B \rightarrow A)}{dA} \quad (3.8.35)$$

as a relation between rates.

Let us also define  $P(A)dA$  as the probability of finding the system in a volume  $dA$  of the multi-particle state phase space. So  $P(A)$  has units of  $V^{N_A}$  where  $N_A$  is the number of particles in the state  $A$ . This follows because each particle has a differential phase space  $d^3\vec{q}$ .

Then the rate of change of  $P(A)$  is

$$\frac{dP(A)}{dt} = \int dB P(B) \frac{d\Gamma(B \rightarrow A)}{dA} - P(A) \int dB \frac{d\Gamma(A \rightarrow B)}{dB} \quad (3.8.36)$$

This just says that the change in  $P(A)$  is given by the rate at which it's increased by scattering minus the rate at which it's depleted. We can see immediately that  $\int P(A)dA$ , the total phase space volume, will be time independent.

Now let us consider the rate of change of the entropy

$$S = - \int dA P(A) \log \left( \frac{P(A)}{V^{N_A}} \right) \quad (3.8.37)$$

It is

$$\frac{dS}{dt} = - \int dA \int dB \left( \log \left( \frac{P(A)}{V^{N_A}} \right) + 1 \right) \left[ P(B) \frac{d\Gamma(B \rightarrow A)}{dA} - P(A) \frac{d\Gamma(A \rightarrow B)}{dB} \right] \quad (3.8.38)$$

The +1 simply vanishes. Interchanging integration labels in the second term we find

$$\frac{dS}{dt} = \int dA \int dB P(B) \log \left( \frac{P(B)V^{N_A}}{P(A)V^{N_B}} \right) \frac{d\Gamma(B \rightarrow A)}{dA} \quad (3.8.39)$$

Now for any positive quantities, the function  $y \log(y/x)$  is convex, so it satisfies

$$y \log \frac{y}{x} \geq y - x \quad (3.8.40)$$

So we can bound the rate of change of the entropy by

$$\begin{aligned} \frac{dS}{dt} &\geq \int dA dB \left( \frac{P(B)}{V^{N_B}} - \frac{P(A)}{V^{N_A}} \right) V^{N_B} \frac{d\Gamma(B \rightarrow A)}{dA} \\ &= \int dA dB \left( \frac{P(B)}{V^{N_B}} \right) \left( V^{N_B} \frac{d\Gamma(B \rightarrow A)}{dA} - V^{N_A} \frac{d\Gamma(A \rightarrow B)}{dB} \right) \end{aligned} \quad (3.8.41)$$

But now the unitarity relation tells us that once we integrate this vanishes, so we conclude

$$\frac{dS}{dt} \geq 0 \quad (3.8.42)$$

We have derived the 2nd law of thermodynamics from unitarity and kinetic theory.

### 3.9 Interlude on Lie Groups and Lie Algebras

We have seen a few examples of Lie Groups, such as the  $SO(1,3)$  Lorentz group, and the  $SU(2) \sim SO(3)$  group of rotations. I also just mentioned  $SO(N)$  global symmetries. Now let us discuss Lie Groups and Lie Algebras a bit more systematically, so that we can use them in discussions of symmetry and dynamics.

#### 3.9.1 Definitions

Lie Groups are both groups and differentiable manifolds. Any group element continuously connected to the identity can be written

$$U = \exp[i\theta_a T^a] \quad (3.9.1)$$

where the  $\theta_a$  parameterize the group and the  $T^a$  are the group generators, which live in the Lie Algebra. We conventionally view  $\theta_a$  as real parameters. If you have an explicit form for  $U$ , say as matrices, then you can figure out  $T^a$  by series expansion.

The generators of  $T^a$ , which generate infinitesimal group transformations, form the Lie Algebra (distinct from the group). The Lie algebra is defined by its commutation relations

$$[T^a, T^b] = i f^{abc} T_c \quad (3.9.2)$$

where  $f^{abc}$  are known as the structure constants. By definition  $f^{abc} = -f^{bac}$ . Note that this ‘commutator’ (also just called a ‘Lie Bracket’) is really an abstract bracket at this point, ie it does not necessarily correspond to the operation  $AB - BA$  on matrices. It only becomes  $AB - BA$  once we have a representation for the Lie Algebra, ie a set of matrices satisfying the algebra.

This explains why we demand that the Lie bracket satisfies the Jacobi identity

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \quad (3.9.3)$$

This identity is trivial once we use  $[A, B] = AB - BA$ , as you can check, but it’s a non-trivial demand to impose on the abstract Lie bracket. The Jacobi identity can be written as

$$f^{abd} f^{dce} + f^{bcd} f^{dae} + f^{cad} f^{dbe} = 0 \quad (3.9.4)$$

using the structure constants.

Lie algebras that we will study are either simple or semi-simple, meaning that they are direct sums of simple algebras. Simple algebras are those that do not contain sub-algebras that commute with all the elements in the algebra (no non-trivial ideals); the example important for physics are mostly  $su(N)$  and  $so(N)$ , which are based on groups that preserve either a complex inner product or a real inner product. The reason we care is that the finite dimensional representations of semi-simple Lie algebras are always Hermitian, so one can find Unitary QM theories based on such algebras.

The complex inner product is

$$U^\dagger U = 1 \quad (3.9.5)$$

defined on  $N$  dimensional complex vector spaces, for  $U(N)$ . Note that in all cases we can write  $U(N) = SU(N) \times U(1)$  where the  $U(1)$  represents an overall phase. There are  $N^2 - 1$  generators for  $SU(N)$ . To see this, let us write the identity infinitesimally as

$$-i\theta_a (T^\dagger)^a + i\theta_a T^a = 0 \implies T = T^\dagger \quad (3.9.6)$$

so we can count the generators by counting  $N \times N$  Hermitian matrices. Such matrices have  $\frac{1}{2}N(N-1)$  imaginary components and  $\frac{1}{2}N(N+1)$  real components, but then we subtract the identity matrix, which just generates  $U(1)$ . In contrast, the real inner product is

$$O^T O = 1 \quad (3.9.7)$$

and there are  $\frac{1}{2}N(N-1)$  generators, because they must satisfy

$$T = -T^T \quad (3.9.8)$$

and thus be anti-symmetric  $N \times N$  matrices. Another group, which we use less often, is the symplectic group  $Sp(2N)$ , which preserves a quaternionic inner product, or equivalently

$$\Omega S = -S^T \Omega \quad (3.9.9)$$

where  $\Omega = i\epsilon_{ab}$  is a  $2 \times 2$  matrix. There are also some exceptional groups,  $G_2, F_4, E_6, E_7, E_8$ .

### 3.9.2 Representations

The groups and algebras discussed above are abstract mathematical objects. We want to have these groups act on quantum states and fields, which are vectors, so we need to *represent* the groups as matrices. There are an infinite number of different representations for a given simple group, as you know from the case of  $SU(2)$ , which gives the spin and rotation representations.

However, there are two obvious and most important representations, which occur most often in physics settings. They are the *fundamental* and *adjoint* representations. The fundamental is the ‘obvious’ representation defining  $SU(N)$  and  $SO(N)$  as  $N \times N$  matrices acting on  $N$  dimensional vectors. It is basically the smallest representation of the algebra, except for cases like  $su(2) \sim so(3)$  and  $su(4) \sim so(6)$ . To write the fundamental formally, we say that  $N$  fields transform under it as

$$\phi_i \rightarrow \phi_i + i\alpha_a (T_f^a)_i^j \phi_j \quad (3.9.10)$$

where  $i = 1, \dots, N$  and the  $\alpha_a$  are real numbers. The complex conjugate fields transform in the anti-fundamental, which is just the conjugate of this

$$\phi_i^* \rightarrow \phi_i^* - i\alpha_a (T_f^a)_i^j \phi_j^* \quad (3.9.11)$$

Note that since  $T_f^a$  are Hermitian, we have  $T_{anti-f} = T_f^*$ , and other trivial relations. In the context of physics the  $i$  labels are often called ‘color’ or ‘flavor’ indices, due to the standard model context.

It’s easy to work out the algebras by just looking at the commutators of Hermitian matrices, or orthogonal matrices in the case of  $so(N)$ . You can check that  $f^{abc} = \epsilon^{abc}$  for  $su(2)$ , as should be familiar. There’s a standard basis used by physicists for  $su(3)$ , called the Gell-Mann matrices.

The normalization of generators is arbitrary and is usually chosen so that

$$f^{acd} f^{bcd} = N \delta^{ab} \quad (3.9.12)$$

but once we choose that, the normalization in all representations is fixed. This implies that in the fundamental representation

$$\text{Tr} (T_f^a T_f^b) = \frac{1}{2} \delta^{ab} \quad (3.9.13)$$

but this relation certainly depends on the representation! It doesn’t even make sense in the abstract, since multiplying generators is not well-defined without a representation.

The other ‘obvious’ representation is the adjoint. The point is to think of the generators themselves as the vectors. So for example in  $su(N)$ , we have  $N^2 - 1$  generators labeled by  $a = 1, \dots, N^2 - 1$ . They are acted on by the Lie bracket, or commutator, so

$$T_{adj}^b \cdot T^a \equiv [T^b, T^a] = i f^{bac} T_c \quad \implies \quad (T_{adj}^a)^{bc} = -i f^{abc} \quad (3.9.14)$$



How can we see that the  $T_{adj}$  actually satisfy the Lie algebra, and thus are really a representation? This is given immediately by the Jacobi identity re-written as

$$f^{cbd}f^{ade} - f^{abd}f^{cde} = -f^{cad}f^{dbe} \implies [T_{adj}^c, T_{adj}^a] = if^{cad}T_{adj}^d \quad (3.9.15)$$

Gauge bosons in gauge theories are always in the adjoint representation of the Lie algebra, hence its importance. It's also one of the simplest representations after the fundamental. Clearly it has dimension  $N^2 - 1$  for  $su(N)$ .

There is a generalization of the notion of  $\vec{J}^2$  in  $su(2)$  (thought of in terms of rotations), it is the Casimir operator or in this case, the quadratic Casimir

$$\sum_a T_R^a T_R^a = C_2(R) \mathbf{1} \quad (3.9.16)$$

where I wrote in the sum just to be clear. This must be proportional to the identity (when acting on a single given irreducible representation) because it commutes with all generators of the group, which follows because

$$[T_R^a T_R^a, T_R^b] = (if^{abc}T_R^c)T_R^a + T_R^a(if^{abc}T_R^c) = 0 \quad (3.9.17)$$

because of anti-symmetry of  $f^{abc}$  in  $ab$ . This is just as in the angular momentum case.

To compute the  $C_2(R)$ , its useful to define an inner product on the generators via

$$\text{Tr} [T_R^a T_R^b] = T(R) \delta^{ab} \quad (3.9.18)$$

The quantity  $T(R)$  is the index of the representation. We have that

$$T(fund) = \frac{1}{2} \quad (3.9.19)$$

and

$$T(adjoint) = N \quad (3.9.20)$$

for  $su(N)$ , given our normalizations. The index relates to the quadratic Casimir by

$$d(R)C_2(R) = T(R)d(G) \quad (3.9.21)$$

where  $d(R)$  is the dimension of the representation, and  $d(G)$  of the algebra, namely  $N^2 - 1$  for  $su(N)$ . Thus

$$C_2(fund) = \frac{N^2 - 1}{2N} \quad (3.9.22)$$

$$C_2(adj) = N \quad (3.9.23)$$

These quantities appear in calculations in Yang-Mills theories, just as the total angular momentum appears in computations involving angular momentum.

### 3.10 Overview of Lie Algebra Classification and Representations

To classify the representations of angular momentum, namely  $su(2)$ , you may recall that we

- First notice that the quadratic Casimir  $\vec{J}^2$  commutes with all the  $J_i$ , so any given irreducible representation must have the same eigenvalue for it; this is  $s(s+1)$ .
- We then note that  $J_{\pm} = J_1 \pm iJ_2$  act as raising and lowering operators for the generator  $J_3$ , in that

$$[J_3, J_{\pm}] = \pm J_{\pm} \quad (3.10.1)$$

$$[J_+, J_-] = J_3 \quad (3.10.2)$$

So we can classify representations in terms of their maximum or minimum  $J_3$  eigenvalue, a highest or lowest weight state. Then all vectors in the representation can be obtained by acting on this highest weight vector with  $J_-$ .

To understand general Lie algebras, one proceeds in a similar fashion. First we find as many commuting generators as possible (just as in QM!), classifying representations according to their eigenvalues. This means that there are many different kinds of raising and lowering operators, and so they must all fit together in a consistent way. Their self-consistency leads to the classification of possible Lie algebras.

#### 3.10.1 Weights

A maximal subset of commuting Lie algebra generators is called a *Cartan subalgebra*. We can write these Cartan generators in any representation  $R$  as

$$[H_i, H_j] = 0, \quad H_i = H_i^\dagger \quad (3.10.3)$$

where they must be Hermitian in a unitary representation. We let  $i = 1, \dots, m$  where  $m$  is called the *rank* of the algebra. Note that

$$\text{Tr} [H_i H_j] = T(R) \delta_{ij} \quad (3.10.4)$$

in our notation above, where  $T(R)$  is the index of the representation.

Now we can write any vector (this would be a specific state in a QM theory) in the representation as a linear combination of the eigenvectors

$$H_i |\mu, x, R\rangle = \mu_i |\mu, x, R\rangle \quad (3.10.5)$$

where  $x$  might include other labels that we need besides the cartan eigenvalues. These Cartan eigenvalues are called the *weights* and  $\mu_i$  is the weight vector. I will often refer to the vectors  $|\mu, x, R\rangle$  as states, since the analogy with quantum mechanics is essential and immediate.

### 3.10.2 Adjoint Representation and Roots

In any Lie algebra, we know that the adjoint representation exists. A basis for the vectors or ‘states’ in the adjoint correspond to the generators of the Lie algebra itself. Labeling such a state as

$$|X_a\rangle \quad (3.10.6)$$

we see that the generator  $X_b$  acts on this as

$$X_b|X_a\rangle = if_{bac}|X_c\rangle = |[X_b, X_a]\rangle \quad (3.10.7)$$

which is just the statement that in the adjoint representation, the Lie algebra acts by commutators.

The *roots* of a Lie algebra are the weights of the adjoint representation. It will turn out that by understanding the possible patterns for the roots we can understand all possible Lie algebras. Note that the states corresponding to the Cartan generators have zero weight vectors, because

$$H_i|H_j\rangle = |[H_i, H_j]\rangle = 0 \quad (3.10.8)$$

Futhermore, all states in the adjoint with zero weight must be in the Cartan subalgebra.

The other states in the adjoint have non-zero weight vectors  $\alpha_i$ , where  $i = 1, \dots, m$ , and they satisfy (by definition)

$$[H_i, E_\alpha] = \alpha_i E_\alpha \quad (3.10.9)$$

where we have diagonalized the Cartan subalgebra. It turns out that in the adjoint representation, the non-zero weights uniquely specify the corresponding generator, so there is no need for any other specification to tell us which generator we are talking about.

The  $E_\alpha$  generators cannot be Hermitian, because

$$[H_i, E_\alpha^\dagger] = -\alpha_i E_\alpha^\dagger \quad (3.10.10)$$

Note that this relation means that

$$E_\alpha^\dagger = E_{-\alpha} \quad (3.10.11)$$

which is reminiscent of  $J_\pm$  in  $su(2)$ . States with different weights must be orthogonal, because they have different Cartan eigenvalues.

### 3.10.3 $E_{\pm\alpha}$ as Raising and Lowering Operators Forming Many $su(2)$ s

In fact, the generators  $E_{\pm\alpha}$  act as raising or lowering operators (in any representation), in that

$$H_i E_{\pm\alpha} |\mu, R\rangle = ([H_i, E_{\pm\alpha}] + E_{\pm\alpha} H_i) |\mu, R\rangle = (\mu \pm \alpha)_i E_{\pm\alpha} |\mu, R\rangle \quad (3.10.12)$$

simply by using the commutation relations. If we specifically consider the adjoint representation, then  $E_\alpha|E_{-\alpha}\rangle$  must be in the Cartan subalgebra, so that

$$[E_\alpha, E_{-\alpha}] = \beta_j H_j \quad (3.10.13)$$

is a linear combination of Cartan generators. In fact we can compute the components of the vector  $\beta_i$  directly via

$$\begin{aligned}
\beta_i &= \langle H_i | [E_\alpha, E_{-\alpha}] \rangle \\
&= \frac{\text{Tr} (H_i [E_\alpha, E_{-\alpha}])}{T(\text{adj})} \\
&= \frac{\text{Tr} (E_{-\alpha} [H_i, E_\alpha])}{T(\text{adj})} \\
&= \alpha_i \frac{\text{Tr} (E_{-\alpha} E_\alpha)}{T(\text{adj})} \\
&= \alpha_i
\end{aligned} \tag{3.10.14}$$

which means that

$$[E_\alpha, E_{-\alpha}] = \alpha_j H_j \tag{3.10.15}$$

which is just like the  $su(2)$  commutation relation  $[J_+, J_-] = J_3$ .

For each  $E_{\pm\alpha}$  there is an  $su(2)$  sub-algebra within our general Lie algebra, where we call ‘ $E_3$ ’ the generator  $\frac{\alpha \cdot H}{\alpha^2}$  and we define  $E_\pm = E_{\pm\alpha}/|\alpha|$ . In fact, we could decompose any irreducible representation  $R$  of the Lie algebra into irreducible representations of each of these little  $su(2)$ s.

Since we already know everything there is to know about irreps of  $su(2)$ , this puts strong constraints on the larger Lie algebra. For example, we can show that the root vectors  $E_\alpha$  are unique. If not, then we can choose  $E_\alpha$  and  $E'_\alpha$  to be orthogonal under the  $\text{Tr} [T_a T_b]$  metric. But this means that

$$E^- |E'_\alpha\rangle = 0 \tag{3.10.16}$$

from the computation above. But we also must have

$$E_3 |E'_\alpha\rangle = |E'_\alpha\rangle \tag{3.10.17}$$

but this is inconsistent with the commutation relations for the  $su(2)$  subalgebra, since the first says that  $|E'_\alpha\rangle$  is a lowest weight state, while the second implies it has  $E_3 = 1$  eigenvalue, so that it’s ‘spin 1’. Via a similar argument we can show that if  $\alpha$  is a root, then there is no root with  $E_{k\alpha}$  except for  $k = \pm 1$ .

### 3.10.4 Structure of the Root Lattice

In general, for any weight  $\mu$  of any representation  $R$ , the  $E_3$  eigenvalue is

$$E_3 |\mu, x, R\rangle = \frac{\alpha \cdot \mu}{\alpha^2} |\mu, x, R\rangle \tag{3.10.18}$$

by definition of  $E_3$  as  $\alpha \cdot H/\alpha^2$ . But this means that in any Lie algebra and any representation, the quantity

$$\frac{2\alpha \cdot \mu}{\alpha^2} \tag{3.10.19}$$

must be an integer. This gives a very strong constraint on the structure of the root vectors  $\alpha$ .

Decompose the general state  $|\mu, xR\rangle$  according to our little  $su(2)$ , with a highest spin state appearing with spin  $\ell$ . Then there is some maximal non-negative integer  $p$  so that

$$(E_+)^p |\mu, xR\rangle \neq 0, \quad (E_+)^{p+1} |\mu, xR\rangle = 0 \quad (3.10.20)$$

This means that the  $E_3$  eigenvalue of the state is

$$\frac{\alpha \cdot (\mu + p\alpha)}{\alpha^2} = \frac{\alpha \cdot \mu}{\alpha^2} + p = j \quad (3.10.21)$$

Similarly, there is some  $q$  associated with lowering, so that

$$\frac{2\alpha \cdot \mu}{\alpha^2} = q - p \quad (3.10.22)$$

The power of these relations can be seen by applying them to two different roots (weights in the adjoint representation). Then we have that

$$\frac{\alpha \cdot \beta}{\alpha^2} = \frac{1}{2}(p - q), \quad \frac{\alpha \cdot \beta}{\beta^2} = \frac{1}{2}(p' - q') \quad (3.10.23)$$

Thus we can multiply the two relations to find

$$\cos^2 \theta_{\alpha\beta} = \frac{(\alpha \cdot \beta)^2}{\alpha^2 \beta^2} = \frac{(p - q)(p' - q')}{4} \quad (3.10.24)$$

where the numerator must obviously be an integer. Thus there are only four possibilities:

$$\theta_{\alpha\beta} = 90, 60 \text{ (120)}, 45 \text{ (135)}, 30 \text{ (150)} \quad (3.10.25)$$

for the angle between any pair of roots. It turns out that the roots of  $su(3)$  form a hexagon, with the two Cartan generators in the center.

Note that for any of the  $su(2)$ s in the algebra, we can flip spin  $\rightarrow$   $-\text{spin}$ , exchanging highest and lowest weight states. This reflects the root lattice in a mirror. The set of all such (combined) mirror reflections for all the various  $su(2)$ s forms the *Weyl group*, a symmetry group of the root lattice.

### 3.10.5 Simple Roots and Dynkin Diagrams

To complete the analogy with  $su(2)$ , we need a notion of ‘positive and negative spin’ so we can differentiate the highest from the lowest weights. It’s sufficient to just define a weight vector  $\mu_i$  as positive if its first component is positive, and negative otherwise. Now we have a notion of highest and lowest weight states.

We do not need to talk about all of the roots of a Lie algebra, since some can be written as sums of others. This we define the *simple roots* as the positive roots that cannot be written as a sum of other positive roots. If a given state is annihilated by all the simple roots, then it is annihilated by all the positive roots, so it is the highest weight state of an irreducible representation.

Some simple and fun to prove statements about simple roots:

- If  $\alpha$  and  $\beta$  are different simple roots, then  $\alpha - \beta$  is not a root.
- It follows that  $\frac{\alpha \cdot \beta}{\alpha^2} = -\frac{p}{2}$  and  $\frac{\alpha \cdot \beta}{\beta^2} = -\frac{p'}{2}$  so knowing  $p$  and  $p'$  tells us the angle between the roots and their relative lengths.
- The angle between pairs of simple roots is in  $[\frac{\pi}{2}, \pi)$ .
- The simple roots are linearly independent.
- The simple roots are complete, so the number of simple roots is equal to  $m$ , the rank of the algebra, or the number of Cartan generators.
- One can construct the entire algebra from the simple roots.

You may have heard of or seen Dynkin diagrams – they are just a short-hand notation for writing down the simple roots. Each simple root is an open circle, and the roots are connected by 1 line if the relative angle between them is 120, 2 lines for 135, 3 lines for 150, and no lines at all if they are orthogonal. So the Dynkin diagram determines all angles between pairs of simple roots. For example, the algebra  $G_2$  consists of two roots at an angle of 150 degrees, while  $su(3)$  just has two roots at an angle of 120 degrees. Of course  $su(2)$  is just one circle.

We have reduced the classification of Lie algebras to geometry. Note that if the Dynkin diagram splits up into two disconnected pieces then all the roots in each piece are orthogonal to all the roots in the other. It follows that the elements of the corresponding Lie algebras commute with each other, giving a pair of simple Lie groups. What remains is to classify consistent connected Dynkin diagrams. You can see Georgi's book on Lie Algebras in Particle Physics, or Fulton and Harris's book on Representation Theory for the details.

### 3.10.6 More on Representations

Representations can be labeled by their unique highest weight state (vector)  $|\mu\rangle$ . All other states in the representation are given by some number of negative simple roots acting on this first state, ie

$$(E_{-\alpha_1})^{n_1} \cdots (E_{-\alpha_m})^{n_m} |\mu\rangle \quad (3.10.26)$$

In fact we can act with the negative simple roots (or just general combinations of any roots, aka any elements of the Lie algebra, since roots are just states in the adjoint which are just Lie algebra generators with definite Cartan subalgebra eigenvalues) in whatever order we want, and sort out the computation using the Lie algebra commutation relations.

*All properties of the representation with highest weight  $|\mu\rangle$  can be worked out algorithmically in this way.* For instance, if we want an inner product of two vectors in the irrep we can write each in terms of lowering operators acting on  $|\mu\rangle$ , then take the inner product, then simplify with commutation relations. This way of working things out is tedious, but deterministic.

This means that irreducible representations are uniquely defined by their highest weight state, and that the highest weight state will be unique within an irrep. If it wasn't, then we could build

two separate sets of states from the two states that each could not be raised, and these separate sets would be independent, and thus would be irreps.

However, as a matter of practice, it's easiest to understand general irreps using *tensor products* of the fundamental representation. Let's consider  $su(N)$ . It has  $m = N - 1$  simple roots, so we can label its irreducible representations with  $m$  numbers which describe the weight of the highest weight state.

In the most familiar case where  $N = 2$ , there is just one number, the spin, characterizing each representation. Or alternatively, we can study a spinor  $u^i$  with  $i = 1, 2$  in the fundamental representation (corresponding respectively to spin up, down), and look at tensor products

$$u^{i_1} \otimes u^{i_2} \otimes \dots \otimes u^{i_k} \quad (3.10.27)$$

How can this be decomposed into irreducible representations? Well when  $i_\alpha = 1$ , we have a state that cannot be raised, and so it will belong to an irreducible representation with  $su(2)$  spin  $k/2$ . However, the tensor product itself is not an irreducible representation, because it contains  $k$  different states with one  $i_\alpha = 2$  and all others  $= 1$ . Whereas the spin  $k/2$  irrep has only has one state with spin- $z$  of  $k/2 - 1$ . In fact, the irrep is what we get if we completely symmetrize over the  $i_\alpha$ . This follows because  $u^1 \otimes \dots \otimes u^1$  is totally symmetric, and acting with  $J_-$  on it will always produce totally symmetric results. In fact, we get all possible  $su(2)$  irreps from symmetric tensors of this form.

Naively you might have expected that there is an 'anti-fundamental' representation acted on by  $v_i \rightarrow (U^\dagger)_i^j v_j$ . This has the property that

$$v_i u^i \quad (3.10.28)$$

is invariant when  $v_i$  is anti-fundamental, and  $u^i$  is a fundamental.

In fact the  $su(2)$  anti-fundamental is just another spin  $1/2$  state, and is therefore just a fundamental. The reason is that the anti-symmetric tensor product of two (anti-)fundamentals

$$x_i \wedge y_j \propto \epsilon_{ij} \quad (3.10.29)$$

is invariant under  $su(2)$ , and so  $\epsilon_{ij} u^j$  transforms in the anti-fundamental whenever  $u^j$  is in the fundamental. (Note I wrote it in this way to make it clear that both the  $i$  and  $j$  indices transform as though they are anti-fundamental rep indices.) Thus it's often said that  $su(2)$  is *real*, because its fundamental and the complex conjugate of the fundamental are the same irrep.

What if we study  $su(3)$ ? Then we have a fundamental representation  $(1, 0)$  and an anti-fundamental representation  $(0, 1)$ , often written by physicists as **3** and  $\bar{\mathbf{3}}$ . But actually, these can be related by noting that the anti-symmetric product of three (anti-)fundamentals

$$x_i \wedge y_j \wedge z_k \propto \epsilon_{ijk} \quad (3.10.30)$$

is invariant under  $su(3)$ . In the context of the strong force, this is a *baryon*, and it is a color singlet (meaning it doesn't transform, or is neutral) state. The reason the proton and neutron are color

neutral is because they are made of three  $su(3)$  fundamental quarks arranged in this way. This means that

$$v_i = \epsilon_{ijk} x^j y^k \quad (3.10.31)$$

is in the  $\bar{\mathbf{3}}$  whenever  $x^i$  and  $y^i$  are in the  $\mathbf{3}$ . This follows because dotting  $u^i v_i$  gives an invariant.

So the fundamental and anti-fundamental are not the same, but we can build the anti-fundamental from the fundamental. In fact, in general for  $su(N)$  the anti-fundamental rep is

$$v_{i_1} = \epsilon_{i_1 \dots i_N} x_2^{i_2} \dots x_N^{i_N} \quad (3.10.32)$$

in terms of  $N - 1$  fundamentals. So we can try to build all irreps from symmetric and anti-symmetric products of fundamentals.

It turns out that this is possible, and the theory for it is based on *Young Tableaux*. These are simply symbols that tell us which indices are to be symmetrized and which are to be anti-symmetrized. For example, in  $su(3)$  we already saw that all tensor products of fundamentals and anti-fundamentals can be written purely in terms of symmetrized and anti-symmetrized fundamentals. All possibilities for mixtures of symmetrization and anti-symmetrization give all irreps for  $su(N)$ . The basic reason why this works is that given a symmetry structure for the highest weight state, the lowering operators will all preserve that symmetry. See e.g. Georgi's book for details.

### 3.11 Spontaneously Broken Global Symmetries and Goldstone Bosons

As a word of philosophy – the most interesting degrees of freedom are those that are *massless*, since they can be seen even at very long distances, and those that are *universal*, in the sense that their existence follows from some general underlying principle. Here we will see that the spontaneous breaking of symmetries leads, universally, to the existence of *Goldstone Bosons*, sometimes called Nambu-Goldstone bosons, which are massless scalar particles.

Spontaneous symmetry breaking occurs when *the equations of motion (or action) have symmetries that are broken by the vacuum state breaks*. Spontaneously broken symmetries are everywhere. Tables break rotational symmetry. They also break translational symmetry, down to the ‘diagonal’, simultaneous translations in space and translations of the material. This spontaneous breaking leads to goldstone bosons, phonons, or sound waves. In fact, that's why sound waves are such a universal phenomenon, because they arise whenever translation symmetry is spontaneously broken, and this happens whenever we have a material object as our local minimum of the energy (vacuum state).

Spontaneous symmetry breaking in field theory is always associated with a degeneracy of vacuum states. As we saw before, linear symmetries

$$\phi_i \rightarrow L_i^j \phi_j \quad (3.11.1)$$

that are respected by the action are also respected by the quantum effective action, ie the exact potential of the theory, so

$$\Gamma[\phi_i] = \Gamma[L_i^j \phi_j] \quad (3.11.2)$$



and in particular, the vacuum state with lowest energy for  $\phi_i$  must be degenerate.

Just because there are multiple minima doesn't mean that the symmetry will be spontaneously broken, because the wavefunction could still just be symmetric. In fact, this happens in many QM examples, and in a small number of spacetime dimensions. But in  $3 + 1$  dimensions and in infinite volume, the states that are stable against small perturbations, and the states that satisfy the cluster decomposition principle, are both those that break the degeneracy, spontaneously breaking the symmetry.

### 3.11.1 Goldstone Bosons

Now let us specialize to continuous symmetries. The canonical example is

$$\mathcal{L} = \partial\phi^\dagger\partial\phi + -\frac{\lambda}{4} (\phi^\dagger\phi - \mu^2)^2 \quad (3.11.3)$$

where  $\mu^2$  is positive. This is the 'Mexican hat potential', and we can write

$$\phi = (\mu + h(x)) e^{i\frac{\pi(x)}{\mu}} \quad (3.11.4)$$

The field  $\pi(x)$  will be massless, representing a Goldstone boson.

Now let us consider the general case. With a position independent VEV for the fields

$$\phi_i(x) = v_i \quad (3.11.5)$$

the fact that there is a symmetry under

$$\phi_i \rightarrow L_i^j \phi_j = \phi_i + \epsilon T_i^j \phi_j \quad (3.11.6)$$

means that the potential (the interaction part of the Lagrangian, or the derivative-less part of the effective action) doesn't change under this transformation, so we must have

$$\frac{\delta V}{\delta \phi_i} T_i^j \phi_j = 0 \quad (3.11.7)$$

If we differentiate this condition a second time with respect to  $\phi_k$ , we get

$$\frac{\delta V}{\delta \phi_i} T_i^k + \frac{\delta^2 V}{\delta \phi_i \delta \phi_k} T_i^j \phi_j = 0 \quad (3.11.8)$$

Now let us assume that the  $\phi_i = v_i$  are at the minimum of the potential. Then the first derivative of  $V$  must vanish, so we find that

$$\frac{\delta^2 V}{\delta \phi_i \delta \phi_k} T_i^j v_j = 0 \quad (3.11.9)$$

But this is just the statement that the directions in field space  $T_i^j v_j$  are massless, since the second derivative of the potential in those directions vanishes.

For example, in the simple case above, we have  $\phi = v + i0$ , and  $T = i$ , so that  $T \cdot v$  rotates purely in the imaginary direction, corresponding to  $\pi \rightarrow \pi + \epsilon$ . Or to put it another way, we can write  $\phi = \phi_1 + i\phi_2$ , in which case  $T_{ij} = \epsilon_{ij}$  acting on  $(\phi_1, \phi_2)$ . So the potential is

$$V(\phi_i) = \frac{\lambda}{4} (\phi_1^2 + \phi_2^2 - \mu^2)^2 \quad (3.11.10)$$

giving

$$\frac{\partial^2 V}{\partial \phi_i \partial \phi_k} \epsilon_{ij} \langle \phi_j \rangle = \frac{\partial^2 V(\mu, 0)}{\partial \phi_k \partial \phi_2} \mu = 0 \quad (3.11.11)$$

for either  $k = 1, 2$ . Thus we see that our simple  $U(1)$  aka  $SO(2)$  example works.

In general, there are massless degrees of freedom corresponding to each of the broken symmetry directions. As we see from equation (3.11.9), the number of Goldstone bosons (scalar degrees of freedom) is the dimensionality of the space

$$T_i^j v_j \quad (3.11.12)$$

For example, say we instead had a theory with  $SO(N)$  symmetry

$$V(\phi_i) = \frac{\lambda}{4} \left( \sum_{i=1}^N \phi_i^2 - \mu^2 \right)^2 \quad (3.11.13)$$

Then we could choose  $\langle \phi_i \rangle = (\mu, 0, 0, \dots, 0)$ . The  $T_i^j$  would be the  $N \times N$  matrices of  $SO(N)$ , so we would have exactly  $N - 1$  Goldstone bosons.

### 3.11.2 Another Proof That Goldstone Bosons Exist

Now let us give a more formal, bare-bones proof that Goldstone bosons exist. By ‘bare-bones’ I mean that we only rely on the fundamentals, without using a Lagrangian description. In particular, we will not assume that any scalar fields exist, but we will see that Goldstone bosons must exist anyway.

What are the fundamentals? Well, we had a continuous symmetry... we could *define* the existence of such a symmetry to be equivalent to the existence of a conserved current

$$\partial_\mu J^\mu = 0 \quad (3.11.14)$$

in the theory, with a charge

$$Q = \int d^3x J^0(x) \quad (3.11.15)$$

that generates the symmetry transformation on all operators, so that

$$[Q, \mathcal{O}_i(x)] = -T_i^j \mathcal{O}_j(x) \quad (3.11.16)$$

for any charged operators. *These operator relations are exact, and are unaffected by spontaneous symmetry breaking!* In other words, these operator relations are true *no matter what state the system is in*, or in other other words, these statements will be true inside any expectation values we wish to compute.

Now we will give an argument that there are Goldstone boson states  $|\pi(\vec{p})\rangle$  such that

$$\int d^3x e^{-i\vec{p}\cdot\vec{x}} \langle \Omega | J^\mu(x) | \pi(\vec{p}) \rangle = iF \frac{p^\mu}{\sqrt{2E_B}} \quad (3.11.17)$$

where  $F$  is called the Goldstone boson decay constant (up to some normalizations). In other words, we will show that when the current acts on the vacuum, it creates a one-Goldstone boson state. This shouldn't be all that surprising. After all, the whole point of spontaneous symmetry breaking is that

$$J^\mu(x)|\Omega\rangle \neq 0 \quad \text{or} \quad Q|\Omega\rangle \neq 0 \quad (3.11.18)$$

because this expression does vanish when the symmetry is unbroken, because that's what it means for the vacuum to be symmetric. The former expression makes more sense than the latter, because acting  $Q$  on the vacuum actually gives infinity.

If we ignore that, we can use these expressions to show that since  $[Q, H] = 0$ , the vacuum and  $Q|\Omega\rangle$  must be degenerate. So if we consider our  $|\pi(\vec{p})\rangle$  state in the limit  $\vec{p} \rightarrow 0$ , this state must have zero energy, since it becomes  $Q|\Omega\rangle$ . This argument is a bit sloppy because these states don't exist.

The right way to proceed is via one of those formal, symmetry and unitarity based arguments, exactly like what we used to derive the Kallen-Lehmann representation (see Weinberg chapter 19.2 for more details). Since it is a commutator of two local operators, we can write

$$\langle [J^\mu(y), \mathcal{O}_i(x)] \rangle = \frac{\partial}{\partial y_\mu} \int d\mu^2 [\rho_i(\mu^2) D_+(y-x, \mu^2) - \rho_i^*(\mu^2) D_+(x-y, \mu^2)] \quad (3.11.19)$$

where

$$D_+(z, \mu^2) = \int \frac{d^4p}{(2\pi)^4} \theta(p^0) \delta(p^2 - \mu^2) e^{ipz} \quad (3.11.20)$$

is the retarded part of the Feynman propagator, and we have assumed that  $J^\mu$  and  $\mathcal{O}_i$  are hermitian.

This expression has to vanish when  $x-y$  is spacelike, so  $\rho_i(\mu^2)$  must be real with

$$\langle [J^\mu(y), \mathcal{O}_i(x)] \rangle = \frac{\partial}{\partial y_\mu} \int d\mu^2 \rho_i(\mu^2) (D_+(y-x, \mu^2) - D_+(x-y, \mu^2)) \quad (3.11.21)$$

Now let us use current conservation, differentiate both sides with respect to  $y$ , and use

$$(\Box_y - \mu^2) D_+(y-x, \mu^2) = 0 \quad (3.11.22)$$

This means that

$$\int d\mu^2 \mu^2 \rho_i(\mu^2) (D_+(y-x, \mu^2) - D_+(x-y, \mu^2)) = 0 \quad (3.11.23)$$

for all  $x$  and  $y$ . This is only possible if

$$\mu^2 \rho_i(\mu^2) = 0 \quad (3.11.24)$$

You might be tempted to conclude that  $\rho_i(\mu^2) = 0$ , but this is not possible, because  $J^\mu$  generates the symmetry transformation. In particular, integrating our first equation says that

$$-T_i^j \langle \mathcal{O}_j \rangle = i \int d\mu^2 \rho_i(\mu^2) \quad (3.11.25)$$

and the right side cannot vanish if any charged operators in the theory get vacuum expectation values.

Thus we must have

$$\rho_i(\mu^2) = i\delta(\mu^2) T_i^j \langle \mathcal{O}_j \rangle \quad (3.11.26)$$

So  $\rho_i$  does not vanish, but instead is proportional to a delta function localized at zero mass. This is our massless goldstone boson pole!

Note that this proof did not use any properties of the ‘fundamental’ theory or the Lagrangian. In particular, although we showed that there exists a scalar particle in the spectrum, we did not assume that there are any elementary or fundamental scalar fields. In fact, in the most relevant HEP phenomenological application, namely to pions, there are no elementary scalars in the theory – the goldstone bosons are bound states of quarks. However, the theorem still applies!

### 3.11.3 Connecting the Two Proofs

Let’s see how these ideas are related in the simple case of a linear sigma model with a  $U(1)$  symmetry. In that case

$$J_\mu = i (\partial_\mu \phi^\dagger \phi - \phi^\dagger \partial_\mu \phi) \quad (3.11.27)$$

Writing this using  $\phi = (v + h)e^{i\pi/v}$ , we see that

$$J_\mu(x) = 2v \partial_\mu \pi(x) \quad (3.11.28)$$

Thus it’s not surprising at all that in the broken phase, when  $v \neq 0$ , we have that

$$J_\mu(p)|0\rangle = 2vp_\mu \pi(p)|0\rangle = 2vp_\mu |p\rangle \quad (3.11.29)$$

and so  $J_\mu(p)$  obviously creates a 1-goldstone boson state!

### 3.11.4 Goldstone Bosons are Everywhere

You know of a lot of light and massless degrees of freedom (probably many of which you haven’t even thought of in that way). Many of them are Goldstone bosons:

- Single free particle states are goldstone bosons of (completely) broken translation invariance. The goldstone mode is a  $1 + 0$  dimensional QFT, otherwise known as a QM theory of that single particle, by itself! Since they break 3 translations, there are 3 degrees of freedom, namely the position of the particle in the  $x$ ,  $y$ , and  $z$  directions.
- Materials and objects spontaneously break translational symmetry everywhere, not just at one point. The associated Goldstone bosons are phonons, which carry sound. This is one way of explaining why sound is so ubiquitous.
- Between these two limits, objects like membranes and strings partially break translation symmetry, and perpendicular waves (vibrations) of these objects are goldstone modes.
- Goldstone bosons are ubiquitous in condensed matter systems (with or without an approximate Lorentz invariance). An interesting example are magnetic systems that spontaneously break rotational invariance by magnetizing in one particular direction. One famous example is the Heisenberg model. The Goldstone bosons are called Magnons.
- As we will discuss, pions are approximate Goldstone bosons because the up and down quarks are light.
- In supersymmetric theories, the spontaneous breaking of supersymmetry leads to particles called ‘Goldstinos’, which are Goldstone fermions.
- The inflaton mode in the early universe probably spontaneously broke time translation symmetry, and therefore its degree of freedom can be viewed as a goldstone boson.
- There is a hypothesized particle called the QCD axion, which would solve something called the strong CP problem. It is an approximate goldstone boson.

### 3.11.5 Pseudo-Goldstone Bosons and Pions

The kinetic terms for the up and down quarks, written in Weyl notation, are

$$\mathcal{L} = iu_L^\dagger \not{D} u_L + iu_R^\dagger \not{D} u_R + id_L^\dagger \not{D} d_L + id_R^\dagger \not{D} d_R + m_u(u_L^\dagger u_R + u_R^\dagger u_L) + m_d(d_L^\dagger d_R + d_R^\dagger d_L)$$

If we set  $m_u = m_d = 0$ , then there is an  $SU(2)_L \times SU(2)_R$  global symmetry where we rotate

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix} \rightarrow U_L \begin{pmatrix} u_L \\ d_L \end{pmatrix} \quad (3.11.30)$$

and similarly for the right-handed Weyl spinors. The mass terms for the quarks break this symmetry, but it turns out that they are small, so this is a good *approximate symmetry* of QCD. It is called a chiral symmetry since it acts differently on left and right handed spinors. We can also write it as  $SU(2)_V \times SU(2)_A$ , for  $V = L + R$  and  $A = L - R$ . Note that the  $SU(2)_V$  acts in the ‘obvious’ way on the Dirac spinors  $(u, d)$ , so  $\bar{u}u + \bar{d}d$  is invariant under  $SU(2)_V$ , but this expression is not invariant under  $SU(2)_A$ .

There are also  $U(1)_L \times U(1)_R$  symmetries as well, usually written as  $U(1)_V \times U(1)_A$ , vector and axial, corresponding to  $L + R$  and  $L - R$ . It turns out that the  $L - R$  symmetry is broken by quantum effects called anomalies, while the  $U(1)_V$  corresponds to baryon number. Baryon number is also anomalous in the full standard model, although  $B - L$ , baryon minus lepton number, is a good symmetry.

Anyway, we are interested in this system because the  $SU(2)_L \times SU(2)_R$  symmetry is spontaneously broken by QCD, at the scale  $\Lambda_{QCD} \sim 300$  MeV. Specifically, there are VEVs

$$\langle \bar{u}u \rangle = \langle \bar{d}d \rangle = V^3 \quad (3.11.31)$$

in the books notation, which translates to

$$\langle u_L^\dagger u_R + u_R^\dagger u_L \rangle = V^3 \quad (3.11.32)$$

and similarly for the down quark, with  $V \sim \Lambda_{QCD}$ . Note that  $SU(2)_V$  is unbroken by these VEVs, but  $SU(2)_A$  is spontaneously broken. This symmetry breaking arises from QCD (the strong force), but there is not any easy analytic way of deriving it – in fact we do not have a theoretical proof. But it does not matter, since Goldstone's theorem only requires the existence of a broken symmetry, not a mechanism for its breaking!

The group  $SU(2)_V$  is often called isospin; it relates protons and neutrons, which are also a doublet, because protons are  $uud$  and neutrons are  $udd$  states. Ignoring electric charge and quark masses (a very good approximation), the proton and neutron are exactly related by a symmetry.

From now on it will be irrelevant what the short-distance physics was that gave rise to the  $SU(2) \times SU(2) \rightarrow SU(2)_V$  symmetry breaking pattern – everything will follow from symmetry and Goldstone's theorem. We will write down a unique effective field theory with these symmetry properties. This means we want to write down the most general possible Lagrangian for the 3 Goldstone bosons of the broken  $SU(2)_A$ , constrained by the remaining  $SU(2)_V$ .

Thus we need a parameterization that encodes the generators of  $SU(2)_A$  as Goldstone bosons, and it has to transform in a nice way under the remaining  $SU(2)_V$ . To do this, let us consider a  $2 \times 2$  matrix of scalar fields  $U_{ij}$  transforming linearly under  $SU(2) \times SU(2)$  via

$$U \rightarrow g_L U g_R^\dagger \quad (3.11.33)$$

We can think of this  $U$  as the quark bilinear

$$U = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \begin{pmatrix} u_R^\dagger & d_R^\dagger \end{pmatrix} = \begin{pmatrix} u_L u_R^\dagger & u_L d_R^\dagger \\ d_L u_R^\dagger & d_L d_R^\dagger \end{pmatrix} \quad (3.11.34)$$

This gets a VEV according to the symmetry breaking structure, so that  $\det U^\dagger U$  is invariant and proportional to the SSB. In other words, it is an order parameter for SSB, and it would vanish if the symmetry were unbroken. The phases can be parameterized by

$$U(x) = e^{2i \frac{\pi_a(x) \tau^a}{F_\pi}} \quad (3.11.35)$$

where we note that the V and A transformations are

$$U \rightarrow g_V U g_V^{-1}, \quad U \rightarrow g_A U g_A \quad (3.11.36)$$

where the first is unbroken by the VEV proportional to the identity, but the latter is broken. Thus under transformations  $g_L = e^{i\theta_L^a \tau_a}$  and  $g_R = e^{i\theta_R^a \tau_a}$  we have

$$\pi^a \rightarrow \pi^a + \frac{F_\pi}{2}(\theta_L^a - \theta_R^a) - \frac{1}{2}f^{abc}(\theta_L^b + \theta_R^b)\pi^c + \dots \quad (3.11.37)$$

This implies that under the unbroken  $SU(2)_V$  the  $\pi^a$  transform in the adjoint representation, but they transform non-linearly under the broken  $SU(2)_A$  – to leading order they shift, as goldstone bosons should. Note that this shift symmetry forbids a mass term for the  $\pi^a$ , as is to be expected.

We conventionally write  $U$  in terms of the pions as

$$\exp \left[ \frac{i}{F_\pi} \begin{pmatrix} \pi^0 & \sqrt{2}\pi^- \\ \sqrt{2}\pi^+ & -\pi^0 \end{pmatrix} \right] \quad (3.11.38)$$

where  $\pi^0$  is the neutral pion, and  $\pi^\pm$  are the charged pions. So we have all 3 of the degrees of freedom

With nothing but symmetry to guide us, we must write down the most general theory of  $U$  that is invariant under  $SU(2) \times SU(2)$ . This looks like

$$\begin{aligned} \mathcal{L} = & \frac{F_\pi^2}{4} \text{Tr} [(D_\mu U)(D^\mu U)^\dagger] + L_1 \text{Tr} [(D_\mu U)(D^\mu U)^\dagger]^2 \\ & + L_2 \text{Tr} [(D_\mu U)(D_\nu U)^\dagger] \text{Tr} [(D^\mu U)(D^\nu U)^\dagger] + L_3 \text{Tr} [(D_\mu U)(D^\mu U)^\dagger (D_\nu U)(D^\nu U)^\dagger] + \dots \end{aligned} \quad (3.11.39)$$

This is the Chiral Lagrangian for pion interactions; all terms must have derivatives because  $U^\dagger U = 1$ . There would be a similar Lagrangian for any  $SU(N) \times SU(N) \rightarrow SU(N)$  symmetry breaking pattern. The interactions necessitated by this theory are strictly constrained by symmetry, for example

$$\frac{F_\pi^2}{4} \text{Tr} [(D_\mu U)(D^\mu U)^\dagger] = \frac{1}{2}(\partial\pi^0)^2 + (D_\mu \pi^+)(D^\mu \pi^-) + \frac{1}{F_\pi^2} \left( -\frac{1}{3}(\pi^0)^2(D_\mu \pi^+)(D^\mu \pi^-) + \dots \right) + \dots$$

The terms involving  $L_i$  all have at least 4 derivatives, so their effects are always subdominant to the first term, which is uniquely fixed in terms of the pion decay constant  $F_\pi$ .

But what about the quark masses, which break the symmetry explicitly? These originally appeared in the Lagrangian as

$$\bar{q} M q \quad (3.11.40)$$

where  $M$  is a  $2 \times 2$  matrix in  $(u, d)$  space. If we imagined that

$$M \rightarrow g_L M g_R^\dagger \quad (3.11.41)$$

under  $SU(2) \times SU(2)$ , then our Lagrangian would be invariant once again. This means that all violations of the symmetries have to show up in invariant combinations, *if we assume this*

transformation rule for  $M$ . The use of this idea is called *the spurion method*, because we view  $M$  as a ‘spurious’ field. Using this  $M$  matrix, we can add terms to the Chiral Lagrangian of the form

$$\begin{aligned}\mathcal{L}_M &= \frac{V^3}{2} \text{Tr} [MU + M^\dagger U^\dagger] \\ &= V^3(m_u + m_d) - \frac{V^3}{2F_\pi^2}(m_u + m_d)(\pi_0^2 + 2\pi^+\pi^-)\end{aligned}\quad (3.11.42)$$

Thus the pion masses are

$$m_\pi^2 = \frac{V^3}{F_\pi^2}(m_u + m_d) \quad (3.11.43)$$

This relation has been confirmed using lattice QCD (to measure  $V$ ). This means that pions are *pseudo-Goldstone bosons*, because they are associated with spontaneous breaking of an approximate symmetry. This also accords from expectations from current conservation, where

$$\partial_\mu J_a^\mu \propto m_q \bar{q} \gamma^5 \tau^a q \neq 0 \quad (3.11.44)$$

sets the pseudo-Goldstone boson mass.

Historically, the theory of the Chiral Lagrangian was very important for the development of effective field theory. Originally, the scattering amplitudes and interactions of pions were computed using something called ‘Current Algebra’, which just means that people computed correlators of the approximately conserved currents that create pion states (Partially Conserved Axial Current, or PCAC). Most researchers thought that QFT could not make sense as a description of pions, essentially due to the fact that their interactions are all ‘irrelevant’, and so are short-distance sensitive. Weinberg and others argued that the pion EFT of the Chiral Lagrangian did in fact make sense as an expansion at low-energy, and in particular, that it was predictive. This, along with Wilson’s ideas, led to a major rethinking of what QFT means.

### 3.11.6 Generalization to $SU(3) \times SU(3)$

We should briefly note that this analysis generalizes to  $SU(3) \times SU(3)$ , by including the strange quark. One then includes the fields

$$U \equiv \exp \left[ \frac{\sqrt{2}i}{F_\pi} \begin{pmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & \pi^- & K^+ \\ \pi^+ & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & K^0 \\ \bar{K}^- & \bar{K}^0 & -\sqrt{\frac{2}{3}}\eta^0 \end{pmatrix} \right] \quad (3.11.45)$$

obtaining interactions for the pions, the eta, and the Kaons. Note that since  $m_s \approx 100$  MeV, this isn’t as good of an approximation for the interactions, but it’s still good for organization.

One can also use this symmetry structure to characterize baryons – the proton and neutron end up in an octet, just like the pions. That looks like

$$\begin{pmatrix} \frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda & \Sigma^+ & P^+ \\ \Sigma^- & -\frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda & N \\ \Xi^- & \Xi^0 & -\sqrt{\frac{2}{3}}\Lambda \end{pmatrix} \quad (3.11.46)$$



Murray Gell-Mann called this the eightfold way.

### 3.12 Non-Abelian Gauge Theories

We have discussed Lie algebra representations, and we used them to talk about global symmetries and their breaking. Now let us move on to a major class of theories – Yang-Mills theories, which involve massless spin 1 bosons transforming according to a non-Abelian Lie group.

We can discover these theories by asking... how can we make massless spin 1 particles carry a gauge charge? Since our bosons are massless, if we are to describe them using local fields  $A_\mu^a(x)$  we must eliminate one degree of freedom, which means we need a gauge invariant Lagrangian. This suggests that we need to use something like

$$F_{\mu\nu}^a \sim D_\mu A_\nu^a - D_\nu A_\mu^a \quad (3.12.1)$$

where we have constructed a gauge covariant  $F$  using a covariant derivative (to be defined). Then we could make an invariant via

$$\sum_a F_{\mu\nu}^a F_a^{\mu\nu} \quad (3.12.2)$$

where  $a$  labels the different charged, massless, spin 1 bosons.

But we run into the problem that if the derivative is really covariant (to couple to some other spin 1 bosons), we must have bosons that transform as

$$A_\mu^a(x) \rightarrow G_b^a(x) A_\mu^b(x) \quad (3.12.3)$$

under the gauge transformation associated with the covariant derivative  $D_\mu$ . Since  $A_\mu$  are real  $G$  must be some Lie Group, with  $A_\mu^a$  in a representation of the group. But the  $A_\mu^a$  must also shift under local gauge transformations

$$A_\mu^a(x) \rightarrow A_\mu^a(x) + \partial_\mu \alpha^a(x) \quad (3.12.4)$$

to preserve masslessness. How can these ideas all be compatible?

The resolution is that  $A_\mu^a$  both rotate and shift under the same gauge transformation – in other words, the bosons  $A_\mu^a$  interact with each other, all as representations of one Lie algebra. Since they are a part of the covariant derivative, and must be able to parameterize the group itself, the  $A_\mu$  must be in the adjoint representation.

Specifically, we must have the infinitesimal transformation rule

$$A_\mu^a(x) \rightarrow A_\mu^a(x) + \frac{1}{g} \partial_\mu \alpha^a(x) - f^{abc} \alpha^b(x) A_\mu^c(x) \quad (3.12.5)$$

This means that the field strength is

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c \quad (3.12.6)$$

The field strength simply transforms like a field in the adjoint representation of the group. Unlike in QED, it is not gauge invariant, but it is gauge covariant, ie it does not shift like the gauge field  $A_\mu^a$ . But we can write a completely gauge invariant action

$$\mathcal{L}_{YM} = -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} \quad (3.12.7)$$

Now we will give a more geometric picture for these Yang-Mills, or Non-Abelian gauge theories.

### 3.12.1 Wilson Lines

Let us begin by noting a problem – how can we compare  $\phi(x)$  and  $\phi(y)$  in scalar QED? Under a gauge transformation

$$\phi(x) - \phi(y) \rightarrow e^{i\alpha(x)}\phi(x) - e^{i\alpha(y)}\phi(y) \quad (3.12.8)$$

and so this difference depends very much on our choice of gauge. We cannot invariantly say  $\phi(x) = \phi(y)$  in scalar QED. This also means that

$$\frac{\phi(x + \epsilon_\mu) - \phi(x)}{\epsilon} = \partial_\mu \phi \quad (3.12.9)$$

is not well-defined, since we cannot compare nearby points.

To fix this problem, we can introduce a new bi-local object

$$W(x, y) \rightarrow e^{i\alpha(x)}W(x, y)e^{-i\alpha(y)} \quad (3.12.10)$$

so that if we study

$$W(x, y)\phi(y) - \phi(x) \quad (3.12.11)$$

then it transforms correctly. In words, the object  $W(x, y)$  connects gauge transformations at  $y$  to those at  $x$ . We can now compute derivatives in a sensible way, as

$$D_\mu \phi(x) \equiv \lim_{\delta x \rightarrow 0} \frac{W(x, x + \delta x)\phi(x + \delta x) - \phi(x)}{\delta x^\mu} \quad (3.12.12)$$

Then  $D_\mu \phi$  transforms in the same way as  $\phi(x)$  under gauge transformations.

We must have that  $W(x, x) = 1$ , and we can expand around this to obtain

$$W(x, x + \delta x) = 1 - ie\delta x^\mu A_\mu(x) + \dots \quad (3.12.13)$$

This defines both  $A_\mu$  and the covariant derivative  $D_\mu$ . Now we see why the gauge field  $A_\mu$ , or sometimes just  $D_\mu$ , are called the *connection*. This is very similar to in GR, where we use the connection to relate vectors at different points on a curved manifold.

We can now exponentiate  $A_\mu$  to write a closed form expression

$$W(x, y) = \exp \left[ ie \int_y^x A_\mu(z(t)) \frac{dz^\mu}{dt} dt \right] \quad (3.12.14)$$

where we mean the line integral from  $y$  to  $x$  in the exponent. This object is known as a Wilson line, on a path from  $y$  to  $x$ . Note that under a gauge transformation it transforms as

$$W \rightarrow \exp \left[ ie \int_y^x (A_\mu(z(t)) + \partial_\mu \alpha(z(t))) \frac{dz^\mu}{dt} dt \right] = e^{i\alpha(x)} W(x, y) e^{-i\alpha(y)} \quad (3.12.15)$$

which is what we wanted. This gauge transformation rule is independent of the path. If we set  $x = y$  then we get a Wilson loop, which is gauge invariant. We can use Stokes theorem to write a Wilson loop as

$$W(x, x) = \exp \left[ i \frac{e}{2} \int_\Sigma F_{\mu\nu} d^2 \sigma^{\mu\nu} \right] \quad (3.12.16)$$

where  $\Sigma$  is a 2-dimensional surface enclosed by the loop's path.

Now note that we can take multiple covariant derivatives, so we can compute

$$[D_\mu, D_\nu] = [\partial_\mu, \partial_\nu] - ie[\partial_\mu, A_\nu] + ie[\partial_\nu, A_\mu] = -ieF_{\mu\nu} \quad (3.12.17)$$

In fact, we can define  $F_{\mu\nu}$  by this commutator. This may be familiar, since it is exactly what we use to define the Riemann tensor in differential geometry and general relativity. It is the result of comparing field values around an infinitesimal closed loop.

**say something about soft factors and Wilson lines?**

### 3.12.2 Non-Abelian Wilson Lines

We can generalize this pretty easily to the non-abelian case. When we studied goldstone bosons, we discussed theories like

$$\mathcal{L} = \sum_{j=1}^N \left[ \partial_\mu \phi_j^\dagger \partial^\mu \phi_j - m^2 \phi_j^\dagger \phi_j \right] - \lambda \left( \sum_j \phi_j^\dagger \phi_j \right)^2 \quad (3.12.18)$$

which have a manifest symmetry under

$$\phi_i \rightarrow U_{ij} \phi_j, \quad \phi_i^\dagger \rightarrow U_{ij}^\dagger \phi_j^\dagger \quad (3.12.19)$$

where  $U$  is a unitary matrix. If we promote this to a local symmetry where  $U \rightarrow U(x)$ , then we can no longer compare  $\phi_j(x)$  and  $\phi_j(y)$  without a Wilson line

$$W(x, y) = P \left\{ e^{ig \int_y^x A_\mu^a(z) T_a dz^\mu} \right\} \quad (3.12.20)$$

where the  $P$  denotes *path ordering*, which is necessary because group generators at different points do not commute. This expression is defined by its Taylor series, and the path ordering works just like time ordering on the  $S$  operator from scattering theory (which you recall we used to derive perturbation theory and Feynman diagrams from the Hamiltonian formalism).

Note that this defines Wilson lines in any representation; we just use  $T_a$  from that representation. Under gauge transformations

$$W \rightarrow U(x)WU^\dagger(x) = e^{i\alpha^a(x)T^a}W e^{-i\alpha^b(y)T^b} \quad (3.12.21)$$

It's also reasonable to view  $\mathbf{A}_\mu = A_\mu^a T_a$  as a matrix valued field.

Now we can derive the gauge transformation properties of  $\mathbf{A}_\mu$  from those of the Wilson line, since infinitesimally (in  $\delta x$ )

$$W(x + \delta x, x) \approx 1 - ig\mathbf{A}_\mu \delta x^\mu \quad (3.12.22)$$

This means that for infinitesimal gauge transformations (in small  $\alpha_a$ )

$$\begin{aligned} W(x + \delta x, x) &\rightarrow (1 + i\alpha^a(x + \delta x)T_a)(1 - ig\mathbf{A}_\mu \delta x^\mu)(1 - i\alpha^a(x)T_a) \\ &\approx i\partial_\mu \alpha^a T_a + g[\alpha^a(x)T_a, \mathbf{A}_\mu] \end{aligned} \quad (3.12.23)$$

We can equivalently write a general version

$$\begin{aligned} W(x + \delta x, x) &\rightarrow U(x + \delta x)(1 - ig\mathbf{A}_\mu \delta x^\mu)U^{-1}(x) \\ &\approx \partial_\mu U U^{-1} - igU\mathbf{A}_\mu U^{-1} \end{aligned} \quad (3.12.24)$$

Thus the transformation rule for the gauge field is

$$\mathbf{A}_\mu \rightarrow U\mathbf{A}_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1} \quad (3.12.25)$$

This also implies that the covariant derivative

$$D_\mu \phi \equiv (\partial_\mu + ig\mathbf{A}_\mu)\phi \quad (3.12.26)$$

transforms in the same way as the field  $\phi$  in any given representation of the group. So we see that, as discussed, non-abelian gauge fields both rotate and shift under the gauge symmetry. They are charged under the force they generate.

Finally, let us look at the commutator of the covariant derivatives

$$\begin{aligned} [D_\mu, D_\nu] &= -ig(\partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu) - g^2[\mathbf{A}_\mu, \mathbf{A}_\nu] \\ &= -igF_{\mu\nu}^a T_a \end{aligned} \quad (3.12.27)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c \quad (3.12.28)$$

In the abelian case this reduces to the usual electromagnetic field strength. Note that  $F_{\mu\nu}^a$  transforms as

$$F_{\mu\nu}^a \rightarrow F_{\mu\nu}^a - f^{abc}\alpha^b F_{\mu\nu}^c \quad (3.12.29)$$

under gauge transformations, which means that it transforms like a charged field (not a gauge field) in the adjoint representation.

Now we can write down a gauge invariant Lagrangian

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu}^a)^2 + \sum_{i,j=1}^N \bar{\psi}_i (\delta_{ij} i \not{D} + g A_{ij}^a T_{ij}^a - m \delta_{ij}) \psi_j \quad (3.12.30)$$

where I used fermion fields instead of scalars for variety. This is a non-abelian gauge theory. In the case of  $N = 3$ , this is precisely the Lagrangian for QCD, the theory of the strong force, which has an  $SU(3)$  gauge symmetry. The three ‘colors’ of quarks are in the fundamental of  $SU(3)$ , and each quark (up, down, strange, charm, top, and bottom) correspond to a separate Dirac spinor field.

Since we’ve mentioned QCD, we can also note that one can include a term

$$\mathcal{L}_\theta = \theta \epsilon^{\mu\nu\alpha\beta} F_{\mu\nu}^a F_{\alpha\beta}^a = 2\theta \partial_\mu (\theta \epsilon^{\mu\nu\alpha\beta} A_\nu^a F_{\alpha\beta}^a) \quad (3.12.31)$$

This term is a total derivative, so it cannot contribute in perturbation theory, but it does have effects. For example,  $\theta$  would give the neutron an electric dipole moment. The absence of such a contribution is known as the *strong CP problem*.

### 3.12.3 Conserved Currents

Our Lagrangians for non-abelian gauge theories have a global symmetry where

$$\psi_i \rightarrow \psi_i + i\alpha^a T_{ij}^a \psi_j \quad (3.12.32)$$

$$A_\mu^a \rightarrow A_\mu^a - f^{abc} \alpha^b A_\mu^c \quad (3.12.33)$$

with infinitesimal constant  $\alpha$ . From Noether’s theorem, there’s an associated conserved current

$$J_\mu^a = -\bar{\psi}_i \gamma_\mu T_{ij}^a \psi_j + f^{abc} A_\nu^b F_{\mu\nu}^c \quad (3.12.34)$$

Using the equations of motion, one can show that this current is conserved.

However, *this conserved current is not gauge-invariant*. Thus it has a very limited physical meaning. In particular, although one can write down associated conserved charges, those charges are not gauge invariant either. So in non-abelian gauge theories, there is no such thing as a classical current (so you cannot run non-abelian gauge current through a wire). We also do not have a Gauss’s law.

We can also define a matter current

$$J_M^{a\mu} = -\bar{\psi}_i \gamma^\mu T_{ij}^a \psi_j \quad (3.12.35)$$

which is gauge covariant. It is not conserved, but instead satisfies

$$D_\mu J_M^{a\mu} = 0 \quad (3.12.36)$$

with a covariant derivative. So it is not conserved, and there’s no associated charge.

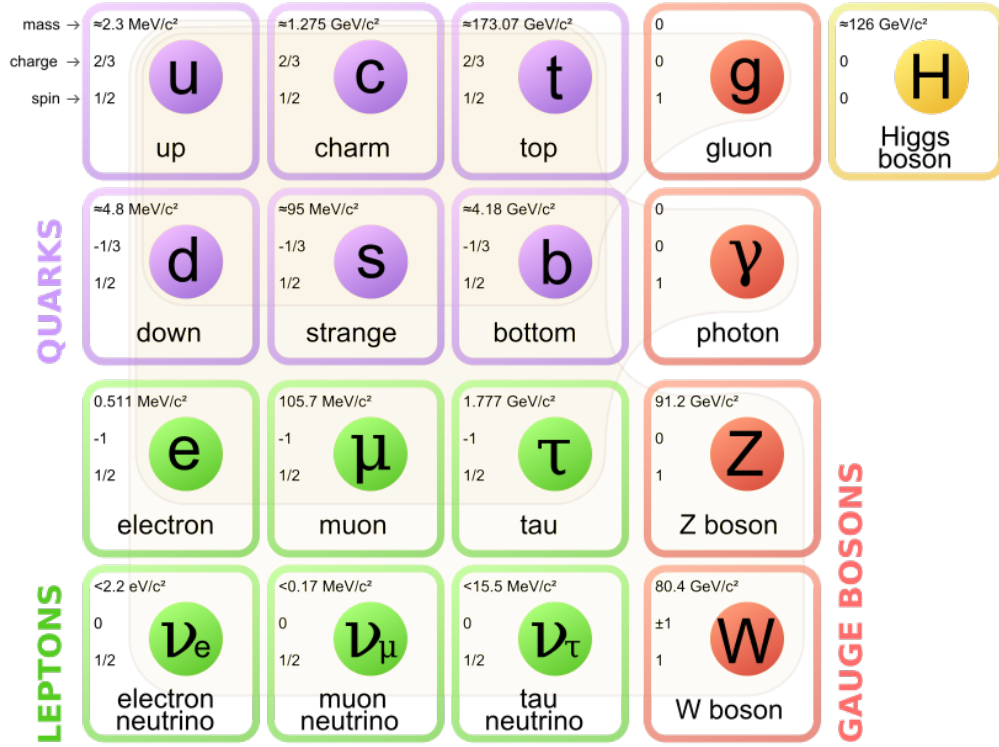


Figure 1: The Standard Model particle content, from Wikipedia.

It's worth noting that these results are not an accident, but follow from a general theorem, called the Weinberg-Witten theorem. It says that a theory with a global non-Abelian symmetry under which a massless spin 1 particle is charged can not have a gauge-invariant conserved current. Similarly, a theory with a conserved and Lorentz-covariant energy-momentum tensor can not have a massless particle of spin 2. This goes a long way towards showing that the graviton cannot be a composite state.

### 3.12.4 Strong and Electromagnetic Forces in the Standard Model

We're now at the point where we can write down the strong and electromagnetic part of the standard model, in the form of a Lagrangian. In fact, we could write down the whole SM; the only part that you'll be missing is how the Higgs gives mass to the  $W$  and  $Z$  bosons. But for simplicity let's leave out the weak interactions for now.

At low energies, below around 80 GeV, we can view the SM as having the gauge group  $SU(3) \times U_{EM}(1)$ , corresponding to the strong force (the gluons) and the electromagnetic force. The quarks are charged under both, the leptons only under electromagnetism. Thus we have four kinds of

covariant derivatives

$$\left( \partial_\mu + ie \frac{2}{3} A_\mu + ig_s G_\mu^a T_{ij}^a \right) U_j \quad (3.12.37)$$

$$\left( \partial_\mu - ie \frac{1}{3} A_\mu + ig_s G_\mu^a T_{ij}^a \right) D_j \quad (3.12.38)$$

$$(\partial_\mu - ie A_\mu) E \quad (3.12.39)$$

$$(\partial_\mu) N \quad (3.12.40)$$

Note that there are three ‘colors’ of quarks (because they are in the fundamental of  $SU(3)$ , meaning that there are way more quark than lepton degrees of freedom. Roughly speaking, the proton is made from up-up-down, and the neutron from up-down-down. The *baryons* are color neutral because they can be written as

$$\epsilon^{ijk} U_i U_j U_k \quad (3.12.41)$$

where the contraction is on their  $SU(3)$  color indices. One can always make baryons from  $N$  fundamental fields in  $SU(N)$ . In contrast, pions can be viewed as composed from combinations like

$$\bar{U}_i U^i + \bar{D}_i D^i \quad (3.12.42)$$

which are invariant in a more familiar way. Note that for  $SU(2)$  these two types of invariants would coincide.

Since we are ignoring the weak interactions, we are free to view these all as dirac spinors. There are three copies of this particle content, the three flavors or three generations. The up-type quarks, down-type quarks, and electron-like leptons have conventional Dirac masses. We don’t know whether the neutrino masses are majorana or Dirac.

Particle contents like what we have above are usually described by contemporary physicists purely in terms of the number of fields with a given charge. For  $U(1)$  gauge groups we need to specify a number (for the charge), whereas for groups like  $SU(N)$  we specify the representation. It’s unusual to see representations besides the fundamental and adjoint, but sometimes they do appear.

The weak interactions complicate things significantly, and are not naturally written in terms of Dirac spinors (although we can easily accomplish that anyway using  $1 \pm \gamma_5$ ).

### 3.13 Quantization of Yang-Mills Theories

Now let us discuss quantization of Yang-Mills theories. The  $R_\xi$  gauges from QED must be generalized. We will see that in order to eliminate the unphysical degrees of freedom, the longitudinal polarizations of the massless spin 1 Yang-Mills bosons, we must introduce *ghosts*, which are essentially ‘negative degrees of freedom’. There are gauges where this isn’t necessary, but they destroy manifest Lorentz invariance, so no one uses them for practical calculations.

Recall that the problem with obtaining a photon propagator immediately is that

$$(g_{\mu\nu} \square - \partial_\mu \partial_\nu) A^\mu = J_\nu \quad (3.13.1)$$

and the differential operator simply isn't invertible. To get around this we added the gauge fixing term

$$\frac{1}{2\xi}(\partial_\mu A^\mu)^2 \quad (3.13.2)$$

to the action. We justified this by multiplying up and down by a certain function that effected the introduction of this gauge fixing term, and then we showed that all it did was change the overall normalization of the path integral.

For non-abelian theories we will use the same trick, but we will see that it is not so innocuous. We start by introducing  $N^2 - 1$  fields  $\pi^a$ , one for each gauge boson DoF. Gauge fields transform as

$$A_\mu^a(x) \rightarrow A_\mu^a(x) + \frac{1}{g}\partial_\mu \alpha^a(x) - f^{abc}\alpha^b(x)A_\mu^c(x) \quad (3.13.3)$$

or equivalently as

$$A_\mu^a(x) \rightarrow A_\mu^a(x) + \frac{1}{g}D_\mu \alpha^a \quad (3.13.4)$$

Thus let us multiply and divide our path integral by

$$f[A] = \int \mathcal{D}\pi \exp \left[ -i \int d^4x \frac{1}{2\xi} (\partial^\mu D_\mu \pi^a)^2 \right] \quad (3.13.5)$$

In the abelian case,  $\pi$  was neutral and  $f$  was just a number, but now it actually depends on  $A_\mu^a$ .

In Non-Abelian theories, we can still define a gauge parameter  $\alpha^a$  that takes us to Lorenz gauge, so that

$$\partial^\mu A_\mu^a = \frac{1}{g}\partial^\mu D_\mu \alpha^a[A] \quad (3.13.6)$$

If we shift  $\pi^a$  by  $\frac{\alpha^a}{g}$  for  $\alpha^a$ , we obtain

$$f[A] = \int \mathcal{D}\pi \exp \left[ -i \int d^4x \frac{1}{2\xi} (\partial^\mu A_\mu^a - \partial^\mu D_\mu \pi^a)^2 \right] \quad (3.13.7)$$

Thus by multiplying the path integral by  $f/f$ , we obtain

$$\begin{aligned} & \int \mathcal{D}\pi \mathcal{D}A_\mu \mathcal{D}\phi_i \frac{1}{f[A]} \exp \left[ iS[A, \phi_i] - i \int d^4x \frac{1}{2\xi} (\partial^\mu A_\mu^a - \partial^\mu D_\mu \pi^a)^2 \right] \\ &= \mathcal{N} \int \mathcal{D}A_\mu \mathcal{D}\phi_i \frac{1}{f[A]} \exp \left[ iS[A, \phi_i] - i \int d^4x \frac{1}{2\xi} (\partial^\mu A_\mu^a)^2 \right] \end{aligned} \quad (3.13.8)$$

where we shifted  $A_\mu$  by  $D_\mu \pi^a$  in the second line, and then the explicit  $\mathcal{D}\pi$  integral gives an overall constant.



The difference from the abelian case is that we have the explicit functional  $f[A]$  in the denominator. But it is just

$$f[A] \propto \sqrt{\frac{1}{\det[(\partial^\mu D_\mu)^2]}} \quad (3.13.9)$$

So its inverse is simply  $\det(\partial^\mu D_\mu)$ . We saw when we discussed the path integral that positive determinant factors are just Grassman integrals. Thus we can write

$$\frac{1}{f[A]} = \int \mathcal{D}c \mathcal{D}\bar{c} \exp \left[ i \int d^4x \bar{c}(x) (-\partial^\mu D_\mu) c(x) \right] \quad (3.13.10)$$

for *ghost fields*  $c$  and  $\bar{c}$ . Thus we have the full Lagrangian for the gauge fixed gauge field and the fermionic ghosts

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 - \frac{1}{2\xi} (\partial^\mu A_\mu^a)^2 + (\partial_\mu \bar{c}^a) (\delta^{ac} \partial_\mu + g f^{abc} A_\mu^b) c^c \quad (3.13.11)$$

This is the Faddeev-Popov Lagrangian for the Yang-Mills field and ghosts in the  $R_\xi$  gauge. We can find an appropriate ghost Lagrangian for any gauge choice, not just this one, as the book and Weinberg explain.

One can think of the ghosts as ‘negative degrees of freedom’ included to cancel the unphysical gauge boson modes. In fact it’s possible to choose a gauge where they do not appear, but *not without breaking manifest Lorentz invariance*. For example, in axial gauge we get

$$-\frac{1}{2\lambda} (r^\mu A_\mu^a)^2 + (r_\mu \bar{c}^a) (\delta^{ac} \partial_\mu + g f^{abc} A_\mu^b) c^c \quad (3.13.12)$$

and so in the limit  $\lambda \rightarrow 0$ , we are forced to have  $r \cdot A = 0$ , and the ghosts decouple. We do not use this gauge for computations because it breaks manifest Lorentz invariance, and also gives the gauge boson propagator a bad high-energy behavior.

The Y-M gauge boson propagator in  $R_\xi$  gauge from Faddeev-Popov is

$$i \frac{-g^{\mu\nu} + (1 - \xi) \frac{p^\mu p^\nu}{p^2}}{p^2 + i\epsilon} \delta^{ab} \quad (3.13.13)$$

which is just the photon propagator times  $\delta^{ab}$ , which means that the propagator is diagonal in the non-Abelian gauge group indices, as one would expect.

### 3.13.1 BRST Symmetry

In the case of QED, we saw that although the gauge fixing term broke gauge symmetry, we were still able to show that correlators and scattering amplitudes are gauge invariant. The reason is that our gauge fixing only modified the photon propagator, and the interactions involved  $A_\mu J^\mu$ , which is gauge invariant because  $\partial^\mu J_\mu = 0$ . For scattering amplitudes, we also used the fact that we could relate  $A_\mu$  to the conserved electromagnetic current  $J_\mu$  (for LSZ).

But in non-abelian theories neither of these tricks work, since the current isn't gauge invariant. This is a major problem, because it suggests that when we renormalize the theory, we may generate UV sensitive gauge symmetry breaking terms, which would spoil the whole theory. Also, in non-abelian theories it's somewhat unclear how to separate the gauge invariant data from the gauge redundancy dependent stuff.

Fortunately, there is a *global symmetry* of the gauge fixed YM and QED Lagrangians that protects the quantum theory from generating gauge-symmetry breaking operators. This is called BRST symmetry, and it exists in both gauge-fixed QED and YM.

For the QED case, consider the gauge-fixed Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{2\xi}(\partial^\mu A_\mu)^2 - \bar{c}\square c \quad (3.13.14)$$

Let us introduce a Grassman number  $\theta$ , and view  $\theta c(x)$  as a bosonic gauge transformation parameter. Then under the transformation

$$A_\mu \rightarrow A_\mu + \frac{1}{g}\theta\partial_\mu c \quad (3.13.15)$$

we generate the terms

$$(\partial^\mu A_\mu)^2 \rightarrow (\partial^\mu A_\mu)^2 + \frac{2}{g}(\partial^\mu A_\mu)(\theta\square c) + \frac{1}{g^2}(\theta\square c)(\theta\square c) \quad (3.13.16)$$

The last term vanishes because  $\theta^2 = 0$ . Thus if we also perform

$$\bar{c} \rightarrow \bar{c} - \frac{1}{g}\theta\frac{1}{\xi}\partial^\mu A_\mu \quad (3.13.17)$$

then the Lagrangian is entirely invariant. This is the abelian version of BRST symmetry. The non-abelian case is more complicated, but similar, and it can be used to constrain renormalization, proving that only a finite number of BRST-invariant counter-terms are needed to renormalize YM theories.

BRST is also associated with a cohomology theory, because the BRST transformation satisfies  $\Delta^2 = 0$ , meaning that we can consider the (closed) states with  $\Delta\psi = 0$  up to differences of the form  $\Delta\chi$  (exact states); these are the closed states modulo the exact states. This gives a precise way to characterize the field configurations that are gauge invariant.

### 3.14 Renormalization in YM and Asymptotic Freedom

For the Feynman rules in YM theory with fermions and scalars, see the Schwartz book, section 26.1 and 26.2.

The renormalized Lagrangian can be written as

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}Z_3(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)^2 - g_R Z_{A^3} f^{abc} \partial_\mu A_\nu^a A^{\mu b} A^{\nu c} - \frac{1}{4}g_R^2 Z_{A^4} (f^{eab} A_\mu^a A_\nu^b)(f^{ecd} A^{\mu c} A^{\nu d}) \\ & + Z_2 \bar{\psi}_i (i\not{D} - Z_m m_R) \psi_i - Z_{3c} \bar{c}^a \square c^a + g_R Z_1 A_\mu^a \bar{\psi}_i \gamma^\mu T_{ij}^a \psi_j + g_R Z_{1c} f^{abc} (\partial_\mu \bar{c}^a) A_\mu^b c^c \end{aligned} \quad (3.14.1)$$

where we have an abundance of  $Z$  factors to allow us to define renormalized fields and couplings.

Just as gauge invariance and the Ward identity demand charge universality in QED, in QCD we must have

$$\frac{Z_1}{Z_2} = \frac{Z_{1c}}{Z_{3c}} = \frac{Z_{A^3}}{Z_3} = \sqrt{\frac{Z_{A^4}}{Z_3}} \quad (3.14.2)$$

as exact non-perturbative statements. Similar equalities hold if we have more species of charged particles, perhaps in more general representations of the  $SU(N)$  gauge group. This means that the strength of the coupling of any charged representation to the YM field  $A_\mu^a$  must be universal, given in terms of the single number  $g_R$ .

### 3.14.1 Vacuum Polarization and the $\beta$ Function

In QCD, there are four types of diagrams, a quark loop, two gauge boson loops, and a ghost loop. The book also draws a diagram for the sum of contributions from counter-terms, which are treated in perturbation theory.

In renormalized perturbation theory we can write the coupling of a fermion to the gauge field as

$$\mu^{\frac{4-d}{2}} g_R Z_1 A_\mu^a \bar{\psi}_i \gamma^\mu T_{ij}^a \psi_j = \mu^{\frac{4-d}{2}} g_R \frac{Z_1}{Z_2 \sqrt{Z_3}} A_{\mu(0)} \bar{\psi}_i^{(0)} \gamma^\mu T_{ij}^a \psi_j^{(0)} \quad (3.14.3)$$

where we are explicit about the scale dependence from dim reg. Thus the bare coupling, which must be independent of the renormalization scale, is

$$g_0 = \mu^{\frac{4-d}{2}} g_R \frac{Z_1}{Z_2 \sqrt{Z_3}} \quad (3.14.4)$$

Differentiating with respect to  $\mu$  and expanding in perturbation theory tells us that

$$\beta(g_R) = \mu \frac{d}{d\mu} g_R = g_R \left[ \left( -\frac{\epsilon}{2} \right) - \mu \frac{d}{d\mu} \left( \delta_1 - \delta_2 - \frac{1}{2} \delta_3 \right) \right] \quad (3.14.5)$$

with  $\epsilon = d - 4$ . But the  $\delta_i$  only depend on  $\mu$  through their  $g_R$  dependence. So we can solve this relation perturbatively giving

$$\beta(g_R) = -\frac{\epsilon}{2} g_R + \frac{\epsilon}{2} g_R^2 \frac{\partial}{\partial g_R} \left( \delta_1 - \delta_2 - \frac{1}{2} \delta_3 \right) \quad (3.14.6)$$

Using the 1-loop counter-term values obtained from modified minimal subtraction, we find that

$$\beta(g_R) = -\frac{\epsilon}{2} g_R - \frac{g_R^3}{16\pi^2} \left[ \frac{11}{3} C_A - \frac{4}{3} n_f T_F \right] \quad (3.14.7)$$

Note that the  $\xi$  dependence cancels, giving a gauge invariant result.

We could have also obtained the same result from

$$\beta(g_R) = -\frac{\epsilon}{2}g_R + \frac{\epsilon}{2}g_R^2 \frac{\partial}{\partial g_R} \left( \delta_{A^3} - \frac{3}{2}\delta_3 \right) \quad (3.14.8)$$

but this gives exactly the same result, due to gauge invariance. There is only one charge setting the strength of the strong interactions.

Specializing to QCD in the standard model, we have  $N = 3$ , so  $C_A = 3$ , and we write  $\alpha_s = \frac{g_s^2}{4\pi}$ . We also have  $T_F = 1/2$ , so at one loop we have

$$\mu \frac{d}{d\mu} \alpha_s = -\frac{\alpha_s^2}{2\pi} \left( 11 - \frac{2n_f}{3} \right) \quad (3.14.9)$$

In nature we have  $n_f = 6$ , but this varies with scale, because they drop out of the RG below the scale of their masses. However, as long as  $n_f < 17$ , note that the  $\beta$  function for  $\alpha_s$  is negative, which means that *the strong couplings gets stronger at low energies, and weaker at high energies*, unlike electromagnetism.

In particular, since  $\alpha_s \rightarrow 0$  at very high energies, QCD exhibits *asymptotic freedom*, which just means that the quarks and gluons are free at high energies. Conversely, QCD gets very strong at a fixed scale  $\Lambda_{QCD}$ , where we can solve the RG to write

$$\alpha_s = \frac{2\pi}{11 - \frac{2n_f}{3}} \frac{1}{\log \frac{\mu}{\Lambda_{QCD}}} \quad (3.14.10)$$

Measuring  $\alpha_s$  at any scale  $\mu$  determines the QCD scale, which in reality is very roughly near 1 GeV. The QCD scale sets the characteristic energy scale of strong-force bound states, such as the proton.

### 3.15 Higgs Mechanism

We can combine our knowledge of Goldstone bosons and non-Abelian gauge theories to understand how to describe massive interacting spin 1 bosons.

Let us try to imagine a theory of massive, charged (interacting) spin 1 particles. In such a theory, one would naively guess that at very high energies  $E \gg m_A$ , the masses of the spin 1 particles should be un-important for the physics. This is certainly true for fermions and scalar particles – at high energies their masses have a negligible effect on physics.

However, with spin 1 particles we run into a problem – we know that massive spin 1 particles have 3 polarization states, while massless spin 1 particles have 2 polarizations. So no matter large  $E/m_A$ , the massive theory must be qualitatively different from the massless theory. However, the only interaction that mixes the 2 transverse polarization states with the longitudinal polarization is

$$m_A^2 A_\mu A^\mu \quad (3.15.1)$$

So at very high energies (or on very short time scales), the transverse and longitudinal polarizations do not mix, and we could separately talk about the transverse and longitudinal polarization.

This suggests that *theories of massive spin 1 bosons must combine the transverse polarizations of a massless spin 1 with some other degrees of freedom that provide the longitudinal component*, since at high energies these two types of degrees of freedom become independent. This is the Higgs mechanism.

Furthermore, note that  $\partial_\mu \pi$  for a scalar  $\pi(x)$  has the correct Lorentz transformation properties to create the longitudinal mode at high-energies. This suggests that scalar degrees of freedom are the ingredient we need. To see this in more detail, simply recall that the polarization states of a massive spin 1 particle with  $p^\mu = (\sqrt{p^2 + m^2}, 0, 0, p)$  are

$$\epsilon_1 = (0, 1, i, 0) \quad (3.15.2)$$

$$\epsilon_2 = (0, 1, -i, 0) \quad (3.15.3)$$

$$\epsilon_L = \frac{1}{m} \left( p, 0, 0, \sqrt{p^2 + m^2} \right) \sim \frac{p^\mu}{m} \quad (3.15.4)$$

where the last approximate equality holds when  $p/m$  is large. This is exactly what we obtain from the Fourier transform of  $\partial_\mu \pi(x)$  for a scalar  $\pi(x)$ . In fact, as we will see, these scalars must be (what would have been) Goldstone bosons.

### 3.15.1 Abelian Case and Superconductivity

Superconductors are materials in which the photon has a mass. All of their phenomenological properties can be derived from this fact. We won't be studying them immediately, however, because the goldstone bosons that the photon 'eats' are particles made from pairs of electrons, and this is more complicated than the simple Abelian Higgs model with which we will begin.

Let us study a theory, the 'Abelian Higgs Model', with Lagrangian

$$L = -\frac{1}{4}F_{\mu\nu}^2 + (\partial_\mu \phi^\dagger - ieA_\mu \phi^\dagger)(\partial_\mu \phi + ieA_\mu \phi) + \frac{\lambda}{4} \left( |\phi|^2 - \frac{v^2}{2} \right)^2 \quad (3.15.5)$$

In this theory,  $\phi$  gets a VEV, and so we can parameterize it as

$$\phi(x) = \frac{1}{\sqrt{2}}(v + H(x))e^{i\frac{\pi(x)}{v}} \quad (3.15.6)$$

Plugging this into the Lagrangian gives

$$\begin{aligned} L &= -\frac{1}{4}F_{\mu\nu}^2 + \frac{(v + H)^2}{2} \left( -i\frac{\partial_\mu \pi}{v} + \frac{\partial_\mu H}{v + H} - ieA_\mu \right) \left( i\frac{\partial_\mu \pi}{v} + \frac{\partial_\mu H}{v + H} + ieA_\mu \right) - \frac{\lambda}{4} \left( \sqrt{2}Hv + \frac{H^2}{2} \right)^2 \\ &\approx -\frac{1}{4}F_{\mu\nu}^2 + \frac{1}{2}(\partial H)^2 + \frac{1}{2}(\partial_\mu \pi + m_A A_\mu)^2 - \frac{\lambda v^2}{2}H^2 + \dots \end{aligned} \quad (3.15.7)$$

where the masses are

$$m_A = ev \quad (3.15.8)$$

$$m_H = \lambda v \quad (3.15.9)$$

for the photon and the ‘Higgs’ field. Note that this means that we can completely decouple the Higgs field by sending  $\lambda \rightarrow \infty$  with  $v$  fixed. Then we are just left with  $\pi$  and  $A_\mu$ .

Now what happens under gauge transformations? If we shift

$$A_\mu \rightarrow A_\mu + \partial_\mu \alpha(x) \quad (3.15.10)$$

where we choose

$$\alpha(x) = -\frac{1}{m_A} \pi(x) \quad (3.15.11)$$

then no matter what value  $\pi(x)$  takes, we can eliminate it from the Lagrangian, leaving

$$L = -\frac{1}{4} F_{\mu\nu}^2 + \frac{1}{2} m_A^2 A_\mu^2 + \dots \quad (3.15.12)$$

Note that this is a gauge choice, called *Unitarity Gauge*. It has this name because it makes it clear what the menu of particles are in our theory – in this case, we have a single massive photon. But  $\pi(x)$ , which wanted to be a goldstone boson, has vanished.

We can also obtain this gauge choice in the original theory by noting that since

$$\phi(x) \rightarrow e^{ie\alpha(x)} \phi(x) \quad \text{or} \quad \frac{1}{\sqrt{2}}(v + H(x))e^{i\frac{\pi(x)}{v}} \rightarrow \frac{1}{\sqrt{2}}(v + H(x))e^{i\frac{\pi(x)}{v} + ie\alpha(x)} \quad (3.15.13)$$

So we can specify a gauge by using  $\phi(x)$  instead of  $A_\mu$ , by stipulating that

$$\phi(x) = \frac{1}{\sqrt{2}}(v + H(x)) \quad (3.15.14)$$

exactly. This leads again to the choice we made above for  $\alpha(x)$ . Note that from this point of view, if  $e = 0$  so that the photon is decoupled, then we would simply have had a goldstone boson  $\pi(x)$  associated to the spontaneous breaking of a global  $U(1)$ . But with  $e \neq 0$  the Lagrangian wasn’t just invariant under

$$\phi(x) \rightarrow e^{i\alpha} \phi(x) \quad \text{but under} \quad \phi(x) \rightarrow e^{i\alpha(x)} \phi(x) \quad (3.15.15)$$

or in other words, we do not have a global symmetry, but a local gauge redundancy. So the redundancy allowed us to make a choice of gauge that ‘rotated away’ or ‘ate’ the Goldstone mode  $\pi(x)$ , eliminating it from the spectrum. Instead of spontaneously breaking a global symmetry and getting a Goldstone boson, we have *spontaneously broken the gauge symmetry* (redundancy), and so we do not have a Goldstone boson, but a massive photon.

Note that the ‘Higgs Boson’ of the Higgs mechanism is the mode  $H(x)$ , not  $\pi(x)$ . And in fact we do not need  $H(x)$  in our abelian theory – the photon ate  $\pi(x)$ , not  $H(x)$ , which we were able to decouple. In fact, all we needed for the Higgs mechanism were the goldstone modes, which we could have represented in a non-linear sigma model (a theory like the chiral Lagrangian, where we only have the  $\pi(x)$  mode around). In the Non-Abelian theory we will discuss next, we cannot

simply have massive interacting spin 1 boson by themselves, because at high energies the theory becomes inconsistent. However, this is not because of the transverse polarizations, but because of the longitudinal goldstone-like polarizations. The real problem is that the non-linear sigma model breaks down at high energies (is not UV or short-distance complete) because it has derivative interactions suppressed by the scale  $v$ .

So why is this a theory of superconductivity? The expulsion of magnetic fields, called the Meissner Effect, is obvious because they would have a very large amount of energy due to the  $m_A^2 A_\mu^2$  (note that  $A_\mu \sim xB$  so  $A_\mu$  would grow spatially). We also see that the  $A_\mu$  field naturally falls off exponentially with distance due to the mass, this gives the penetration depth.

In general, it's possible that  $\langle\phi(x)\rangle$  might be  $x$ -dependent, and in particular, there can be regions where  $\langle\phi(x)\rangle = 0$ . This can happen if we apply a lot of magnetic field to the superconductor, so that it becomes energetically favorable to let magnetic field lines penetrate in little bundles or 'vortices'. Depending on the relationship of the 'Higgs mass' (mass associated with Cooper pairs) to the photon mass, we have type I vs type II superconductors, which are primarily differentiated by how vortices behave.

### 3.15.2 Non-Abelian Higgs Mechanism

Now let us discuss the Non-Abelian version.

Let us imagine that we have a theory such that when the gauge coupling  $g = 0$ , we would have a global symmetry with scalar fields transforming as

$$\phi_i(x) \rightarrow U_i^j \phi_j(x) \quad (3.15.16)$$

Now if that symmetry were spontaneously broken, so that  $\langle\phi_i\rangle = v_i$ , then we have a set of massless goldstone modes, plus other degrees of freedom that are generically massive. So we can write

$$\phi_i(x) = e^{i\pi^a(x)t_{ij}^a}(v_j + h_j(x)) \quad (3.15.17)$$

where we have separated out the goldstone modes  $\pi^a(x)$  from the massive modes  $h_j(x)$  by rotating the  $\phi_i(x)$  to a canonical basis.

Now let us turn on the gauge coupling  $g$ , so that the global symmetry suddenly becomes a gauge redundancy. This means that the transformation

$$\phi_i(x) \rightarrow \left[ e^{i\pi^a(x)t_{ij}^a} \right]^{-1} \phi_i(x) = v_j + h_j(x) \quad (3.15.18)$$

that rotated the fields to the special basis  $v_j$  is just a part of the gauge redundancy. This means that the goldstone boson fields  $\pi^a(x)$  are no longer physical fields. But it also means that we have eliminated a part of the redundancy by fixing it using the  $\phi_i$  VEVs. We have chosen *unitarity gauge*.

Given that the gauge redundancy is the same thing as the original global symmetry, it must be that the Lagrangian contains terms

$$\frac{1}{2} |\partial_\mu h_i(x) - ig T_{ij}^a A_\mu^a (v_j + h_j(x))|^2 \supset \frac{1}{2} g^2 v_i v_j T_{ik}^a T_{kj}^b A_\mu^a A^{b\mu} \quad (3.15.19)$$

which gives the gauge bosons a mass term. For each would-be massless goldstone boson there is a massive spin 1 boson. Note that

$$\partial_\mu h_i T_{ij}^a A^{a\mu} (v_j + h_j(x)) = 0 \quad (3.15.20)$$

by definition of the Higgs and goldstone directions (they are orthogonal), so the Higgs bosons do not mix with the gauge bosons.

Let us consider an example with an  $SO(3)$  gauge group and an  $SO(3)$  fundamental scalar field. It gets a VEV

$$\langle (\phi_1, \phi_2, \phi_3) \rangle = (0, 0, v) \quad (3.15.21)$$

The gauge boson acts in this representation as

$$A_\mu^a T_{ij}^a = \begin{pmatrix} 0 & -A_\mu^3 & A_\mu^2 \\ A_\mu^3 & 0 & -A_\mu^1 \\ -A_\mu^2 & A_\mu^1 & 0 \end{pmatrix} \quad (3.15.22)$$

and the mass term looks like

$$\frac{g^2 v^2}{2} (A_\mu^1 A^{\mu 1} + A_\mu^2 A^{\mu 2}) \quad (3.15.23)$$

Thus we get two massive gauge bosons and one massless gauge boson. What about the remaining massless ‘photon’ (we call it that because it’s a  $U(1)$ ) – how does it interact?

The Higgs field, which was the  $\phi_3$  direction of  $\phi_i$ , will not be charged under the remaining  $A_\mu^3$  – this must be the case because it’s merely a real scalar field, and those cannot have charge. However, note that we can write

$$W_\mu^\pm = A_\mu^1 \pm A_\mu^2 \quad (3.15.24)$$

has charge  $\pm 1$  with respect to the remaining  $A_\mu^3$  photon. So the massive gauge bosons are charged. In fact, this model was once proposed as a potential description of the weak interactions.

It’s worth noting, as the book does, that one can break  $SU(5)$  to  $SU(3) \times SU(2) \times U(1)$  via an adjoint Higgs field

$$v_a T^a = \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix} \quad (3.15.25)$$

That’s how grand unification works. Note that this leads to massive ‘X-Y’ gauge bosons at the GUT scale, leaving our Standard Model gauge bosons massless.



### 3.15.3 Renormalizability and UV Completeness

Historically, proving the renormalizability of theories of massive spin 1 boson with non-abelian interactions was very important. In part this was because renormalizability was viewed as necessary for the consistency of QFT, although we have seen that it's not. However, it is important to know how the theory behaves at short distances, whether more counter-terms will be needed to absorb short-distance sensitivity, whether the pattern of massive and massless particles is consistent and natural, etc.

The naive problem is that a massive vector boson propagator would appear to be (in Unitarity gauge)

$$\frac{i}{p^2 - m_A^2} \left( -g^{\mu\nu} + \frac{p^\mu p^\nu}{m_A^2} \right) \delta^{ab} \quad (3.15.26)$$

Why is this bad? The reason is that this seems to behave as a constant as  $p \rightarrow \infty$ , instead of falling off as  $1/p^2$ . If true, this would mean that loop corrections are completely unsuppressed, no matter how many propagators are included. Thus one would have an extreme amount of short-distance sensitivity, and the theory would be completely unmanageable.

The technical resolution to this is to restore the gauge redundancy, and then eliminate it in a different way. This gives rise to

$$L = -\frac{1}{4}(F_{\mu\nu}^a)^2 + \frac{1}{2}m_A^2 \left( A_\mu^a + \frac{\partial_\mu \pi^a}{m_A} \right)^2 - \frac{1}{2\xi} (\partial^\mu A_\mu^a - \xi m_A \pi^a)^2 + \bar{c}^a (-\partial^\mu D_\mu + \xi m_A^2) c^a$$

when we use  $\partial^\mu A_\mu^a - \xi m_A \pi^a$  as our gauge fixing functional and repeat our Faddeev-Popov derivation. This gives a vector boson propagator

$$\frac{i}{p^2 - m_A^2} \left( -g^{\mu\nu} + \frac{p^\mu p^\nu}{p^2 - \xi m_A^2} (1 - \xi) \right) \delta^{ab} \quad (3.15.27)$$

But we also have fictitious goldstone bosons and ghosts present in the theory, with fictitious masses  $\xi m_A^2$ . The advantage is that we maintain manifest renormalizability. If we take  $\xi \rightarrow \infty$  these extra fictitious states disappear, and we're back to unitarity gauge. The most convenient gauge for calculations is  $\xi = 1$  though, where propagators become very simple.

These gauge tricks do not completely resolve issues with renormalizability. However, renormalizability is really a question of whether the theory needs a short-distance completion. But at very short distances the longitudinal and transverse polarizations of the vector bosons decouple, and we approach the regime of applicability of the *Goldstone boson equivalence theorem*. This theorem says that the high-energy behavior of longitudinal gauge boson polarizations is identical to the high energy behavior of the would-be Goldstone bosons in a theory where the gauge coupling  $g = 0$ . In particular, *if the theory of the Goldstone bosons is UV complete, so that it makes sense to arbitrarily short distances, then the same will be true of the massive vector boson theory. If not, the theory will be incomplete, but this is entirely a problem of UV completing the Goldstone theory.* The transverse polarizations of the vector bosons in YM theories will always be well-behaved on their own.

Note that this means that any theory consisting only of massive non-abelian gauge bosons, without any other particles, *must be incomplete*. This is because an interacting non-linear sigma model (any model of interacting goldstone bosons) will be non-renormalizable, since derivative interactions are always non-renormalizable (the lowest dimension operator would be something of the schematic form  $\pi(\partial\pi)^2/F_\pi$ , and this already dimension 5).

This simple observation has important implications. First of all, it means that if we get our Goldstone bosons from a linear sigma model, ie a standard ‘Higgs’ theory with a  $V(h) = -m^2 h^2 + \lambda h^4$ , then since that scalar theory is renormalizable, our theory of Higgsed massive vector bosons will also be renormalizable. However, we needn’t get Goldstone modes in this way – for example, we saw earlier that in QCD, the pion is an approximate Goldstone boson. If we had a theory of QCD with massless quarks, but the same pattern of spontaneous  $SU(N)_L \times SU(N)_R$  chiral symmetry breaking, then the resulting degrees of freedom could be used to produce the longitudinal modes of an  $SU(N)$  vector boson theory. In fact, this was a (now-defunct) proposal for the  $W$  and  $Z$  bosons, called technicolor. In such a theory, there would be no Higgs boson at all, but only the massive  $W$  and  $Z$ .

### 3.15.4 The Higgs Mechanism and Fermions

It’s often said that the Higgs mechanism gives rise to fermion masses. This is true in the Standard Model, but it’s something of a side effect. One can have models where fermion masses appear directly, where they arise from the spontaneous breaking of a global symmetry, or where they arise from the Higgs mechanism (giving gauge bosons a mass at the same time).

The idea is the following. Consider a theory where there are left handed spinor fermions  $Q_i$  (it’s easiest to think of them as Weyl spinors) that transform as a doublet (fundamental) under  $SU(2)$ . But assume that there are no right-handed spinor doublets; instead we have left-handed spinor singlets  $U$  and  $D$ . Then it’s impossible to write an  $SU(2)$  invariant mass term, because it would have to involve both  $Q$  and either  $U$  or  $D$ , but this is forbidden by symmetry. To be specific,  $Q^\dagger Q, U^\dagger U, D^\dagger D$  could be  $SU(2)$  invariant, but they are prohibited by Lorentz invariance (recall mass terms must couple two left-handed or two right-handed spinors), while terms like  $QU$  or  $Q^\dagger D^\dagger$  are prohibited by  $SU(2)$ .

However, if we have an  $SU(2)$  double scalar field  $H_i$ , then we can write down

$$L \supset \lambda_U Q_i H_i^\dagger U + \lambda_D \epsilon^{ij} Q_i H_j D \quad (3.15.28)$$

So far, these are not mass terms, but are simply yukawa couplings. However, if  $H_i$  gets a VEV  $\langle H_i \rangle = v_i$ , then these Yukawa couplings turn into mass terms! This is how all of the elementary fermions in the Standard Model get masses.

Note that the  $SU(2)$  could have been either a global symmetry or a gauge redundancy. In the SM it happens to be the  $SU(2)$  Weak-force gauge redundancy. We can also use a wide variety of symmetries besides  $SU(2)$  in the same way. However, this idea can only work in a *chiral* theory, where we treat left and right-handed spinors differently. Without a chiral symmetry (which breaks parity) one cannot prohibit fermion masses.

For a nice discussion of the electroweak theory itself, where these ideas are concretely realized, see Weinberg chapter 21.3.

### 3.16 Lattice Gauge Theory, QCD, and Confinement

We have discussed Wilson loops, but let us now discuss their physical relationship with confinement.

When we studied QED we were able to probe the potential by looking at

$$V(r) = \langle \Omega | T \{ J^0(r) J^0(0) \} | \Omega \rangle \approx \frac{e^2}{4\pi r} \quad (3.16.1)$$

where  $J^0(r) = \delta^3(\vec{r} - \vec{x})$  and  $J^0(0) = \delta^3(\vec{x})$ . We computed quantum corrections to this quantity, namely the vacuum polarization Feynman diagram, and we saw how the electromagnetic force changes with scale. However, we cannot compute a directly analogous quantity in QCD because it wouldn't be gauge invariant.

However, we can probe the potential by devising a suitable setup. So consider an experiment where we create a pair of charges, pull them apart to some finite distance  $D$ , let them exist for a very long time  $T$ , and then re-annihilate them. In the limit where the charges act like fixed external classical sources, the action for this process is exactly given by the expectation value of a rectangular Wilson loop with space-like sides of length  $D$  and time-like sides of length  $T$ . Thus if we can compute Wilson loop expectation values we can probe the potential in the theory!

So the formal claim is that

$$V(R) = \lim_{T \rightarrow \infty} \frac{1}{iT} \log \langle \Omega | \text{Tr} \{ W_{loop} \} | \Omega \rangle \quad (3.16.2)$$

where  $W$  is a Wilson loop (path ordered) integrated along a rectangle with sides of length  $R$  and  $T$  in space and time, respectively. We can write it as

$$W_{loop} = P \left\{ e^{ig \oint_P A_\mu^a T_{ij}^a dx^\mu} \right\} \quad (3.16.3)$$

We can compute this in any representation; for particles like quarks we would use the fundamental, but we could also use the adjoint or another representation.

This can be done analytically in  $2 + 1$  dimensional QED; see Polyakov's book 'Gauge Fields and Strings'. One finds that in fact, that theory is a confining gauge theory, meaning that the potential between two charges grows linearly with the distance between charges.

Let us first see how in perturbation theory it gives the correct result. If we choose

$$J^0(x) = \delta(x)\delta(y)\delta(z - R) - \delta(x)\delta(y)\delta(z) \quad (3.16.4)$$

where we turn this on and off at times  $\pm \frac{T}{2}$ , and connect up the 'ends' to make the Wilson loop. Thus we write

$$e^{iET} = \langle \Omega | e^{iHT} | \Omega \rangle = \frac{\int \mathcal{D}A e^{iS[A] + \int e A_\mu J^\mu}}{\int \mathcal{D}A e^{iS[A]}} \quad (3.16.5)$$

We can compute this exactly in QED, to obtain

$$e^{iET} = e^{i \int d^4x d^4y \frac{e^2}{2} J^\mu(x) D_{\mu\nu}(x, y) J^\nu(y)} \quad (3.16.6)$$

where  $D_{\mu\nu}$  is the position space Feynman propagator for the photon. There are  $R$ -independent divergences (associated to self-energies of charges), but the only  $R$  dependence at large  $T$  comes from the two opposite sides of the loop, giving

$$iET = -\frac{e^2}{4\pi} \int_{-T/2}^{T/2} dt \int_{-\infty}^{\infty} dt' \frac{1}{(t-t')^2 - R^2 - i\epsilon} = i\frac{e^2 T}{4\pi R} \quad (3.16.7)$$

giving the expected  $E = V(R)$ . We took  $T \rightarrow \infty$  in the second integral for simplicity, since this does not alter the leading  $T$  behavior.

A similar computation has been performed at 1-loop in pure YM (and at higher-loops), giving

$$\tilde{V}(\vec{q}) = -C_F \frac{g^2}{\vec{q}^2} \left( 1 + \frac{g^2}{16\pi^2} \left( \frac{11}{3} C_A \log \frac{\vec{q}^2}{\mu^2} + \dots \right) + \dots \right) \quad (3.16.8)$$

This is a gauge-invariant result that shows the effect of the running coupling on the strength of the YM (or in the case of QCD, the strong) force.

### 3.16.1 Lattice Gauge Theories

Now let us discuss how one can compute without using perturbation theory, by turning QFT into a finite system and using a computer (although we won't be writing any programs). We will see is that one can derive suggestive non-perturbative results via a rough analytic argument.

Let us make spacetime into a 4d lattice with  $n_{sites}$  sites in each direction, with a spacing  $a$ . This reduces QFT to QM with  $n_{sites}^4$  degrees of freedom. As we saw long ago in the case of a scalar field, we can naturally put matter fields like  $\phi(x) \rightarrow \phi(n_\mu)$ , where  $n_\mu$  labels the lattice points. Notationally, we can denote  $\hat{\mu}$  as the unit vector (length  $a$ ) in the  $\mu$  direction, so that  $\phi(n + \hat{\mu})$  and  $\phi(n)$  are nearest neighbors.

Gauge transformations are also discrete and act at lattice sites, so

$$\phi_a(n) \rightarrow U_a^b(n) \phi(n) \quad (3.16.9)$$

would be a gauge transformation where the matrix  $U(n)$  is in some appropriate representation of the gauge group. To compare fields at difference sites, we need discrete Wilson loops, transforming as

$$W_\mu(n) \rightarrow U(n) W_\mu(n) U^\dagger(n + \hat{\mu}) \quad (3.16.10)$$

so that

$$\phi^\dagger(n) W_\mu(n) \phi(n + \hat{\mu}) \quad (3.16.11)$$

is gauge invariant. We can multiply together many of the

$$W_\mu(n) = e^{iaA_\mu(n)} \quad (3.16.12)$$

on successive lattice sites in order to connect distant fields. Thus it's more natural to view  $W_\mu(n)$  as living on the links between lattice sites, and they are called *link fields*. We have absorbed the coupling  $g$  into the  $A_\mu(n)$  field.

The gauge invariant YM field operators are Wilson loops. The simplest non-trivial example is a *plaquette* which is defined as

$$\begin{aligned} W_{\mu\nu}(n) &= W_{-\nu}(n + \hat{\nu})W_{-\mu}(n + \hat{\mu} + \hat{\nu})W_{\nu}(n + \hat{\mu})W_{\mu}(n) \\ &= W_{\nu}^{\dagger}(n)W_{\mu}^{\dagger}(n + \hat{\nu})W_{\nu}(n + \hat{\mu})W_{\mu}(n) \end{aligned} \quad (3.16.13)$$

We can connect this with the continuum by using

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\dots} \quad (3.16.14)$$

so that to order  $a^2$  we have

$$\begin{aligned} \log W_{\mu\nu}(n) &= ia(A_{\mu}(n) + A_{\nu}(n + \hat{\mu}) - A_{\mu}(n + \hat{\nu}) - A_{\nu}(n)) \\ &\quad + \frac{a^2}{2} [A_{\nu}(n) + A_{\mu}(n + \hat{\nu}), A_{\nu}(n + \hat{\mu}) + A_{\mu}(n)] \\ &\quad - \frac{a^2}{2} [A_{\nu}(n), A_{\mu}(n + \hat{\nu})] - \frac{a^2}{2} [A_{\nu}(n + \hat{\mu}), A_{\mu}(n)] \end{aligned} \quad (3.16.15)$$

Now we can write

$$A_{\nu}(n + \hat{\mu}) = A_{\nu}(n) + a\partial_{\mu}A_{\nu}(n) + \dots \quad (3.16.16)$$

to compute that

$$W_{\mu\nu}(n) = e^{ia^2 F_{\mu\nu}(n) + \dots} \quad \text{with} \quad F_{\mu\nu}(n) = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} - i[A_{\mu}, A_{\nu}] \quad (3.16.17)$$

This means that we can expand to see that

$$W_{\mu\nu}(n) = 1 + ia^2 F_{\mu\nu}(n) - \frac{a^4}{2} F_{\mu\nu}(n) F_{\mu\nu}(n) + \dots \quad (3.16.18)$$

where we recall that these are matrices, so the third term is a product of gauge-group matrices. We can therefore write the lattice action

$$S[W_{\mu\nu}] = -\frac{i}{2g^2 N} \sum_{n,\mu,\nu} \text{Re Tr} [1 - W_{\mu\nu}(n)] \quad (3.16.19)$$

Now one can perform the path integral by directly summing over lattice field configurations in terms of the link fields  $W_{\mu}(n)$ .

One can use this to compute correlators by inserting operators. For example, by computing 2-pt functions and looking at their  $x$ -dependence, we can extract the masses of particles such as the pion. As discussed above, we can also probe confinement by computing Wilson loop expectation values.

### 3.16.2 Strong Coupling Expansion and Confinement

Now let us discuss what happens to our Wilson loop computation if we take the limit of strong coupling,  $g \rightarrow \infty$ . The result will be equivalent for any gauge theory, abelian or non-abelian.

We want to evaluate the expectation value of a large Wilson loop with sides  $R$  and  $T$

$$\begin{aligned} \langle W \rangle &= \frac{\int \mathcal{D}A e^{iS[A]} P \left\{ e^{i \oint_P A_\mu^a T_{ij}^a dx^\mu} \right\}}{\int \mathcal{D}A e^{iS[A]}} \\ &= \frac{\prod_n \int dW_\mu(n) e^{\frac{i}{g^2} S[W_\mu(n)]} P \left\{ e^{i \sum_C W_\mu^a(n) T_{ij}^a} \right\}}{\prod_n \int dW_\mu(n) e^{\frac{i}{g^2} S[W_\mu(n)]}} \end{aligned} \quad (3.16.20)$$

where we have normalized the fields so that the coupling appears as an overall factor outside the action. We want to consider the limit  $g^2$  large, which means that we can expand the exponential in powers of  $\frac{1}{g^2}$ . Thus we get

$$\prod_n \int dW_\mu(n) \left[ 1 + \frac{i}{g^2} S[W_\mu(n)] + \frac{1}{2} \left( \frac{i}{g^2} S[W_\mu(n)] \right)^2 + \dots \right] P \left\{ e^{i \sum_C W_\mu^a(n) T_{ij}^a} \right\} \quad (3.16.21)$$

When the loop contour  $C$  is large, the low orders in this expansion all vanish. This is because segments of the loop include many factors of

$$\int dW_\mu(n) e^{iW_\mu(n)} = 0 \quad (3.16.22)$$

which are pure phases, and vanish. It is only when these phases are cancelled that we can obtain a non-vanishing result.

These phases can only be cancelled if we expand to such a high order in  $1/g^2$  that we get a contribution from *all plaquettes inside the Wilson loop*. If one tiles the inside of the loop with plaquettes, all contributions cancel internally, and we are left with the contributions along the perimeter of the loop. These cancel the explicit phases from

$$P \left\{ e^{i \sum_C W_\mu^a(n) T_{ij}^a} \right\} \quad (3.16.23)$$

This requires us to pull down a factor of

$$\left( \frac{1}{g^2} \right)^{TR} = e^{iTR \log(g^2)} \quad (3.16.24)$$

which means that the expectation value of the Wilson loop is proportional to the exponential of its *area*  $= RT$ . Note that we are pulling down one plaquette each inside the loop, so there is no  $1/n!$ , or in other words, the combinatorics of all the possible plaquette contributions cancels this  $1/n!$  with a numerator factor of  $n!$ .

To obtain the potential between two quarks, we equate this with  $e^{iET}$ , and so we learn that

$$E = V(R) = R \log(g^2) \quad (3.16.25)$$

This means that the potential between quarks (charges) grows linearly with their separation. Infinitely separated quarks would require an infinite energy. This is confinement, derived from the strong coupling  $g \rightarrow \infty$  expansion. This argument makes it plausible that when the QCD coupling runs to become large, we enter into this regime, and obtain the confinement of quarks and gluons observed in nature.

### 3.17 Parton Model and Deep Inelastic Scattering

Now we will think simultaneously about the consequences of confinement and jets (our discussion will be rather anachronistic, skipping around in history). Naively, one might expect that confinement and the ‘strength’ of the strong force (QCD) makes it impossible to use perturbation theory for anything. However, due to asymptotic freedom (running of  $\alpha_s$ ), at large energies (or short distances) the quarks and gluons of QCD interact weakly. In what ways can we make use of this?

The dumbest way is through processes like  $e^+e^- \rightarrow$  hadrons. Although the specific products one obtains depend a great deal on the details of QCD, at large center of mass energies, the total differential cross section can be reliably estimated by computing  $e^+e^- \rightarrow q\bar{q}, q\bar{q}g$ , etc, even though quarks and gluons can never be found by themselves. Conceptually, the reason this (should) work is because one first makes a few very high energy quarks in the perturbative regime, and then, as they spread apart, the effective strength of  $\alpha_s$  increases (and there is a lot of soft radiation), until the various partons reach separations of order  $1/\Lambda_{QCD}$ , at which point *hadronization* occurs, and the quarks and gluons form strongly interacting bound states. But since the total cross section is limited by the rate for the very first process, it can be estimated perturbatively, with uncomputable corrections that go like powers of  $\Lambda_{QCD}/E$ .

Once again, it’s a separation of scales that makes it possible to do a reliable perturbative computation. There is in fact a whole effective field theory concerned with these kinds of processes, called Soft-Collinear Effective Theory (SCET). The author of the textbook is a SCET expert, so it is discussed in the book, if you’re curious.

What about scattering of protons, neutrons, and mesons? If the proton was an elementary particle it would scatter off of electrons just as predicted by QED, with

$$\frac{d\sigma}{d\Omega} = \frac{\alpha_e^2}{4E^2 \sin^4 \frac{\theta}{2}} \frac{E'}{E} \left( \cos^2 \frac{\theta}{2} - \frac{q^2}{2m_p^2} \sin^2 \frac{\theta}{2} \right) \quad (3.17.1)$$

in the lab frame, where  $E'$  and  $E$  are the final and initial electron energies. At low energies this is correct. As you derived on a problem set, we can parameterize the general photon vertex with  $F_1$  and  $F_2$ , although it turns out that

$$g_p = 5.58, \quad g_n = -3.82 \quad (3.17.2)$$

so the proton and neutron have magnetic moments that are very different from what one would expect from a truly elementary particle. In fact, just from the form factors one can see that the proton seems to have a characteristic ‘size’ of about  $1/(0.84\text{GeV})$ , so it is not point-like.

In the 1950s, no one knew what this substructure meant. Some interactions of the proton seemed weaker than one might expect. For example, proton-proton collisions at 10 GeV led to the production of large numbers of pions, but they mostly flew in the same direction as the original incident protons, as though the proton itself was breaking apart into blobs flying in the same direction. The probability of making pions that fly off perpendicular to the beam axis was very suppressed. This led to the idea that perhaps the proton is goopy (with many many constituents), and so in collisions it just flies into a cloud of pieces. Mathematically, since the pieces stay collinear, we can characterize this ‘goopy’ hypothesis by the claim that although the nearly light-like momenta  $q$  exchanged among the constituents may be large, the invariant  $q^2$  would always be very small.

These ideas were decisively tested via the 1960s SLAC-MIT *Deep Inelastic Scattering* experiments, which involved 20 GeV electron beams scattering off of protons. The direction of the outgoing electron gave a direct indicator of  $q^2$ . If protons were goop, then all electron deflection angles would be small (think of a bullet going through water, vs two bullets colliding). In fact, large deflection angles often occurred, but protons themselves rarely survived. But the overall rate was roughly in accord with what one would have gotten if the proton was an elementary particle.

To explain these observations, Bjorken and Feynman proposed the *parton model* – that protons are made of a loosely bound assemblage of partons (as we know now, these are quarks and gluons). Electrons just scatter off of one of these quarks, possibly exchanging a large  $q^2$ , but the rest of the proton is just left behind, and propagates collinearly with the proton beam. Only the ejected quark and its hadronization products have large scattering angles. Note that the parton model isn’t really so new – it’s also a good description of a hydrogen atom.

Let us see what this model predicts at the most basic (quantitative) level. The square of the electron-quark matrix element is

$$\frac{1}{4} \sum_{\text{spins}} |M|^2 = \frac{8e^4 Q_i^2}{t^2} \left( \frac{s^2 + u^2}{4} \right) \quad (3.17.3)$$

where  $s, t, u$  are mandelstam invariants, and since we’re at high energy, we ignore all masses. The parameter  $Q_i$  is the quark (parton) charge. This leads to a cross section that can be written as

$$\frac{d\sigma}{dt} = \frac{2\pi\alpha^2 Q_i^2}{s^2} \frac{s^2 + (s+t)^2}{t^2} \quad (3.17.4)$$

using  $t = -s(1 - \cos\theta)/2$  in the CoM frame. We can write  $t = -Q^2$ , where  $Q^2 = -q^2$  is positive. It is directly measurable in terms of the electron’s initial and final momentum.

Writing  $s$  in terms of easily measurable quantities is more difficult. Working in the electron-proton center of mass frame and ignoring masses, we can characterize a given parton by the fraction of the proton’s total momentum that it carries. We call this the longitudinal momentum fraction  $x$ , with  $0 < x < 1$ . For each species of parton, there is a *parton distribution function*  $f(x)$  that expresses the probability that the proton contains a parton of this type with longitudinal fraction  $x$ . The momentum of the parton is  $p = xP$ , where  $P$  is the proton momentum.



Physically, the idea is that the momentum transfers back and forth among the various partons, and so on average there is some probability distribution to have a given kind of parton with a given momentum. This is obviously wrong quantum mechanically – really the proton just has a wavefunction – but the separation of scales between  $Q$  and  $\Lambda_{QCD}$  justifies a decoherent, statistical approach. In fact it's very challenging to prove rigorously that this *factorization* is correct.

If  $k$  is the initial electron momentum then

$$s = (p + k)^2 = 2p \cdot k = 2xP \cdot k = xs_{pe} \quad (3.17.5)$$

where  $s_{pe}$  is the square of the electron-proton center of mass energy. If we assume that the electron-parton scattering is elastic, then we can determine  $x$  via

$$0 \approx (p + q)^2 = 2p \cdot q + q^2 = 2xP \cdot q - Q^2 \quad (3.17.6)$$

so that

$$x = \frac{Q^2}{2P \cdot q} \quad (3.17.7)$$

For each scattered electron, we can determine the value of  $Q^2$  and  $x$  for the process, and the parton model predicts the event distribution in the  $x$ - $Q^2$  plane. Note that  $x$  is determined from  $q$  and  $P$ , which are known.

Using the parton distribution functions  $f_i(x)$ , we have

$$\frac{d^2\sigma}{dx dQ^2} = \sum_i f_i(x) e_i^2 \frac{2\pi\alpha^2}{Q^4} \left( 1 + \left( 1 - \frac{Q^2}{xs} \right)^2 \right) \quad (3.17.8)$$

where the PDFs  $f_i(x)$  depend on the detailed structure of the proton, where  $e_i$  is the charge of parton  $i$ .

But this formula still makes a prediction, namely that when we divide the DIS cross section by the factor in parentheses and multiply by  $Q^4$ , we should get a quantity that depends only on  $x$  but is independent of  $Q^2$ . This is called *Bjorken scaling*, and it was observed to about 10% accuracy above 1 GeV in the SLAC-MIT experiment. Bjorken scaling is essentially the statment that the proton looks the same to an electromagnetic probe no matter how hard the proton is struck, and that the partons themselves act like elementary free particles, without complicated form factors. In the proton frame the energy of the incident (effective) photon is

$$q^0 = \frac{P \cdot q}{m} = \frac{Q^2}{2xm_p} \quad (3.17.9)$$

The inverse of this gives a time scale for interaction. DIS is when  $q^0 \gg m_p$ , so that the scattering is very rapid on the characteristic time scale of the proton. Bjorken scaling assumes that during this very rapid scattering, the interactions among the constituents of the proton can be ignored.

### 3.17.1 PDF Sum Rules

So for any bound state, there are PDFs that indicate how much of each type of particle one can ‘find’ inside. We often say that the proton is a  $uud$  bound state. If it had exactly one down quark, then  $\int dx f_d(x) = 1$ , but in fact this is false, because we can make down anti-down pairs. However, what we do know is that

$$\int dx (f_d(x) - f_{\bar{d}}(x)) = 1 \quad (3.17.10)$$

$$\int dx (f_u(x) - f_{\bar{u}}(x)) = 2 \quad (3.17.11)$$

$$\int dx (f_s(x) - f_{\bar{s}}(x)) = 0 \quad (3.17.12)$$

and similarly for the other quarks, because quark number conservation is an approximate symmetry. This is what we mean when we say that a proton is  $uud$ , it has 2 up and 1 down *valence quark*.

We also have by momentum conservation that

$$\sum_j \int dx x f_j(x) = 1 \quad (3.17.13)$$

since  $xP$  is the momentum fraction of each parton, and the total momentum must be  $P$ . All of these sum rules can be derived from the conservation of a current.

Note that

$$\int dx x (f_u(x) + f_d(x)) \approx 0.38 \quad (3.17.14)$$

so in fact, most of the momentum of the proton is not carried by the quarks! Most of the rest (35 to 50 %) is made up by gluons, and the remainder comes from ‘sea quarks’, so quark anti-quark pairs that can be created.

Actually, as we will now discuss, all of these numbers depend on the energy scale at which we probe the proton. What the proton is made of depends on how you look at it!

The form of the equations for the evolution of the PDFs, often called Alterelli-Parisi or DGLAP equations, is very much like the Boltzmann equations in thermodynamics. You can think of the proton as a box containing quarks and gluons, in which all possible interactions occur. Thus quarks can emit gluons, and gluons can split into quarks. The number of quarks and gluons will be proportional to the number that are produced minus the number that disappear. Equilibrium will be the situation where these effects balance. Finally, you can think of the energy scale (or perhaps, more accurately, its logarithm) at which we probe the proton as analogous to the temperature – as we slowly increase the temperature, we get more and more stuff, and we obtain a different equilibrium balance. This analogy works pretty well for understanding the PDFs and their energy dependent evolution.

### 3.17.2 Equivalent Photons and the ‘Parton Model’ for QED

On your problem set you studied the equivalent photon approximation – the idea that a high energy electron can also be viewed as a source for high-energy, nearly on-shell photons. This is the QED equivalent of the parton model, since it suggests that a single electron really isn’t so simple, but can actually be regarded as a smeared bundle composed of an electron, some photons, and some  $e^+e^-$  pairs.

You showed that there is a photon ‘PDF’

$$f_\gamma(z) = \frac{\alpha}{2\pi} \log \frac{s}{m_e^2} \left( \frac{1 + (1-z)^2}{z} \right) \quad (3.17.15)$$

representing the distribution of photons with momentum  $zp$  ‘inside’ an electron with momentum  $p$ . It can be used to compute

$$\sigma(e^-X \rightarrow e^-Y) = \int_0^1 dz f_\gamma(z) \sigma(\gamma X \rightarrow Y) \quad (3.17.16)$$

so we can get electron scattering from the photon PDF times the cross section for photons to interact.

In fact, we can go on to work out the consequence of these ideas to higher orders. This means we can ask for the probability of finding an electron ‘inside’ an electron, by including the possibility that electrons produce photons that split into electrons, or that free photons contain  $e^+e^-$  pairs. Following the same steps that led to the equivalent photon approximation tells us that

$$f_e(x) = \delta(1-x) + \frac{\alpha}{2\pi} \log \frac{s}{m_e^2} \left( \frac{1+x^2}{(1-x)_+} + \frac{3}{2} \delta(1-x) \right) \quad (3.17.17)$$

There are actually several subtleties here. First of all, the  $\delta(1-x)$  encodes the leading order expectation that we find exactly one electron with momentum  $p$ . This means that the sub-leading term has to be normalized to maintain

$$\int_0^1 dx [f_e(x) - f_{\bar{e}}(x)] = 1 \quad (3.17.18)$$

as expected from sum rules. That is how the second  $\delta(1-x)$  was determined. Relatedly, the first factor contains a  $1/(1-x)$  singularity that has to be regulated; the ‘+’ indicates that we are doing this by

$$\int_0^1 dx \frac{f(x)}{(1-x)_+} \equiv \int_0^1 dx \frac{f(x) - f(1)}{1-x} \quad (3.17.19)$$

Even with this regulator, the distribution is still very singular near  $x = 1$ , and will receive corrections from multiple photons.

In fact, if we look at multiple photon emissions with

$$p_{1\perp} \gg p_{2\perp} \gg p_{3\perp} \gg \dots \quad (3.17.20)$$

we find that the electron propagators and interactions squared and integrated over phase space contribute

$$\left(\frac{\alpha}{2\pi}\right)^k \int_{m_e^2}^s \frac{dp_{1\perp}^2}{p_{1\perp}^2} \int_{m_e^2}^{p_{1\perp}^2} \frac{dp_{2\perp}^2}{p_{2\perp}^2} \int_{m_e^2}^{p_{2\perp}^2} \frac{dp_{3\perp}^2}{p_{3\perp}^2} \dots = \frac{1}{k!} \left(\frac{\alpha}{2\pi}\right)^k \log^k \frac{s}{m_e^2} \quad (3.17.21)$$

This is just our usual exponentiation of soft and collinear emissions. Note that the ordering is required because otherwise there isn't a denominator  $p_{i\perp}^2$  for each  $i = 1, \dots, k$ , and we do not get the leading logarithmic behavior. This is called the strongly ordered region of phase space.

We can interpret this sequence of photon emissions as probing larger and larger values of  $p_{\perp}^2$ , or in other words, they are probing the electron structure at different energy scales. Viewing emission as a continuous process, we can obtain an RG equation for the PDFs.

Thus we let  $f_{\gamma}(x, Q)$  and  $f_e(x, Q)$  be the probabilities of finding a photon or electron of longitudinal momentum fraction  $x$  inside a physical electron, taking into account collinear photon emissions with transverse momenta  $p_{\perp} < Q$ . As we change  $Q$ , we must account for  $Q < p_{\perp} < Q + \delta Q$ . The differential probability for an electron to split off such a photon is just

$$\frac{\alpha}{2\pi} \frac{1 + (1-z)^2}{z} \frac{dp_{\perp}^2}{p_{\perp}^2} \quad (3.17.22)$$

We can use this to compute that

$$\begin{aligned} f_{\gamma}(x, Q + dQ) &= f_{\gamma}(x, Q) + \int_0^1 dx' \int_0^1 dz \left[ \frac{\alpha}{2\pi} \frac{1 + (1-z)^2}{z} \frac{d(Q^2)}{Q^2} \right] f_e(x', p_{\perp}) \delta(x - zx') \\ &= f_{\gamma}(x, Q) + \frac{dQ}{Q} \int_x^1 \frac{dz}{z} \left[ \frac{\alpha}{\pi} \frac{1 + (1-z)^2}{z} \right] f_e\left(\frac{x}{z}, p_{\perp}\right) \end{aligned}$$

This leads to the differential equation

$$\frac{d}{d \log Q} f_{\gamma}(x, Q) = \int_x^1 \frac{dz}{z} \frac{\alpha}{\pi} \frac{1 + (1-z)^2}{z} f_e\left(\frac{x}{z}, Q\right) \quad (3.17.23)$$

Similarly, the electron distribution evolves as

$$\frac{d}{d \log Q} f_e(x, Q) = \int_x^1 \frac{dz}{z} \frac{\alpha}{\pi} \left[ \frac{1 + z^2}{(1-z)_+} + \frac{3}{2} \delta(1-z) \right] f_e\left(\frac{x}{z}, Q\right) \quad (3.17.24)$$

If we integrate these equations from  $Q \sim m_e$  with the boundary condition that at that scale, there are no photons and only one electron at  $x = 1$ , we get PDFs at other energy scales.

There is another effect, namely that of  $\gamma \rightarrow e^+ e^-$ , which you can find in Peskin and Schroder, which leads to combined differential equations for  $f_{\gamma}$ ,  $f_e$ , and  $f_{\bar{e}}$ . Since we aren't going to do the explicit calculation I won't write out the details.

These PDFs can be used to compute reaction rates via

$$\sigma(e^- X \rightarrow e^- + n\gamma + Y) = \int_0^1 dx f_{\gamma}(x, Q) \sigma(\gamma X \rightarrow Y) \quad (3.17.25)$$

$$\sigma(e^- X \rightarrow n\gamma + Y) = \int_0^1 dx f_e(x, Q) \sigma(e^- X \rightarrow Y) \quad (3.17.26)$$

where the cross sections assume a photon or electron with longitudinal momentum fraction  $x$ .

But the central point here is that we have found *renormalization flow* equations for the PDFs that determine the *internal structure* of the electron, where the scale is that at which we probe the electron. The reason for this flow is that there are logarithms that contribute at all scales.

### 3.17.3 DIS and the DGLAP Equations for QCD

Peskin and Schroder's discussion of Parton Distribution Functions in their chapter 17.5 is nicer and more complete than the discussion in Schwartz, and so is recommended.

Let us study electron-proton scattering, while remaining agnostic about the products – so we study  $e^-p^+ \rightarrow e^-X$  for all  $X$ . This is called *Deep Inelastic Scattering* (DIS). However, we will be assuming that the parton model is valid. We can parameterize the process via the scattering angle of the electron and its initial and final energies,  $E$  and  $E'$ , via

$$\frac{d\sigma}{d\Omega dE'} = \frac{\alpha_e^2}{4\pi m_p q^4} \frac{E'}{E} L^{\mu\nu} W_{\mu\nu} \quad (3.17.27)$$

where we are in the lab frame,  $L^{\mu\nu}$  is the leptonic tensor (it encodes information about the electron polarizations) and  $W_{\mu\nu}$  is the much more complicated hadronic tensor, which is what we're really interested in.

For unpolarized beams, the lepton tensor is just

$$L_{\mu\nu} = \frac{1}{2} \text{Tr} [k' \gamma^\mu k \gamma^\nu] = 2(k'^\mu k^\nu + k'^\nu k^\mu - k \cdot k' g^{\mu\nu}) \quad (3.17.28)$$

where  $k$  and  $k'$  are initial and final electron momenta. This came from  $\bar{u}(k')\gamma^\mu u(k)$  squared and summed over polarizations.

The hadronic tensor must necessarily include an integral over the phase space of all possible final state  $X$ . Since the electron actually only probes the proton with a photon, we really have

$$e^2 \epsilon_\mu \epsilon_\nu^* W^{\mu\nu} = \frac{1}{2} \int dX (2\pi)^4 \delta^4(q + P - p_X) |\mathcal{M}(\gamma^* p^+ \rightarrow X)|^2 \quad (3.17.29)$$

where  $\epsilon_\mu$  is an (off-shell) photon polarization. Since we've summed over final states, this can only depend on  $P$  and  $q$ , and it must be symmetric and have  $q_\mu W^{\mu\nu} = 0$  by the Ward identity. Thus

$$W^{\mu\nu} = W_1 \left( -g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) + W_2 \left( P^\mu - \frac{P \cdot q}{q^2} q^\mu \right) \left( P^\nu - \frac{P \cdot q}{q^2} q^\nu \right) \quad (3.17.30)$$

where  $W_1$  and  $W_2$  only depend on  $q^2$  and  $P \cdot q$ , and we recall that  $x = -q^2/(2P \cdot q)$ .

Now let us look at this parton-by-parton. We can write  $\hat{W}_{\mu\nu}(z, Q)$  where

$$z \equiv \frac{Q^2}{2p_i \cdot q} \quad (3.17.31)$$

where  $p_i$  is the  $i$ th parton momentum, so that

$$p_i = \xi P^\mu \quad (3.17.32)$$

with  $0 < \xi < 1$ . In other words  $x = z\xi$ , and our PDFs for the  $i$ th parton are  $f_i(\xi)$ . So

$$\begin{aligned} W^{\mu\nu}(x, Q) &= \sum_i \int_0^1 dz \int_0^1 d\xi f_i(\xi) \hat{W}_i^{\mu\nu}(z, Q) \delta(x - z\xi) \\ &= \sum_i \int_x^1 \frac{d\xi}{\xi} f_i(\xi) \hat{W}_i^{\mu\nu}\left(\frac{x}{\xi}, Q\right) \end{aligned} \quad (3.17.33)$$

At leading order, the only partonic process that contributes is  $\gamma^* q \rightarrow q$ . This has a trivial matrix element (identical to the leptonic  $L_{\mu\nu}$  matrix), which just gives

$$\hat{W}_1 = 2\pi e_i^2 \delta(1 - z) = \frac{Q^2}{4z} \hat{W}_2 \quad (3.17.34)$$

This just means that at leading order, we have a quark with  $z = 1$  identically.

At higher order, a given parton can split into other partons, making a non-trivial contribution to the PDFs  $f_i(x)$ . In particular, there is a loop correction to  $\gamma^* q \rightarrow q$ , and more interestingly, there are graphs for

$$\gamma^* q_i \rightarrow q_i g \quad (3.17.35)$$

$$g \rightarrow q_i \bar{q}_i \quad (3.17.36)$$

$$g \rightarrow gg \quad (3.17.37)$$

where  $i$  now labels the various quarks,  $u, d, s$ , etc. Note that  $\gamma^*$  can only interact directly with the quarks at leading order. This leads to the DGLAP evolution equations

$$\mu \frac{d}{d\mu} \begin{pmatrix} f_i(x, \mu) \\ f_g(x, \mu) \end{pmatrix} = \sum_j \frac{\alpha_s}{\pi} \int_x^1 \frac{d\xi}{\xi} \begin{pmatrix} P_{q_i q_j}(\frac{x}{\xi}) & P_{q_i g}(\frac{x}{\xi}) \\ P_{g q_j}(\frac{x}{\xi}) & P_{gg}(\frac{x}{\xi}) \end{pmatrix} \begin{pmatrix} f_j(\xi, \mu) \\ f_g(\xi, \mu) \end{pmatrix} \quad (3.17.38)$$

conveniently packaged in matrix form. The formulas for the  $P(z)$  functions are given in the book; they are all derived by studying the simple processes mentioned above, and can be found derived in more detail in Peskin and Schroder.

These equations tell us how the PDFs change with energy scale  $\mu$ . As we mentioned above, their form is quite analogous to that of the Boltzmann equations for the distributions of matter in statistical mechanics.

The splitting amplitudes are universal, and are identical to the soft and collinear photon/gluon/etc amplitudes that we discussed when we talked about soft and collinear divergences. This means they also tell us about rates for processes that occur after hard quarks are produced in a high-energy, when they *shower*, producing many lower energy gluons and quarks that form into *jets*. These jets then *hadronize*. So these splitting functions are used as probabilities in *Parton Showers*, which are computer codes that try to model the distribution of jets and radiation in QCD, as the book discusses.

### 3.18 DIS, CFTs, and the OPE

Lots of acronyms – Deep Inelastic Scattering, Conformal Field Theory, and the Operator Product Expansion.

### 3.18.1 Reformulating DIS in terms of Operators

Let's begin again, by studying an electron scattering off of a proton. Recall that we showed that instead of using the photon field  $A_\mu$  to compute photon scattering amplitudes, we can instead just compute correlators of the electric current  $J_\mu$ . At a direct level, this follows because if we couple to spinors then the current has matrix elements

$$\langle p' | J^\mu(x) | p \rangle = \bar{u}_2(p') \gamma^\mu u_1(p) e^{i(p'-p) \cdot x} \quad (3.18.1)$$

which follows immediately since  $J^\mu \supset \bar{\psi}(x) \gamma^\mu \psi(x)$  for fermions.

To study DIS, we are interested in  $\gamma^* p^+ \rightarrow X$ , where  $\gamma^*$  is an off-shell photon. So we can just use matrix elements of this current to get

$$\mathcal{M}(\gamma^* p^+ \rightarrow X) = e \epsilon^\mu \langle X | J_\mu(0) | P \rangle \quad (3.18.2)$$

This means that the hadronic tensor for DIS is just

$$\begin{aligned} W_{\mu\nu}(\frac{1}{x}, Q) &= \int dX \langle P | J_\mu(0) | X \rangle \langle X | J_\nu(0) | P \rangle (2\pi)^4 \delta^4(q^\mu + P^\mu - p_X^\mu) \\ &= \int dX d^4y \langle P | J_\mu(0) | X \rangle \langle X | J_\nu(y) | P \rangle e^{i(q+P-p_X) \cdot y} \\ &= 2\text{Im} \left[ \int d^4y \langle P | T \{ J_\mu(y) J_\nu(0) \} | P \rangle e^{iq \cdot y} \right] \end{aligned} \quad (3.18.3)$$

Something very nice has occurred – the square of the matrix elements turned into the correlator of  $J_\mu(y) J_\nu(0)$  in a 1-proton state with momentum  $P$ , after summing over  $X$ ! This was guaranteed by the optical theorem, since we are summing over all states  $X$  (we took the imaginary part in order to justify the time ordering symbol inside the correlator; we could also have left it without the time ordering symbol and without the imaginary part operation).

We would like to use this expression to study DIS at large  $Q^2 = -q^2$ . When  $q$  is very large, one might expect that we are dominated by  $y^2 \rightarrow 0$ . This does not quite mean that  $y_\mu \rightarrow 0$ , since near the lightcone we have  $y_\mu$  large but  $y^2 \approx 0$ . However, for most purposes it will be good enough to study  $y_\mu$  small. In particular, when we study integrals over the partonic  $x$  variable associated with the PDFs, our large  $Q$  expansion will make sense. Recall

$$x \equiv -\frac{q^2}{2P \cdot q} \quad (3.18.4)$$

where  $q$  is the off-shell photon momentum and  $P$  is the proton momentum, and the physical region is  $0 < x < 1$ . To integrate  $W_{\mu\nu}$  over this region, we need to integrate the imaginary part of the time-ordered correlator, which we can get from a discontinuity in the complex  $x$  plane across the real axis. But we can deform this contour of integration to pick up only poles at  $x = \infty$ . This justifies series expanding around  $Q^2 = 0$ .

This means we need to understand

$$J_\mu(y) J_\nu(0) \sim \sum C_n y^{\Delta_n - 2\Delta_J} \mathcal{O}_n^{\mu\nu}(0) \quad (3.18.5)$$

where we have organized an expansion in powers  $\Delta_n$  of  $y$ , so that we can focus on the terms that are most important at small  $y^2$ . Let us now discuss and justify such an *Operator Product Expansion*.

### 3.18.2 CFTs, Radial Quantization, and the OPE

Can we actually expand  $J_\mu(y)J_\nu(0)$  as written above?

This is a rather ambitious expansion to attempt. After all, expanding around  $y \rightarrow 0$  means that we are probing arbitrarily short-distance physics, which is something that we’ve almost always remained agnostic about. For this OPE to be well-defined, we need to understand our theory at very short distances. If, for example, the theory becomes strongly coupled at very short distances in a way that we do not understand, then this OPE will be very hard to compute.

In particular, if there is a non-trivial renormalization flow that continues as we go to shorter and shorter distances, then it’s hard to imagine the OPE can be defined or understood. Thus we can only attempt an OPE expansion in theories where the renormalization flow slows and stops at short distances. So the OPE only makes sense if the theory becomes *scale invariant*. It turns out that essentially all QFTs that are scale invariant have an even larger set of symmetries, called *conformal symmetries*. QFTs with such a symmetry pattern are called *Conformal Field Theories*.

[Aside: However, if the theory is nearly scale invariant over a wide range of distance scales, so that it’s well-described by a scale-invariant theory, we can just use the OPE of that theory and get nearly-correct results. So even in EFTs that eventually break down, the OPE can still be useful, and we needn’t abandon our EFT intuition that very very short-distance physics is ‘irrelevant’. Practically speaking, we will study QCD and use  $Q \gg \Lambda_{QCD}$ , but we needn’t take  $Q$  all the way  $\rightarrow \infty$ , where, e.g., grand unification and quantum gravity would become important.]

In fact, we do expect that QCD has this property of scale invariance at short distances, because it is asymptotically free. It turns into a very simple kind of CFT at short distances – it becomes a free theory. So the OPE does make sense in QCD, and we can compute it by starting with a free theory of quarks and adding perturbative corrections, which are small at very short distances.

The best and most general way to think about the OPE is via a radical re-interpretation of the path integral, known as *radial quantization*.

Let us study our theory in Euclidean space, so the metric is ‘all +’. If our theory makes sense down to arbitrarily short distances, then we can choose a special origin, and consider spheres about that origin. We *define* our notion of ‘time’ to be the (logarithm of) the radius of these spheres. This means that *states live on concentric spheres*.

What happens if we evolve in ‘time’ down to arbitrarily small distances? Then we left with a sphere of radius  $\epsilon$  where any given state  $\psi$  lives. But since the state lives on an arbitrarily small sphere, we can take the  $t \rightarrow -\infty$  limit and view it as living exactly at a point, the origin. Thus we identify states at the origin with operators  $\mathcal{O}_\psi$  where

$$\mathcal{O}_\psi(0)|0\rangle = |\psi\rangle \quad (3.18.6)$$

This is the *operator/state correspondence*, which is an isomorphism in CFTs.

Now consider two operators  $\mathcal{O}_1(x)\mathcal{O}_2(0)$ . We can surround them by a sphere, and thus study the state  $\psi_{\mathcal{O}_1, \mathcal{O}_2}(x, y)$  that results on that sphere. If we shrink that state down to zero size, we get another local operator, which will be a linear combination of any set of basis operators in the theory. Thus we have derived the *Operator Product Expansion* (OPE). The specific form of the OPE will be strongly constrained by symmetry.



### 3.18.3 Applying the OPE to DIS

Now we want to use the OPE to evaluate the Euclidean correlator

$$X_{\mu\nu} = i \int d^4y \langle P | J_\mu(y) J_\nu(0) | P \rangle e^{iq \cdot y} \quad (3.18.7)$$

We have removed the Lorentzian time ordering symbol because (unordered) Euclidean correlators automatically analytically continue to time-ordered Lorentzian correlators.

The OPE is an operator relation (so it's true when sandwiched inside any pair of states), so we want to evaluate

$$J_\mu(y) J_\nu(0) = \bar{q}(y) \gamma_\mu q(y) \bar{q}(0) \gamma_\nu q(0) \quad (3.18.8)$$

We get no non-trivial contribution to  $X_{\mu\nu}$  if we fully contract the fields, so the leading contribution comes from contracting a  $\bar{q}(y)q(0)$  or  $q(y)\bar{q}(0)$ , which just give free quark propagators to leading order. Note that the quarks have to be of identical types for us to make these contractions and get singular terms as  $y \rightarrow 0$ .

We need to take a Fourier transform, so it's most convenient to write the result as

$$\int d^4y e^{iq \cdot y} \bar{\psi}(y) \gamma_\mu \psi(y) \bar{\psi}(0) \gamma_\nu \psi(0) = \bar{\psi}(q) \gamma_\mu \frac{i(i\partial + \not{q})}{(i\partial + q)^2} \gamma_\nu \psi(0) \quad (3.18.9)$$

where I am writing the quark field as  $\psi$  to differentiate it from the momentum.

We can write the denominator as

$$\frac{1}{(i\partial + q)^2} = \frac{-1}{Q^2 - 2iq \cdot \partial + \partial^2} = -\frac{1}{Q^2} \sum_{n=0}^{\infty} \left( \frac{2iq \cdot \partial - \partial^2}{Q^2} \right)^n \quad (3.18.10)$$

Naively one would expect that all terms aside from  $n = 0$  should be suppressed at large  $Q$ , but in fact we will see that although we can ignore  $\partial^2 \sim P^2 \sim m_p^2$ , we must keep

$$\frac{2iq \cdot \partial}{Q^2} \sim \frac{2xq \cdot P}{Q^2} \sim 1 \quad (3.18.11)$$

where we have inserted a parton model 'x' type variable. Roughly speaking, we cannot neglect these terms because although we are taking  $Q$  large, we are not just probing short distances, but short distances *near the lightcone*. This is another way of saying that we are interested in protons that are highly boosted.

One can simplify the gamma matrices to give

$$\int d^4y e^{iq \cdot y} \bar{\psi}(y) \gamma_\mu \psi(y) \bar{\psi}(0) \gamma_\nu \psi(0) = -i \bar{\psi} (2\gamma^\mu (i\partial^\nu) - g^{\mu\nu} \not{q}) \frac{1}{Q^2} \sum_{n=0}^{\infty} \left( \frac{2iq \cdot \partial}{Q^2} \right)^n \psi \quad (3.18.12)$$

Thus we see that the operators that dominate the OPE for DIS are

$$\bar{\psi} \gamma^{\mu_1} (i\partial^{\mu_2}) (i\partial^{\mu_3}) \cdots (i\partial^{\mu_\ell}) \psi \quad (3.18.13)$$

where the indices  $\mu_i$  are completely symmetric. This should be made gauge invariant, giving

$$\mathcal{O}_\ell^{\mu_i} = \bar{\psi} \gamma^{\mu_1} (iD^{\mu_2}) (iD^{\mu_3}) \cdots (iD^{\mu_\ell}) \psi \quad (3.18.14)$$

Using these operators, we can write a general expression for

$$\int d^4y e^{iq \cdot y} J^\mu(x) J^\nu(0) = \sum_{n=2}^{\infty} 4 \frac{(2q_{\mu_1}) \cdots (2q_{\mu_{\ell-2}})}{Q^{2\ell-2}} \mathcal{O}_\ell^{\mu\nu\mu_i} - g^{\mu\nu} \frac{(2q_{\mu_1}) \cdots (2q_{\mu_{\ell-2}})}{Q^{2\ell-2}} \mathcal{O}_\ell^{\mu_i} \quad (3.18.15)$$

Let us use dimensional analysis to observe the relative importance of various operators. Due to the extra indices of spin  $\ell$  operators, the overall contribution must be

$$\left( \frac{P \cdot q}{Q^2} \right)^\ell \frac{1}{Q^{\Delta-\ell-2}} \propto x^\ell Q^{2+\ell-\Delta} \quad (3.18.16)$$

This means that at large  $Q$  with fixed  $x$ , the process is dominated by the operators of smallest *twist*  $\tau \equiv \Delta - \ell$ , where  $\Delta$  is the scaling dimension of the operator.

One can now analyze DIS by looking directly at the renormalization of the operators  $\mathcal{O}_\ell$  and keeping large logarithms. You can read much more in the textbook or in Peskin and Schroder (section 18.5). The punchline is that these operators control *moments* of the PDFs, ie quantities like

$$\int_0^1 dx x^{\ell-1} f_i(x; Q) \quad (3.18.17)$$

and the renormalization of  $\mathcal{O}_\ell$  tells us about the  $Q$  dependence. Thus *the DGLAP equations and the operator product expansion analysis of deep inelastic scattering encode exactly the same physics.*

There's a universal term from the stress energy tensor  $T^{\mu\nu}$  of QCD, which gives the  $\ell = 2$  contribution and dominates at very very large energies. Its presence tells us

$$\int_0^1 dx x f_i(x; Q) = \frac{1}{\frac{16}{3} + N_f} \quad (3.18.18)$$

where  $N_f$  is the number of flavors of quarks. This tells us how much momentum is carried by quarks and gluons in the very very large energy limit. One can compare all of this directly to the DGLAP equations by taking moments of the PDFs derived from them.

### 3.19 Anomalies – A Summary

This is the most important subject that we haven't yet covered. The essential point is that in some situations, symmetries of the classical action can be broken by quantum effects. The word 'anomaly' is used in many different contexts to indicate a few different effects:

- There is the *chiral anomaly*, whereby chiral symmetries – those that rotate left-handed and right-handed spinors  $\psi_L$  and  $\psi_R$  spinors differently (oppositely) – are violated by quantum effects. These anomalies have important consequences for gauge theories and gravity.

- Sometimes the quantum violation of scale invariance is referred to as an anomaly. For example, QED was classically scale invariant, so the renormalization flow of  $\alpha(\mu)$  can be viewed as an anomaly. This viewpoint is not usually well-developed, although it's valid.
- And as a more advanced comment... certain contact terms in correlators are referred to as anomaly terms. The most famous examples come from conformal field theories. Usually conformal invariance is associated with the statement that the trace  $T^\mu_\mu = 0$  in conformally invariant theories (one can show this by deriving the currents associated with scale invariance and special conformal transformations). However, if we study a CFT in a spacetime that is not flat, then  $T^\mu_\mu$  can get contributions proportional to the spacetime curvature. This corresponds to the statement that correlators of  $T^\mu_\mu$  with one or more other stress tensors may not vanish. The coefficients are called ‘conformal anomaly coefficients’.

### **An outline of some important theoretical points:**

- Chiral anomalies have both an UV and IR aspects, and are related to closely related to topology. They connect topological solutions in gauge theories, called instantons, to (integer) fermion numbers. The simplest example of an anomaly occurs in 1 + 1 dimensional QED (this is discussed in Peskin's textbook and also in TASI lectures by Harvey).
- The UV aspect can be seen because the anomaly shows up at short distances, in contact terms, and in a kind of residue from the regulator of loop diagrams. Chiral anomalies also involve a tension between Lorentz invariance and symmetry – we could eliminate the anomaly by breaking Lorentz invariance.
- The IR aspect is that chiral anomalies can only be produced by massless particles, and in fact they are one-loop exact. This can be seen by integrating the anomalous current conservation equation

$$\partial^\mu J^\mu_5 = \frac{g^2}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} \quad (3.19.1)$$

On the LHS we find  $N_R - N_L$ , the difference in the number of left and right chiral fermions. On the RHS we find a topological invariant of the gauge field. Both sides must be integers, so there cannot be any further perturbative corrections to the anomaly.

- From the path integral perspective, we saw that classical symmetries will become quantum symmetries if they leave the action and *the measure of the path integral* invariant. It is the non-invariance of the path integral measure that leads to anomalies.
- Anomalies impose 't Hooft *anomaly matching conditions* between the UV and IR descriptions of theories. These give information on IR theories of goldstone bosons, powerful constraints on the possible compositeness of quarks and leptons, and in the case of conformal theories, they constrain renormalization flows via the *c*-theorem and *a*-theorem of 2d and 4d QFT.

### **A few important phenomenological implications:**

- There is no  $\eta'$  particle. This is a putative goldstone boson of the  $U(1)$  chiral global symmetry of  $SU(2) \times SU(2)$ , but it is absent because that symmetry is anomalous (and thus isn't a symmetry).
- The decay  $\pi^0 \rightarrow \gamma\gamma$  is possible and predicted by the anomaly, although classically this process is impossible. The vertex is

$$\frac{N_c e^2}{48\pi^2} \frac{\pi^0(x)}{F_\pi} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}(x) F_{\rho\sigma}(x) \quad (3.19.2)$$

derived from studying the effect of anomalies on goldstone bosons. Note that this affords a direct measurement of  $N_c$ ! Historically, it was one of the original ways that physicists determined  $N_c = 3$  in the standard model.

- The standard model does not have any anomalies in its gauge symmetries; this is necessary for the self-consistency of these theories. The absence of anomalies puts non-trivial constraints on beyond the standard model fermions (such as supersymmetric models that postulate new fermion partners).
- Anomaly matching conditions strongly constrains goldstone bosons in the IR theory, and can be used to prove that the  $SU(3) \times SU(3)$  global symmetry must be spontaneously broken in the standard model. The idea here is that the low-energy theory of goldstone bosons must include anomaly terms that match the UV theory (of fermions).

### 3.20 Anomalies as Almost Local Effects

The most important feature of anomalies is that they represent effects that are almost but not quite local, as we will now describe.

One way to talk about anomalies is by introducing sources for any and all operators that one wants to study. The main use of this formalism is that it makes it easy to distinguish between local and non-local effects. If we want to study the correlation functions of conserved currents, then we should introduce background vector fields  $A_\mu$  corresponding to each current  $J_\mu$ , and then study the functional

$$Z[A_\mu(x)] = \int \mathcal{D}\phi e^{iS[\phi] + \int d^d x J_\mu(x) A^\mu(x)} \quad (3.20.1)$$

This object will be extremely complicated and non-local. By the words ‘non-local’ I just mean that when we vary with respect to  $A_\mu(x)$ , ie via

$$\frac{\delta}{\delta A_\mu(x)} \frac{\delta}{\delta A_\nu(y)} Z \Big|_{A=0} = \langle J_\mu(x) J_\nu(y) \rangle \quad (3.20.2)$$

and this is non-zero when  $x \neq y$ . We will not be so ambitious as to try to explicitly discuss or classify the non-local structures that can appear in  $Z[A]$ .

However, we might also expect that  $Z[A]$  will have purely local terms. These would only contribute to correlators like  $\langle J_\mu(x)J_\nu(y) \rangle$  when  $x = y$ , which means that they produce pure delta functions in these correlators. For example we might have a term of the form

$$Z[A] \supset e^{\int d^d x A_\mu(x) A^\mu(x)} \quad (3.20.3)$$

which would produce a correlator

$$\langle J_\mu(x)J_\nu(y) \rangle \supset g_{\mu\nu} \delta^4(x - y) \quad (3.20.4)$$

for the physical currents  $J_\mu$  (which are the genuine operators in our theory; remember that we are viewing the  $A_\mu$  as background sources).

Naively, from the point of view of low-energy physics and/or effective field theory, we should expect that these local terms are not very interesting. This follows from their definition – since they are local or ‘contact’ terms, they are only important when operators are at exactly the same point, which seems to mean that they must be UV sensitive. In particular, we would expect these terms to depend very explicitly on what regulator we use to define the very short-distance physics of our theory. This suggests that they are not at all universal. Anomaly terms are the exceptions to this intuition, which is one reason why they are interesting.

Anomalies represent a violation of the symmetry, so before discussing them we need to understand how the symmetry appears in  $Z[A_\mu(x)]$ . This is relatively simple. The symmetry says that the current is conserved, so

$$\partial^\mu J_\mu = 0 \quad (3.20.5)$$

This means that  $Z[A_\mu(x)]$  should be invariant under transformations that look like gauge transformations, insofar as

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \alpha(x) \quad (3.20.6)$$

just takes

$$\begin{aligned} \delta \int J^\mu A_\mu &= \int d^d x J^\mu(x) \partial_\mu \alpha(x) \\ &= - \int d^d x \partial_\mu J^\mu(x) \alpha(x) = 0 \end{aligned} \quad (3.20.7)$$

as long as  $\alpha(x)$  vanishes at infinity (note that we use more complicated non-linear transformations for some non-abelian symmetries). Thus an anomaly must show up as a violation of the gauge invariance of  $Z[A_\mu]$ . Ordinary violations of the symmetry can show up as non-local terms in  $Z$ .

Anomalies are special terms that are *almost* local. If they really were entirely local, then they would just be cutoff-dependent objects that could be removed by a shift in the regulator (ie by adding purely local counter-terms to  $Z[A]$ ). If they were entirely non-local, they would involve separate points and the symmetry would be completely and obviously broken. How can something be ‘almost’ local? The answer is that anomaly terms are non-local, but their variation under

$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \alpha(x)$  is in fact purely local. What could this mean? Well consider a function like

$$X = \int d^d x \frac{\partial^\mu A_\mu}{\partial^2} G[A; x] \quad (3.20.8)$$

where  $G$  is some general function of sources or fields (it depends on  $A$ , or perhaps on other similar sources for other currents and operators). Then under  $A_\mu \rightarrow A_\mu + \partial_\mu \alpha$ , we see that this term becomes the purely local object

$$\delta X = \int d^d x \alpha(x) G[A; x] \quad (3.20.9)$$

Thus it has a local variation under the symmetry transformation, even though it's not local itself. It also has the implication that

$$\langle \partial^\mu J_\mu \rangle = \partial^\mu \frac{\delta X}{\delta A_\mu} = G[A; x] \quad (3.20.10)$$

so that  $G$  does appear as an ‘anomaly’, ie as a violation of the usual current conservation condition. As a more complete example, the simplest type of anomaly, that associated with three  $U(1)$  currents, can appear as

$$\int d^4 x \frac{\partial^\mu A_\mu}{\partial^2} \epsilon^{abcd} F_{ab} F_{cd} \quad (3.20.11)$$

Note that  $\epsilon^{abcd} F_{ab} F_{cd} = \partial_a (\epsilon^{abcd} A_b F_{cd})$  and so it is a total derivative. Thus derivatives in this expression can be moved around using integration by parts. It's also worth noting that our anomaly term is formally scale-invariant, ie it does not require a dimensionful coefficient.

This *almost local* property of anomalies is what makes them special, important, and powerful. Usually anomalies are discussed as quantum effects that violate classical symmetries. This is correct of course, but by itself it doesn't explain why they're subtle. The point is that fully non-local violations of a symmetry would be very obvious, and would already be visible at the classical level. Purely local symmetry violations (such as what we would obtain from choosing a regulator that breaks the symmetry) could just be removed by a redefinition of the regulator. Anomalies straddle these two unsubtle limits.

Since the anomalies are non-local, they can only be produced by massless particles. This follows because massive particle could just be integrated out of the theory, resulting in purely local terms suppressed by  $\frac{1}{M^2}$ . However, because anomalies are ‘almost local’, they can still be seen in calculations at very high energies; as we observed they are essentially scale-invariant. Thus we see the famous fact that anomalies are in some sense simultaneously UV and IR phenomena, which makes it possible to match anomalies between UV and IR descriptions of a single theory (e.g. between a theory of quarks and the chiral Lagrangian for goldstone bosons).

Of course it's useful to pierce all this abstraction with a concrete example. The simplest possible anomaly occurs in  $1 + 1$  dimensions, where the anomaly term for QED is simply

$$\int d^2 x \frac{\partial^\mu A_\mu}{\partial^2} \epsilon^{ab} F_{ab} \quad (3.20.12)$$

corresponding to the statement that the axial current

$$\langle \partial_\mu J_{axial}^\mu \rangle = \frac{1}{2\pi} \epsilon^{ab} F_{ab} \quad (3.20.13)$$

We can compute this directly by calculating a fermion loop.

For this purpose, note that we can take the 2 gamma matrices (so we are dealing with a 1 + 1 dimensional Dirac spinor) to be

$$\gamma^0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \bar{\gamma} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.20.14)$$

where  $\bar{\gamma} = \gamma^0 \gamma^1$  is the equivalent of  $\gamma^5$  in 3 + 1 dimensions. Then the vector and axial currents are

$$\begin{aligned} J_\mu^V &= \bar{\psi} \gamma_\mu \psi \\ J_\mu^A &= \bar{\psi} \gamma_\mu \bar{\gamma} \psi \end{aligned} \quad (3.20.15)$$

and we simply have

$$J_\mu^A = -\epsilon^{\mu\nu} J_\nu^V \quad (3.20.16)$$

because  $\gamma^\mu \bar{\gamma} = -\epsilon^{\mu\nu} \gamma_\nu$ . The fermion loop simply gives

$$i\Pi^{\mu\nu}(q) = i(q^2 \eta^{\mu\nu} - q^\mu q^\nu) \frac{1}{\pi q^2} \quad (3.20.17)$$

The vector current is conserved, but the axial vector current has

$$q_\mu \langle J_A^\mu(q) \rangle = \frac{1}{\pi} \epsilon^{\mu\nu} q_\mu A_\nu \quad (3.20.18)$$

in the presence of a non-zero background electromagnetic field  $A_\mu$ .

### 3.21 A Lecture on Cosmological Perturbation Theory

Why is the universe inhomogeneous? We think the answer is that it began being very, but not completely, homogeneous and those inhomogeneities collapsed. So where did those initial inhomogeneities come from? We think they arose from quantum fluctuations during an inflationary phase. To discuss this, we will cover four topics:

1. FRW Cosmology and Inflation (Classically)
2. Quantizing a scalar field in deSitter space and the 2-pt correlator.
3. Gauge fixing issues and perturbations from a shift of the inflationary ‘clock’
4. Linear growth of density perturbations.

### 3.21.1 Classical FRW Cosmology and Inflation

Einstein's equations have solutions corresponding to homogeneous matter and energy called FRW solutions, with spacetime metric:

$$ds^2 = dt^2 - a^2(t)dx_i^2 \quad (3.21.1)$$

where  $a(t)$  is called the scale factor, and the Hubble parameter (not always a constant) is  $H = \frac{\dot{a}}{a}$ . Here we have already assumed that fixed time surfaces are spatially flat. There are two other possibilities – positive curved (spheres) or negatively curved (hyperboloids) but I'll ignore that for reasons that will eventually be explained.

The equations of motion for  $a(t)$  depend on the matter and energy in the universe. This isn't a course on cosmology or GR, so I'll be brief with all that. If we write the stress energy tensor of matter  $T_{\mu\nu}$  in the perfect fluid form then

$$\begin{aligned} \dot{\rho} &= -3\frac{\dot{a}}{a}(\rho + p) \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p) \end{aligned} \quad (3.21.2)$$

The first equation tells us how the energy density evolves with time, while the second tells us how the scalar factor responds to the energy density. Note that  $3\frac{\dot{a}}{a}$  is the relative rate of change of the spatial volume, ie it is  $\frac{\dot{V}}{V}$ .

It's conventional to define an equation of state for the perfect fluid as

$$p = w\rho \quad (3.21.3)$$

where we take  $w$  to be a constant. Free particles, aka 'dust' or non-interacting matter corresponds to  $w = 0$ , while radiation (relativistic particles) have  $w = \frac{1}{3}$ . This simplifies the equations further to

$$\begin{aligned} \frac{\dot{\rho}}{\rho} &= -3\frac{\dot{a}}{a}(1 + w) \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}\rho(1 + 3w) \end{aligned} \quad (3.21.4)$$

Note that there are two very special values of  $w$ , where the behavior of the equations changes qualitatively –  $w = -1$  and  $w = -\frac{1}{3}$ . The former corresponds to a cosmological constant (the latter indicates the border between 'accelerating' and 'decelerating' cosmological expansion). Note that a cosmological constant does what its name implies –  $\dot{\rho} = 0$  for  $w = -1$ . It is believed that all forms of matter satisfy  $w \geq -1$ , so the cosmological constant is at the far end of a spectrum.

DeSitter space corresponds to the solution  $w = -1$ . In this case, since  $\rho$  is constant, we simply have

$$\ddot{a} = \frac{8\pi G\rho}{3}a \quad (3.21.5)$$



which has exponentially growing and shrinking solutions

$$a(t) = e^{Ht} \quad (3.21.6)$$

where the (constant)  $H^2 = \frac{8\pi G}{3}\rho$ . This is the paradigmatic example of an *inflating* spacetime. Since our universe has a cosmological constant, it will eventually (in many billions of years) look like this.

Next let's consider a classical scalar field, which provides a very specific kind of matter. Let's take

$$L = \frac{1}{2}(\partial\phi)^2 - V(\phi) \quad (3.21.7)$$

as usual. If we include gravity, then we obtain FRW equations involving  $a(t)$  and  $\phi$ . Let's assume that  $\phi$  is rolling down its potential homogeneously, so that  $\phi(t)$ . Then its equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} = -V'(\phi) \quad (3.21.8)$$

along with

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ p &= \frac{1}{2}\dot{\phi}^2 - V(\phi) \end{aligned} \quad (3.21.9)$$

Let's think about these equations. The last two say that  $\dot{\phi}^2 \ll V(\phi)$  then the scalar field basically acts like a cosmological constant. The first equation tells us that if  $\ddot{\phi} \ll H\dot{\phi}$  then the scalar field's motion is dominated by friction – specifically, what's called ‘Hubble friction’. In this friction dominated, slow-roll limit we can write

$$\dot{\phi} = -\frac{V'}{3H} = \frac{V'}{\sqrt{24\pi GV}} \quad (3.21.10)$$

Slow roll inflation occurs when  $\dot{\phi}$  is small and the motion of  $\phi$  is dominated by Hubble friction. We can translate this into the statement that

$$\begin{aligned} \frac{\dot{V}}{HV} &\ll 1 \\ \frac{\ddot{\phi}}{H\dot{\phi}} &\ll 1 \end{aligned} \quad (3.21.11)$$

With some algebra we can translate these statements into conditions on the potential  $V$  and its derivatives, noting that  $\dot{V} = V'\dot{\phi}$ .

### 3.21.2 A Free Scalar Field in DeSitter Space

Now let's discuss the quantization of a free massless scalar field in DeSitter space. This may be an interesting problem in its own right, but as we'll explain in the next section, it's the crucial ingredient we need to understand how density perturbations are generated during (slow-roll) inflation.

For this purpose, it's easiest if we change coordinates and write the pure dS metric as

$$ds^2 = \frac{1}{H^2\eta^2} (d\eta^2 - dx_i^2) \quad (3.21.12)$$

It's easy to get here from  $dt^2 - e^{2Ht} dx_i^2$  by pulling out an overall factor of  $e^{2Ht}$ . Note that  $\eta \in (-\infty, 0)$ .

Now we just proceed with canonical quantization according to the usual rules. In this metric, the action for our field is

$$\int \frac{d\eta d^3x_i}{H^2\eta^2} \left( (\partial_\eta \phi)^2 - (\vec{\nabla} \phi)^2 \right) \quad (3.21.13)$$

The equations of motion are very simple to derive. We can go to (spatial) momentum space, in which case the solutions are

$$f(\eta, k) = \frac{H}{\sqrt{2k^3}} (1 - ik\eta) e^{ik\eta} \quad (3.21.14)$$

and their complex conjugates. Here we have normalized the solutions so that the quantum operators

$$\phi(\eta, x) = \int d^3k \left( f(\eta, k) a_k + f^\dagger(\eta, k) a_k^\dagger \right) \quad (3.21.15)$$

have canonical commutation relations with the canonical momentum fields

$$\pi(\eta, x) = \frac{\partial_\eta \phi}{H^2\eta^2} \quad (3.21.16)$$

To check this, note that

$$\begin{aligned} [\pi(\eta, x), \phi(\eta, y)] &= \frac{1}{H^2\eta^2} \int d^3k \frac{H^2}{2k^3} e^{ik(x-y)} ((1 + ik\eta)k^2\eta - (1 - ik\eta)k^2\eta) \\ &= i\delta^3(\vec{x} - \vec{y}) \end{aligned} \quad (3.21.17)$$

and so we have canonically quantized correctly.

Now we can compute the physical 2-pt correlator of the fields  $\phi$  at the same time. This is

$$\begin{aligned} \langle \phi(\eta, k) \phi(\eta, k') \rangle &= (2\pi)^3 \delta^3(\vec{k} + \vec{k}') |f_k(\eta)|^2 \\ &= (2\pi)^3 \delta^3(\vec{k} + \vec{k}') \frac{H^2}{2k^3} (1 + k^2\eta^2) \end{aligned} \quad (3.21.18)$$

This is the crucial result. Note that when  $k\eta \ll 1$  this has an even simpler form. The fact that it is proportional to  $k^{-3}$  means that naively, in position space there isn't any dependence on  $\vec{x} - \vec{y}$  at

all! This is the *scale invariance* that is often referenced with regards to density perturbations and fluctuations in deSitter spacetime.

Note that in fact the Fourier transform of the 2-pt correlator contains a logarithm, so it isn't purely constant. Thus

$$\langle \phi(\eta, x) \phi(\eta, y) \rangle \propto H^2 \log |\vec{x} - \vec{y}| \quad (3.21.19)$$

This is in comoving coordinates. But in physical coordinates and the more physical time  $t$ ,  $|\vec{x} - \vec{y}| \propto e^{Ht}$  for  $x$  and  $y$  on geodesics. This means that the 2-pt function between two points of separate timelike geodesics behaves as

$$\langle \phi(t) \phi(t) \rangle \sim H^3 t \quad (3.21.20)$$

which means that the 2-pt function grows linearly in physical time! This is exactly the behavior expected for a random walk. This is no accident – the correlators in deSitter space are thermal, with Hawking temperature  $T = \frac{H}{2\pi}$ , and we are seeing a random walk due to this quantum temperature.

### 3.21.3 Time at the End of Inflation as the Origin of Inhomogeneities

Now we have the task of relating the very simple canonical quantization result from the last section to the actual, physical spectrum of density perturbations from slow-roll inflation. To do this precisely and correctly is actually more difficult and subtle than the simple computation in the previous section (see astro-ph/0210603 for the full analysis). Here we will just rely on a physical argument and explanation.

We are studying a slow-roll model, which means that

$$\begin{aligned} \frac{\dot{V}}{HV} &\ll 1 \\ \frac{\ddot{\phi}}{H\dot{\phi}} &\ll 1 \end{aligned} \quad (3.21.21)$$

This can be translated into a statement about the standard *slow roll parameters* defined as

$$\begin{aligned} \epsilon &\equiv \frac{1}{2G_N} \left( \frac{V'}{V} \right)^2 \sim \frac{G_N \dot{\phi}^2}{2H^2} \ll 1 \\ \eta &\equiv \frac{V''}{G_N V} \sim -\frac{\ddot{\phi}}{H\dot{\phi}} + \frac{G_N \dot{\phi}^2}{2H^2} \ll 1 \end{aligned} \quad (3.21.22)$$

This also means that

$$\frac{1}{2} \dot{\phi}^2 \ll V(\phi) \quad (3.21.23)$$

in the slow roll regime, so the energy density of the universe is dominated by the potential  $V$  rather than by the kinetic energy of the scalar. This means that  $p \approx -\rho$  or  $w \approx -1$ , and so the spacetime metric will be approximately equal to that of deSitter space.

Thus we can view  $\phi(t, x) = \phi_{cl}(t) + \delta\phi(t, x)$  where  $\delta\phi$  is a quantum fluctuation about the time-dependent classical solution. Simply by virtue of the fact that  $\phi_{cl}(t)$  is a solution to the classical equations of motion, if we expand  $\phi$  in this form then  $\delta\phi$  will not contribute at zeroth or first order, and so will just look like a massless scalar field in deSitter. We quantized such a field in the last section, so we know the behavior of the quantum fluctuations  $\delta\phi$ . Very roughly speaking, we expect that  $\delta\phi \sim \frac{H}{2\pi}$ , and that correlations of  $\delta\phi$  are scale invariant.

The question we now face is – how do we translate the behavior of  $\delta\phi$  into a prediction for the spectrum of *density perturbations*, the primordial inhomogeneities?

Naively, you might expect that this is easy. Fluctuations  $\delta\phi$  change the value of  $V(\phi) = V(\phi_{cl} + \delta\phi)$ , and this alters the energy density in the universe. This would lead to a prediction

$$\delta\rho \stackrel{?}{\approx} V'\delta\phi \sim \frac{V'H}{2\pi} \sim \sqrt{\epsilon} \frac{H}{M_{pl}} V \quad (3.21.24)$$

Thus we see that if this is the whole story, then  $\frac{\delta\rho}{\rho} \sim \sqrt{\epsilon} \frac{H}{M_{pl}}$ , so the density perturbations would be double suppressed.

However, this answer is completely wrong. The reason is that the primordial inhomogeneities are not dominated by shifts in the local energy density. Rather, they are dominated by the effect of shifts in the total time that the universe inflates (as a function of local position), which leads to more or less expansion (and thus dilution) in different regions. Thus we have that

$$\delta\rho \approx \rho(e^{3H\delta t} - 1) \approx \rho 3H \frac{\delta\phi}{\dot{\phi}} \approx \frac{3H^2}{2\pi\dot{\phi}} \rho \approx \frac{H}{\sqrt{\epsilon}M_{pl}} \rho \quad (3.21.25)$$

and this last result is *enhanced* by  $\frac{1}{\sqrt{\epsilon}}$ . This is the correct result.

Why did we have to guess the answer this way? The reason is that to do the calculation correctly, one must couple  $\phi$  to the full gravitational theory in perturbation theory, fix the gauge (general relativity has a huge gauge freedom), eliminate some auxiliary fields (like solving for one non-propagating component of  $A_\mu$  after gauge fixing), then finally find an action for the remaining scalar fields in a physically sensible gauge. This is technically complicated and potentially confusing (see astro-ph/0210603 if you're interested in how it's really done).

### 3.21.4 Growth of Linear Perturbations

## References

- [1] J. Polchinski, “Effective field theory and the Fermi surface,” [arXiv:hep-th/9210046](#) [hep-th].
- [2] J. L. Cardy, “Scaling and renormalization in statistical physics,”.
- [3] S. Weinberg, “The Quantum theory of fields. Vol. 1: Foundations,”.