# Physics 215B: Particles and Fields
# Winter 2018

Lecturer: McGreevy

These lecture notes live here. Please email corrections to mcgreevy at physics dot ucsd dot edu.

Last updated: 2018/03/23, 12:32:33

# Contents

## 0.1 Sources

The material in these notes is collected from many places, among which I should mention in particular the following:

Peskin and Schroeder, *An introduction to quantum field theory* (Wiley)

Zee, *Quantum Field Theory* (Princeton, 2d Edition)

Banks, *Modern Quantum Field Theory: A Concise Introduction* (Cambridge)

Schwartz, *Quantum field theory and the standard model* (Cambridge)

David Tong's lecture notes

Many other bits of wisdom come from the Berkeley QFT courses of Prof. L. Hall and Prof. M. Halpern.

## 0.2 Conventions

Following most QFT books, I am going to use the $+ - --$ signature convention for the Minkowski metric. I am used to the other convention, where time is the weird one, so I'll need your help checking my signs. More explicitly, denoting a small spacetime displacement as $dx^\mu \equiv (dt, d\vec{x})^\mu$, the Lorentz-invariant distance is:

$$ds^2 = +dt^2 - d\vec{x} \cdot d\vec{x} = \eta_{\mu\nu} dx^\mu dx^\nu \quad \text{with} \quad \eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}_{\mu\nu} .$$

(spacelike is negative). We will also write $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\partial_t, \vec{\nabla}_x\right)^\mu$, and $\partial^\mu \equiv \eta^{\mu\nu}\partial_\nu$. I'll use $\mu, \nu...$ for Lorentz indices, and $i, k, ...$ for spatial indices.

The convention that repeated indices are summed is always in effect unless otherwise indicated.

A consequence of the fact that english and math are written from left to right is that time goes to the left.

A useful generalization of the shorthand $\hbar \equiv \frac{h}{2\pi}$ is

$$đk \equiv \frac{\mathrm{d}k}{2\pi}.$$

I will also write $\not{\delta}^d(q) \equiv (2\pi)^d \delta^{(d)}(q)$. I will try to be consistent about writing Fourier transforms as

$$\int \frac{\mathrm{d}^d k}{(2\pi)^d} e^{ikx} \tilde{f}(k) \equiv \int đ^d k \; e^{ikx} \tilde{f}(k) \equiv f(x).$$

IFF $\equiv$ if and only if.
RHS $\equiv$ right-hand side. LHS $\equiv$ left-hand side. BHS $\equiv$ both-hand side.
IBP $\equiv$ integration by parts. WLOG $\equiv$ without loss of generality.
$+\mathcal{O}(x^n) \equiv$ plus terms which go like $x^n$ (and higher powers) when $x$ is small.
$+h.c. \equiv$ plus hermitian conjugate.
$\mathcal{L} \ni \mathcal{O}$ means the object $\mathcal{L}$ contains the term $\mathcal{O}$.

We work in units where $\hbar$ and the speed of light, $c$, are equal to one unless otherwise noted. When I say 'Peskin' I usually mean 'Peskin & Schroeder'.

Please tell me if you find typos or errors or violations of the rules above.

# 7 To infinity and beyond

Last quarter we ended at a high point, computing the amplitudes and cross-sections for many processes using QED. More precisely, we studied the leading-order-in-$\alpha$ amplitudes, using Feynman diagrams which were trees – no loops. The natural next step is to look at the next terms in the perturbation expansion in $\alpha$, which come from diagrams with one loop. When we do that we're going to encounter some confusing stuff, in fact some of the same confusing stuff we encountered in thinking about Casimir forces at the beginning of last quarter.

We didn't encounter these short-distance issues in studying tree-level diagrams because in a tree-level diagram, the quantum numbers (and in particular the momenta) of the intermediate states are fixed by the external states. In contrast, once there is a loop, there are undetermined momenta which must be summed, and this sum includes, it seems, arbitrarily high momentum modes, about which surely we have no information yet.

In order to put ourselves in the right frame of mind to think about that stuff, let's make a brief retreat to systems with finitely many degrees of freedom.

Then we'll apply some of these lessons to a toy field theory example (scalar field theory). Then we'll come back to perturbation theory in QED. Reading assignment for this chapter: Zee §III.

## 7.1 A parable from quantum mechanics on the breaking of scale invariance

Recall that the coupling constant in $\phi^4$ theory in $D = 3+1$ spacetime dimensions is dimensionless, and the same is true of the electromagnetic coupling $e$ in QED in $D = 3+1$ spacetime dimensions. In fact, the mass parameters are the only dimensionful quantities in those theories, at least in their classical avatars. This means that if we ignore the masses, for example because we are interested in physics at much higher energies, then these models seem to possess *scale invariance*: the physics is unchanged under zooming in.

Here we will study a simple quantum mechanical example (that is: an example with a finite number of degrees of freedom)[1] with such (classical) scale invariance. It exhibits many interesting features that can happen in strongly interacting quantum field theory – asymptotic freedom, dimensional transmutation. Because the model is simple, we can understand these phenomena without resort to perturbation theory.

[1] I learned this example from Marty Halpern.

They will nevertheless illuminate some ways of thinking which we'll need in examples where perturbating is our only option.

Consider the following ('bare') action:

$$S[q] = \int dt \left( \frac{1}{2}\dot{\vec{q}}^2 + g_0 \delta^{(2)}(\vec{q}) \right) \equiv \int dt \left( \frac{1}{2}\dot{\vec{q}}^2 - V(\vec{q}) \right)$$

where $\vec{q} = (x, y)$ are two coordinates of a quantum particle, and the potential involves $\delta^{(2)}(\vec{q}) \equiv \delta(x)\delta(y)$, a Dirac delta function. I chose the sign so that $g_0 > 0$ is attractive. (Notice that I have absorbed the inertial mass $m$ in $\frac{1}{2}mv^2$ into a redefinition of the variable $q$, $q \to \sqrt{m}q$.)

---

First, let's do dimensional analysis (always a good idea). Since $\hbar = c = 1$, all dimensionful quantites are some power of a length. Let $-[X]$ denote the number of powers of length in the units of the quantity $X$; that is, if $X \sim (\text{length})^{\nu(X)}$ then we have $[X] = -\nu(X)$, a number. We have:

$$[t] = [\text{length}/c] = -1 \implies [dt] = -1.$$

The action appears in the exponent in the path integrand, and is therefore dimensionless (it has units of $\hbar$), so we had better have:

$$0 = [S] = [\hbar]$$

and this applies to each term in the action. We begin with the kinetic term:

$$0 = [\int dt \dot{\vec{q}}^2] \implies$$

$$[\dot{\vec{q}}^2] = +1 \implies [\dot{\vec{q}}] = +\frac{1}{2} \implies [\vec{q}] = -\frac{1}{2}.$$

Since $1 = \int dq \delta(q)$, we have $0 = [dq] + [\delta(q)]$ and

$$[\delta^D(\vec{q})] = -[q]D = \frac{D}{2}, \quad \text{and in particular } [\delta^2(\vec{q})] = 1.$$

This implies that the naive ("engineering") dimensions of the coupling constant $g_0$ are $[g_0] = 0$ – it is dimensionless. Classically, the theory does not have a special length scale; it is scale invariant.

---

The Hamiltonian associated with the Lagrangian above is

$$H = \frac{1}{2}\left(p_x^2 + p_y^2\right) + V(\vec{q}).$$

Now we treat this as a quantum system. Acting in the position basis, the quantum Hamiltonian operator is

$$\mathbf{H} = -\frac{\hbar^2}{2}\left(\partial_x^2 + \partial_y^2\right) - g_0 \delta^{(2)}(\vec{q})$$

So in the Schrödinger equation $\mathbf{H}\psi = \left(-\frac{\hbar^2}{2}\nabla^2 + V(\vec{q})\right)\psi = E\psi$, the second term on the LHS is

$$V(\vec{q})\psi(\vec{q}) = -g_0 \delta^{(2)}(\vec{q})\psi(0).$$

To make it look more like we are doing QFT, let's solve it in momentum space:

$$\psi(\vec{q}) \equiv \int \frac{d^2p}{(2\pi\hbar)^2}\, e^{\mathbf{i}\vec{p}\cdot\vec{q}/\hbar}\varphi(\vec{p})$$

The delta function is

$$\delta^{(2)}(q) = \int \frac{d^2p}{(2\pi\hbar)^2}\, e^{\mathbf{i}\vec{p}\cdot\vec{q}/\hbar}.$$

So the Schrödinger equation says

$$\begin{aligned} \left(-\frac{1}{2}\nabla^2 - E\right)\psi(q) &= -V(q)\psi(q) \\ \int đ^2 p e^{ip\cdot q}\left(\frac{p^2}{2} - E\right)\varphi(p) &= +g_0\delta^2(q)\psi(0) \\ &= +g_0\left(\int đ^2 p e^{\mathbf{i}p\cdot q}\right)\psi(0) \end{aligned} \tag{7.1}$$

which (integrating the both-hand side of (7.1) over $q$: $\int d^2 q e^{-\mathbf{i}p\cdot q}\left((7.1)\right)$ ) says

$$\left(\frac{\vec{p}^2}{2} - E\right)\varphi(\vec{p}) = +g_0 \underbrace{\int \frac{d^2p'}{(2\pi\hbar)^2}\varphi(\vec{p}')}_{=\psi(0)}$$

There are two cases to consider:

- $\psi(\vec{q}=0) = \int đ^2 p\varphi(\vec{p}) = 0$. Then this case is the same as a free theory, with the constraint that $\psi(0) = 0$,

  $$\left(\frac{\vec{p}^2}{2} - E\right)\varphi(\vec{p}) = 0$$

  *i.e.* plane waves which vanish at the origin, *e.g.* $\psi \propto \sin\frac{p_x x}{\hbar}e^{\pm\mathbf{i}p_y y/\hbar}$. These scattering solutions don't see the delta-function potential at all.

- $\psi(0) \equiv \alpha \neq 0$, some constant to be determined. This means $\vec{p}^2/2 - E \neq 0$, so we can divide by it :

$$\varphi(\vec{p}) = \frac{g_0}{\frac{\vec{p}^2}{2} - E}\left(\int đ^2 p' \varphi(\vec{p}')\right) = \frac{g_0}{\frac{\vec{p}^2}{2} - E}\alpha.$$

The integral of the RHS (for $\psi(0) = \alpha$) is a little problematic if $E > 0$, since then there is some value of $p$ where $p^2 = 2E$. Avoid this singularity by going to the boundstate region: consider $E = -\epsilon_B < 0$. So:

$$\varphi(\vec{p}) = \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B}\alpha.$$

What happens if we integrate this $\int đ^2 p$ to check self-consistency – the LHS should give $\alpha$ again:

$$0 \stackrel{!}{=} \underbrace{\int đ^2 p \varphi(\vec{p})}_{=\psi(0)=\alpha\neq 0}\left(1 - \int đ^2 p \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B}\right)$$

$$\implies \qquad \boxed{\int đ^2 p \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B} = 1}$$

is a condition on the energy $\epsilon_B$ of possible boundstates.

But there's a problem: the integral on the LHS behaves at large $p$ like

$$\int \frac{d^2 p}{p^2} = \infty \ .$$

At this point in an undergrad QM class, you would give up on this model. In QFT we don't have that luxury, because this kind of thing happens all over the place. Here's what we do instead.

We cut off the integral at some large $p = \Lambda$:

$$\int^{\Lambda} \frac{d^2 p}{p^2} \sim \log \Lambda \ .$$

This our first example of the general principle that a classically scale invariant system will exhibit logarithmic divergences (rather: logarithmic dependence on the cutoff). It's the only kind allowed by dimensional analysis.

The introduction of the cutoff can be thought of in many ways: we could say there are no momentum states with $|p| > \Lambda$, or maybe we could say that the potential is not really a delta function if we look more closely. The choice of narrative here shouldn't affect our answers to physics questions.

More precisely:

$$\int^{\Lambda} \frac{d^2 p}{\frac{p^2}{2} + \epsilon_B} = 2\pi \int_0^{\Lambda} \frac{pdp}{\frac{p^2}{2} + \epsilon_B} = 2\pi \log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right) .$$

So in our cutoff theory, the boundstate condition is:

$$1 = g_0 \int^{\Lambda} \frac{đ^2 p}{\frac{p^2}{2} + \epsilon_B} = \frac{g_0}{2\pi\hbar^2} \log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right) .$$

A solution only exists for $g_0 > 0$. This makes sense since only then is the potential attractive (recall that $V = -g_0 \delta$).

Now here's a trivial-seeming step that offers a dramatic new vista: solve for $\epsilon_B$.

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} . \tag{7.2}$$

As we remove the cutoff ($\Lambda \to \infty$), we see that $E = -\epsilon_B \to -\infty$, the boundstate becomes more and more bound – the potential is too attractive.

Suppose we insist that the boundstate energy $\epsilon_B$ is a fixed thing – imagine we've *measured* it to be 200 MeV[2]. We should express everything in terms of the measured quantity. Then, given some cutoff $\Lambda$, we should solve for $g_0(\Lambda)$ to get the boundstate energy we have measured:

$$g_0(\Lambda) = \frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)} .$$

This is the crucial step: this silly symbol $g_0$ which appeared in our action doesn't mean anything to anyone (see Zee's dialogue with the S.E. in section III). *We are allowing $g_0 \equiv$ the bare coupling to be cutoff-dependent.*

Instead of a dimensionless coupling $g_0$, the useful theory contains an arbitrary *dimensionful* coupling constant (here $\epsilon_B$). This phenomenon is called *dimensional transmutation* (d.t.). The cutoff is supposed to go away in observables, which depend on $\epsilon_B$ instead.

In QCD we expect that in an identical way, an arbitrary scale $\Lambda_{QCD}$ will enter into physical quantities. (If QCD were the theory of the whole world, we would work in units where it was one.) This can be taken to be the rest mass of some mesons – boundstates of quarks. Unlike this example, in QCD there are many boundstates, but their energies are dimensionless multiplies of the one dimensionful scale, $\Lambda_{QCD}$. Nature chooses $\Lambda_{QCD} \simeq 200$ MeV.

[2]Spoiler alert: I picked this value of energy to stress the analogy with QCD.

[This d.t. phenomenon was maybe first seen in a perturbative field theory in S. Coleman, E. Weinberg, *Phys Rev* **D7** (1973) 1898. We'll come back to their example.]

There are more lessons in this example. Go back to (7.2):

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} \neq \sum_{n=0}^{\infty} g_0^n f_n(\Lambda)$$

it is *not* analytic (*i.e.* a power series) in $g_0(\Lambda)$ near small $g_0$; rather, there is an essential singularity in $g_0$. (All derivatives of $\epsilon_B$ with respect to $g_0$ vanish at $g_0 = 0$.) You can't expand the dimensionful parameter in powers of the coupling. This means that you'll never see it in perturbation theory in $g_0$. Dimensional transmutation is an inherently non-perturbative phenomenon.

Look at how the bare coupling depends on the cutoff in this example:

$$g_0(\Lambda) = \frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)} \overset{\Lambda^2 \gg \epsilon_B}{\to} \frac{2\pi\hbar^2}{\log\left(\frac{\Lambda^2}{2\epsilon_B}\right)} \overset{\Lambda^2 \gg \epsilon_B}{\to} 0$$

– the bare coupling vanishes in this limit, since we are insisting that the parameter $\epsilon_B$ is fixed. This is called *asymptotic freedom* (AF): the bare coupling goes to zero (*i.e.* the theory becomes free) as the cutoff is removed. This also happens in QCD.

**RG flow equations.** Define the *beta-function* as the logarithmic derivative of the bare coupling with respect to the cutoff:

$$\textbf{Def:} \qquad \beta(g_0) \equiv \Lambda \frac{\partial}{\partial \Lambda} g_0(\Lambda) \ .$$

For this theory

$$\beta(g_0) = \Lambda \frac{\partial}{\partial \Lambda} \left( \frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)} \right) \overset{\text{calculate}}{=} -\frac{g_0^2}{\pi\hbar^2} \left( \underbrace{1}_{\text{perturbative}} - \underbrace{e^{-2\pi\hbar^2/g_0}}_{\text{not perturbative}} \right) .$$

Notice that it's a function only of $g_0$, and not explicitly of $\Lambda$. Also, in this simple toy theory, the perturbation series for the beta function happens to stop at order $g_0^2$.

$\beta$ measures the failure of the cutoff to disappear from our discussion – it signals a quantum mechanical violation of scale invariance. What's $\beta$ for? Flow equations:

$$\dot{g}_0 = \beta(g_0).$$

[3] This is a tautology. The dot is

$$\dot{A} = \partial_s A, \quad s \equiv \log \Lambda/\Lambda_0 \implies \partial_s = \Lambda\partial_\Lambda.$$

($\Lambda_0$ is some reference scale.) But forget for the moment that this is just a definition:

$$\dot{g}_0 = -\frac{g_0^2}{\pi\hbar^2}\left(1 - e^{-2\pi\hbar^2/g_0}\right) .$$

This equation tells you how $g_0$ changes as you change the cutoff. Think of it as a nonlinear dynamical system (fixed points, limit cycles...)

**Def:** A fixed point $g_0^\star$ of a flow is a point where the flow stops:

$$\boxed{0 = \dot{g}_0|_{g_0^\star} = \beta(g_0^\star)} ,$$

a zero of the beta function. (Note: if we have many couplings $g_i$, then we have such an equation for each $g$: $\dot{g}_i = \beta_i(g)$. So $\beta_i$ is (locally) a vector field on the space of couplings.)

Where are the fixed points in our example?

$$\beta(g_0) = -\frac{g_0^2}{\pi\hbar^2}\left(1 - e^{-2\pi\hbar^2/g_0}\right) .$$

There's only one: $g_0^\star = 0$, near which $\beta(g_0) \sim -\frac{g_0^2}{\pi\hbar}$, the non-perturbative terms are small. What does the flow look like near this point? For $g_0 > 0$, $\dot{g}_0 = \beta(g_0) < 0$. With this (high-energy) definition of the direction of flow, $g_0 = 0$ is an attractive fixed point:

```
*<-<-<-<-<-<-<-<-<-<-<------------------------ g_0
```

$g_0^\star = 0$.

We already knew this. It just says $g_0(\Lambda) \sim \frac{1}{\log \Lambda^2} \to 0$ at large $\Lambda$. A lesson is that in the vicinity of such an AF fixed point, the non-perturbatuve stuff $e^{\frac{-2\pi\hbar^2}{g_0}}$ is small. So we can get good results near the fixed point from the perturbative part of $\beta$. That is: we can compute the behavior of the flow of couplings near an AF fixed point *perturbatively*, and be sure that it is an AF fixed point. This is the situation in QCD.

[3]Warning: The sign in this definition carries a great deal of cultural baggage. With the definition given here, the flow (increasing $s$) is toward the UV, toward high energy. This is the high-energy particle physics perspective, where we learn more physics by going to higher energies. As we will see, there is a strong argument to be made for the other perspective, that the flow should be regarded as going from UV to IR, since we lose information as we move in that direction – in fact, the IR behavior does not determine the UV behavior in general, but UV does determine IR.

On the other hand, the d.t. phenomenon that we've shown here is something that we can't prove in QCD. However, the circumstantial evidence is very strong!

---

Another example where this happens is quantum mechanics in any number of variables with a central potential $V = -\frac{g_0^2}{r^2}$. It is also classically scale invariant:

$$[r] = -\frac{1}{2}, \quad \left[\frac{1}{r^2}\right] = +1 \quad \Longrightarrow \quad [g_0] = 0.$$

This model was studied in K.M. Case, *Phys Rev* **80** (1950) 797 and you will study it on the first homework. The resulting boundstates and d.t. phenomenon are called Efimov states; this model preserves a *discrete* scale invariance.

Here's a quote from Marty Halpern from his lecture on this subject:

*I want you to study this set of examples very carefully, because it's the only time in your career when you will understand what is going on.*

In my experience it's been basically true. For real QFTs, you get distracted by Feynman diagrams, gauge invariance, regularization and renormalization schemes, and the fact that you can only do perturbation theory.

## 7.2 A simple example of perturbative renormalization in QFT

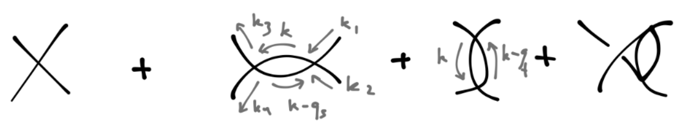
[Zee §III.1, Schwartz §15.4] Now let's consider an actual field theory but a simple one, namely the theory of a real scalar field in four dimensions, with

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - m^2\phi^2 - \frac{g}{4!}\phi^4. \tag{7.3}$$

Recall that $[\phi] = \frac{D-2}{2}$ so $[m] = 1$ and $[g] = \frac{4-D}{2}$, so $g$ is dimensionless in $D = 4$. As above, this will mean *logarithms*!

Let's do $2 \leftarrow 2$ scattering of $\phi$ particles.

$\mathbf{i}\mathcal{M}_{2\leftarrow 2} =$




$+ \quad \mathcal{O}(g^3)$

$$= -\mathbf{i}g \quad + \quad \mathbf{i}\mathcal{M}_s \quad + \quad \mathbf{i}\mathcal{M}_t \quad + \quad \mathbf{i}\mathcal{M}_u \quad + \quad \mathcal{O}(g^3)$$

where, in terms of $q_s \equiv k_1 + k_2$, the $s$-channel 1-loop amplitude is

$$\mathbf{i}\mathcal{M}_s = \frac{1}{2}(-\mathbf{i}g)^2 \int đ^4k \frac{\mathbf{i}}{k^2 - m^2 + \mathbf{i}\epsilon} \frac{\mathbf{i}}{(q_s - k)^2 - m^2 + \mathbf{i}\epsilon} \quad \sim \int^\Lambda \frac{d^4k}{k^4}.$$

**Parametrizing ignorance.** Recall our discovery of the scalar field at the beginning of last quarter by starting with a chain of springs, and looking at the long-wavelength (small-wavenumber) modes. In the sum, $\int d^4k$, the region of integration that's causing the trouble is *not* the part where the system looks most like a field theory. That is: if we look closely enough (small enough $1/k$), we will see that the mattress is made of springs. In terms of the microscopic description with springs, there is a smallest wavelength, of order the inverse lattice spacing: the sum stops.

Field theories arise from many such models, which may differ dramatically in their short-distance physics. We'd like to not worry too much about which one, but rather say things which do not depend on this choice. Recall the discussion of the Casimir force from §1: in that calculation, many different choices of regulators for the mode sum corresponded to different material properties of the conducting plates. The leading Casimir force was independent of this choice; more generally, it is an important part of the physics problem to identify which quantities are UV sensitive and which are not.

Parametrizing ignorance is another way to say 'doing science'. In the context of field theory, at least in the high-energy community it is called 'regularization'.

Now we need to talk about the integral a little more. The part which is causing the trouble is the bit with large $k$, which might as well be $|k| \sim \Lambda \gg m$, so let's set $m = 0$ for simplicity.

We'll spend lots of time learning to do integrals below. Here's the answer:

$$\mathbf{i}\mathcal{M} = -\mathbf{i}g + \mathbf{i}Cg^2 \left( \log \frac{\Lambda^2}{s} + \log \frac{\Lambda^2}{t} + \log \frac{\Lambda^2}{u} \right) + \mathcal{O}(g^3)$$

If you must know, $C = \frac{1}{16\pi^2}$.

**Observables can be predicted from other observables.** Again, the boldface statement might sound like some content-free tweet from some boring philosophy-of-science twitter feed, but actually it's a very important thing to remember here.

What is $g$? As Zee's Smart Experimentalist says, it is just a letter in some theorist's lagrangian, and it doesn't help anyone to write physical quantities in terms of it. Much more useful would be to say what is the scattering amplitude in terms of things that can be measured. So, suppose someone scatters $\phi$ particles at some given $(s, t, u) = (s_0, t_0, u_0)$, and finds for the amplitude $\mathbf{i}\mathcal{M}(s_0, t_0, u_0) = -\mathbf{i}g_P$ where $P$ is for 'physical'.[4] *This* we can relate to our theory letters:

$$-\mathbf{i}g_P = \mathbf{i}\mathcal{M}(s_0, t_0, u_0) = -\mathbf{i}g + \mathbf{i}Cg^2 L_0 + \mathcal{O}(g^3) \tag{7.4}$$

where $L_0 \equiv \log \frac{\Lambda^2}{s_0} + \log \frac{\Lambda^2}{t_0} + \log \frac{\Lambda^2}{u_0}$. (Note that quantities like $g_P$ are often called $g_R$ where 'R' is for 'renormalized,' whatever that is.)

[End of Lecture 21]

**Renormalization.** Now here comes the big gestalt shift: Solve this equation (7.4) for the stupid letter $g$

$$\begin{aligned} -\mathbf{i}g &= -\mathbf{i}g_P - \mathbf{i}Cg^2 L_0 + \mathcal{O}(g^3) \\ &= -\mathbf{i}g_P - \mathbf{i}Cg_P^2 L_0 + \mathcal{O}(g_P^3). \end{aligned} \tag{7.5}$$

and eliminate $g$ from the discussion:

$$\begin{aligned} \mathbf{i}\mathcal{M}(s,t,u) &= -\mathbf{i}g + \mathbf{i}Cg^2 L + \mathcal{O}(g^3) \\ &\overset{(7.5)}{=} -\mathbf{i}g_P - \mathbf{i}Cg_P^2 L_0 + \mathbf{i}Cg_P^2 L + \mathcal{O}(g_P^3) \end{aligned}$$

[4] You might hesitate here about my referring to the amplitude $\mathcal{M}$ as an 'observable'. The difficult and interesting question of what can actually be measured in experiments can be decoupled a bit from this discussion. I'll say more later, but if you are impatient see the beginning of Schwartz, chapter 18.

$$= -\mathbf{i}g_P + \mathbf{i}Cg_P^2\left(\log\frac{s_0}{s} + \log\frac{t_0}{t} + \log\frac{u_0}{u}\right) + \mathcal{O}(g_P^3). \tag{7.6}$$

This expresses the amplitude at any momenta (within the range of validity of the theory!) in terms of measured quantities, $g_P, s_0, t_0, u_0$. The cutoff $\Lambda$ is gone! Just like in our parable in §7.1, it was eliminated by letting the coupling vary with it, $g = g(\Lambda)$, according to (7.5). We'll say a lot more about how to think about that dependence.

**Renormalized perturbation theory.** To slick up this machinery, consider the following Lagrangian density (in fact the same as (7.3), with $m = 0$ for simplicity):

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 \tag{7.7}$$

but written in terms of the measured coupling $g_P$, and some as-yet-undetermined 'counterterm' $\delta_g$. Then

$$\mathcal{M}(s,t,u) = -g_P - \delta_g - Cg_P^2\left(\log\frac{s}{\Lambda^2} + \log\frac{t}{\Lambda^2} + \log\frac{u}{\Lambda^2}\right) + \mathcal{O}(g_P^3).$$

If, in order to enforce the *renormalization condition* $\mathcal{M}(s_0,t_0,u_0) = -g_P$, we choose

$$\delta_g = -g_P^2 C\left(\log\frac{s_0}{\Lambda^2} + \log\frac{t_0}{\Lambda^2} + \log\frac{u_0}{\Lambda^2}\right)$$

then we find

$$\mathcal{M}(s,t,u) = -g_P - Cg_P^2\left(\log\frac{s}{s_0} + \log\frac{t}{t_0} + \log\frac{u}{u_0}\right) + \mathcal{O}(g_P^3)$$

– all the dependence on the unknown cutoff is gone, we satisfy the observational demand $\mathcal{M}(s_0,t_0,u_0) = -g_P$, and we can predict the scattering amplitude (and others!) at any momenta.

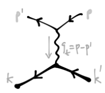
The only price is that the 'bare coupling' $g$ depends on the cutoff, and becomes infinite if we pretend that there is no cutoff. Happily, we didn't care about $g$ anyway. We can just let it go.

The step whereby we were able to absorb all the dependence on the cutoff into the bare coupling constant involved some apparent magic. It is not so clear that the same magic will happen if we study the next order $\mathcal{O}(g_P^3)$ terms, or if we study other amplitudes. A QFT where all the cutoff dependence to all orders can be removed with a finite number of counterterms is called 'renormalizable'. As we will see, such a field theory is less useful because it allows us to pretend that it is valid up to arbitrarily high energies. The alternative, where we must add more counterterms (such as something like $\frac{\delta_6}{\Lambda^2}\phi^6$) at each order in perturbation theory, is called an *effective field theory*, which is a field theory that has the decency to predict its regime of validity.

## 7.3 Radiative corrections to the Mott formula

Recall from last quarter that by studying scattering of an electron from a heavy charged fermion (a muon is convenient) we reconstructed the cross section for scattering off a Coulomb potential (named after Mott). Our next goal is to figure out how this cross section is corrected by other QED processes.

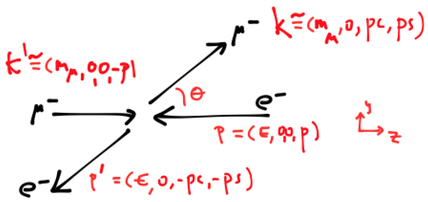
Recall that

$$\mathbf{i}\mathcal{M} = \quad$$  $$= (-\mathbf{i}e\bar{u}(p')\gamma^\mu u(p)))_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q^t_\mu q^t_\nu}{q_t^2}\right)}{q_t^2} (-\mathbf{i}e\bar{u}(k)\gamma^\nu u(k'))_{\text{muons}} \quad (7.8)$$

with $q_t \equiv p - p' = k - k'$. After the spin sum,

$$\frac{1}{4}\sum_{s,s',r,r'} |\mathcal{M}|^2 = 4\frac{e^4}{t^2}\left(-p_\mu p'_\nu - p'_\mu p_\nu - \eta_{\mu\nu}(-p\cdot p' + m_e^2)\right) \cdot \left(-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k\cdot k' + m_\mu^2)\right) \quad (7.9)$$

Consider the limit where the target $\mu$ particle is much heavier than the electron. 'Heavy' here means that we can approximate the CoM frame by its rest frame, and its initial and final energy as $k'_0 = m_\mu, k_0 = \sqrt{m_\mu^2 + \vec{k}^2} = m_\mu + \frac{1}{2}\vec{k}^2/m_\mu + \cdots \simeq m_\mu$. Also, this means the collision is approximately elastic. In the diagram of the kinematics at right, $c \equiv \cos\theta, s \equiv \sin\theta$.

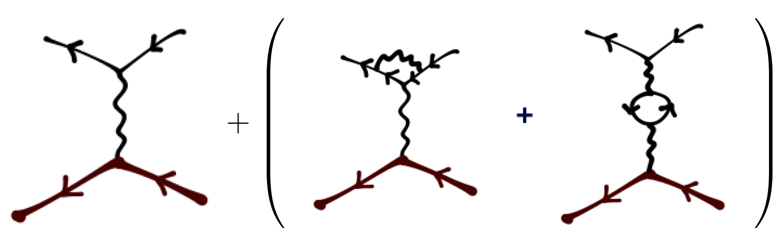


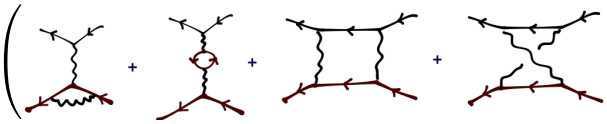
The answer we found after some boiling was:

$$\frac{d\sigma}{d\Omega}_{\text{Mott}} = \frac{\alpha^2(1 - \beta^2\sin^2\theta/2)}{4\beta^2\vec{p}^2\sin^4\theta/2}.$$

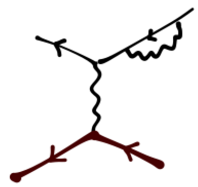
If we take $\beta \ll 1$ in this formula we get the Rutherford formula.

**Radiative corrections.** Now it's time to think about perturbative corrections to this cross section. Given that the leading-order calculation reproduced the classical physics of the Coulomb potential, you can think of what we are doing as effectively discovering (high-energy or short-distance) quantum corrections to the Coulomb law. The diagrams we must include are these (I made the muon lines thicker and also red):

$$\mathbf{i}\mathcal{M}_{e\mu\leftarrow e\mu} =$$ 

$$+ \Bigg( \quad$$  $$\Bigg) + \mathcal{O}(e^6)$$

• What do the one-loop diagrams in the second line have in common? They have an internal muon line. Why does this matter? When the energy going through the line is much smaller than the muon mass, then the propagator is $\frac{\mathbf{i}(\not{k}+m_\mu)}{k^2-m_\mu^2} \sim \frac{1}{m_\mu}$ and its relative contribution is down by $k/m_\mu \ll 1$. So let's neglect these for now.

• Why don't we include diagrams like  ? The LSZ formula tells us that their effects on the $S$-matrix are accounted for by the wavefunction renormalization factors $Z$

$$S_{e\mu\leftarrow e\mu} = \sqrt{Z_e}^2 \sqrt{Z_\mu}^2 \Bigg($$  $$+ \cdots \Bigg)_{\text{amputated, on-shell}}$$

and in determining the locations of the poles whose residues are the $S$-matrix elements.

• Notice that the one-loop *amplitudes* are suppressed relative to the tree level amplitude by two factors of $e$, hence one factor of the fine structure constant $\alpha = \frac{e^2}{4\pi}$. Their leading effects on the cross section come from

$$\sigma \sim \Big|$$  $$+ \cdots \Big|^2 \sim \sigma_{\text{tree}} + \mathcal{O}(\alpha^3)$$

from the cross term between the tree and one-loop amplitudes.

In the above discussion, we encounter all three 'primitive' one-loop divergent amplitudes of QED, which we'll study in turn:

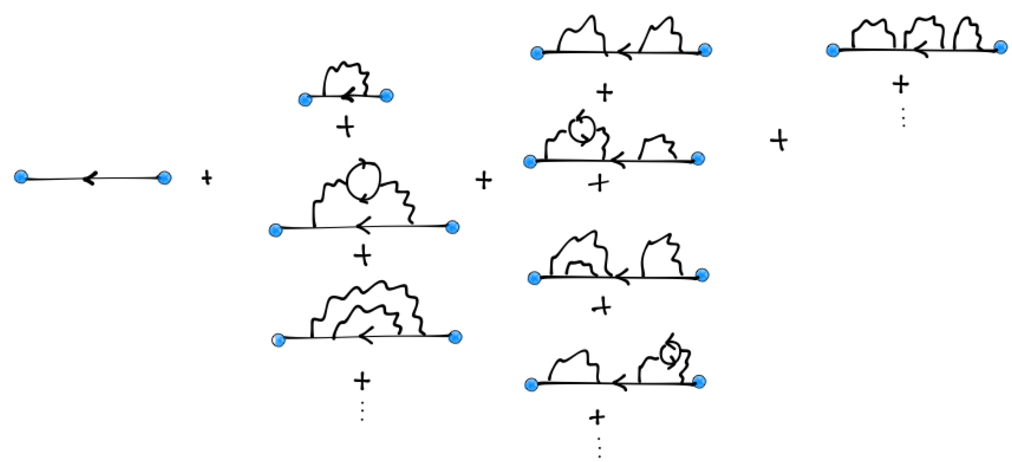
• electron self-energy: 

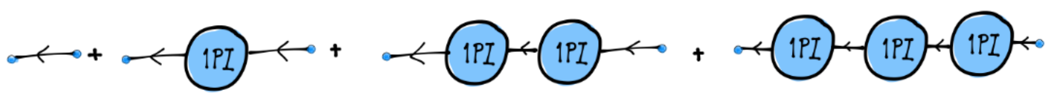
• vertex correction: 

• vacuum polarization (photon self-energy): 

## 7.4 Electron self-energy in QED

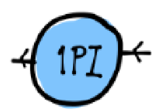
Let's think about the electron two-point function in momentum space:

$\tilde{G}^{(2)}(p) =$  $+ \cdots$

$=$  $+ \cdots$

(7.10)

As we did for the scalar field theory in §3 last quarter, we will denote the 1PI two-point function by

$-\mathbf{i}\Sigma(p) \equiv$ 


a blob with nubbins; for fermions with conserved particle number, the nubbins carry arrows indicating the particle number flow. Let me call the tree level propagator

$$\mathbf{i}S(p) \equiv \frac{\mathbf{i}(\not{p} + m_0)}{p^2 - m_0^2 + \mathbf{i}\epsilon} = \frac{\mathbf{i}}{\not{p} - m_0}$$

– notice that I added a demeaning subscript to the notation for the mass appearing in the Lagrangian. Foreshadowing.

The full two point function is then:

$$\begin{aligned} \tilde{G}^{(2)}(p) &= \mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \cdots \\ &= \mathbf{i}S\left(1 + \Sigma S + \Sigma S \Sigma S + \cdots\right) = \mathbf{i}S\frac{1}{1 - \Sigma S} \\ &= \frac{\mathbf{i}}{\not{p} - m_0}\frac{1}{1 - \Sigma\frac{1}{\not{p} - m_0}} = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)}. \end{aligned} \tag{7.11}$$

Are you worried about these manipulations because $\Sigma$ and $S$ are matrices in the spinor indices? Don't be: they are both made entirely from $\not{p}$, and therefore they commute;

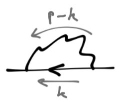
we could do these manipulations in the eigenbasis of $\not{p}$. This fully corrected propagator has a pole at

$$\not{p} = m \equiv m_0 + \Sigma(m) \ . \tag{7.12}$$

This means that the actual mass of the particle is this new quantity $m$. But what is $m$ (it is called the 'renormalized mass')? To figure it out, we need to know about $\Sigma$.

In QED we must study $\Sigma$ in perturbation theory. As you can see from (7.10), the leading (one-loop) contribution is

$$-\mathbf{i}\Sigma_2(p) =$$




$$= (-\mathbf{i}e)^2 \int đ^4k\ \gamma^\mu \frac{\mathbf{i}(\not{k} + m_0)}{k^2 - m_0^2 + \mathbf{i}\epsilon} \gamma^\nu \frac{-\mathbf{i}\eta_{\mu\nu}}{(p-k)^2 - \mu^2 + \mathbf{i}\epsilon} \ .$$

Notice that I am relying on the Ward identity to enforce the fact that only the traverse bit of the photon propagator matters. Also, I added a mass $\mu$ for the photon as an IR regulator. We must keep the external momentum $p$ arbitrary, since we don't even know where the mass-shell is!

Finally, I can't put it off any longer: how are we going to do this loop-momentum integral?

Step 1: Feynman parameter trick. It is a good idea to consider the integral

$$\int_0^1 dx \frac{1}{(xA + (1-x)B)^2} = \int_0^1 dx \frac{1}{(x(A-B) + B)^2} = \frac{1}{A-B} \frac{-1}{x(A-B)+B}\Bigg|_{x=0}^{x=1}$$
$$= \frac{1}{A-B}\left(-\frac{1}{A} + \frac{1}{B}\right) = \frac{1}{AB}.$$

This allows us to combine the denominators into one:

$$\mathcal{I} = \underbrace{\frac{1}{k^2 - m_0^2 + \mathbf{i}\epsilon}}_{B} \underbrace{\frac{1}{(p-k)^2 - \mu^2 + \mathbf{i}\epsilon}}_{A} = \int_0^1 dx \frac{1}{\left(x\left((p^2 - 2pk + k^2) - \mu^2 + \mathbf{i}\epsilon\right) + (1-x)(k^2 - m_0^2 + \mathbf{i}\epsilon)\right)^2}$$

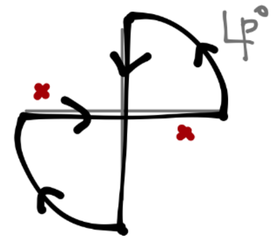
Step 2: Now we can complete the square

$$\mathcal{I} = \int_0^1 dx \frac{1}{\left(\underbrace{(k - px)}_{\equiv \ell}{}^2 - \Delta + \mathbf{i}\epsilon\right)^2}$$

with

$$\ell^\mu \equiv k^\mu - p^\mu x, \quad \Delta \equiv +p^2x^2 + x\mu^2 - xp^2 + (1-x)m_0^2 = x\mu^2 + (1-x)m_0^2 - x(1-x)p^2.$$

Step 3: Wick rotate. Because of the $\mathbf{i}\epsilon$ we've been dutifully carrying around, the poles of the $p^0$ integral don't occur in the first and third octants of the complex $p^0$ plane. (And the integrand decays at large $|p^0|$.) This means that we can rotate the contour to euclidean time for free: $\ell^0 \equiv \mathbf{i}\ell^4$. Equivalently: the integral over the contour at right vanishes, so the real time contour gives the same answer as the (upward-directed) Euclidean contour.



Notice that $\ell^2 = -\ell_E^2$. Altogether

$$-\mathbf{i}\Sigma_2(p) = -e^2 \int đ^4\ell \int_0^1 dx \frac{N}{(\ell^2 - \Delta + \mathbf{i}\epsilon)^2} = -e^2 \int_0^1 dx \mathbf{i} \int đ^4\ell_E \frac{N}{(\ell_E^2 + \Delta)^2}$$

where the numerator is

$$N = \gamma^\mu \left(\not{\ell} + x\not{p} + m_0\right) \gamma_\mu = -2\left(\not{\ell} + x\not{p}\right) + 4m_0.$$

Here I used two Clifford algebra facts: $\gamma^\mu\gamma_\mu = 4$ and $\gamma^\mu \not{p} \gamma_\mu = -2\not{p}$. Think about the contribution from the term with $\not{\ell}$ in the numerator: everything else is invariant under rotations of $\ell$

$$đ^4\ell_E = \frac{1}{(2\pi)^4} d\Omega_3 \ell^3 d\ell = \frac{d\Omega_3}{(2\pi)^4} \ell^2 \frac{d\ell^2}{2},$$

so this averages to zero. The rest is of the form (using $\int_{S^3} d\Omega_3 = 2\pi^2$)

$$\begin{aligned} \Sigma_2(p) &= e^2 \int_0^1 dx \int \frac{\ell^2 d\ell^2}{2} \frac{(2\pi^2)}{(2\pi)^4} \frac{2(2m_0 - x\not{p})}{(\ell^2 + \Delta)^2} \\ &= \frac{e^2}{8\pi^2} \int_0^1 dx (2m_0 - x\not{p}) \mathcal{J} \end{aligned} \tag{7.13}$$

with

$$\mathcal{J} = \int_0^\infty d\ell^2 \frac{\ell^2}{(\ell^2 + \Delta)^2}.$$

In the large $\ell$ part of the integrand this is

$$\int^\Lambda \frac{d\ell^2}{\ell^2} \sim \log \Lambda.$$

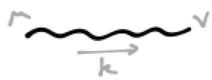
You knew this UV divergence was coming. To be more precise, let's add zero:

$$\begin{aligned} \mathcal{J} &= \int d\ell^2 \left( \frac{\ell^2 + \Delta}{(\ell^2 + \Delta)^2} - \frac{\Delta}{(\ell^2 + \Delta)^2} \right) \\ &= \int_0^\infty d\ell^2 \left( \frac{1}{\ell^2 + \Delta} - \frac{\Delta}{(\ell^2 + \Delta)^2} \right) = \ln(\ell^2 + \Delta)\Big|_{\ell^2=0}^\infty + \frac{\Delta}{\ell^2 + \Delta}\Big|_{\ell^2=0}^\infty = \ln(\ell^2 + \Delta)\Big|_{\ell^2=0}^\infty - 1. \end{aligned}$$

Recall that

$$\Delta = x\mu^2 + (1 - x)m_0^2 - x(1 - x)p^2 \equiv \Delta(\mu^2).$$

**Pauli-Villars regularization.** Here is a convenient fiction: when you exchange a photon, you also exchange a very heavy particle, with mass $m^2 = \Lambda^2$, with an extra $(-1)$ in its propagator. This means that (in this Pauli-Villars regulation scheme) the Feynman rule for the wiggly line is instead




$$\begin{aligned} &= -\mathbf{i}\eta_{\mu\nu}\left(\frac{1}{k^2-\mu^2+\mathbf{i}\epsilon} - \frac{1}{k^2-\Lambda^2+\mathbf{i}\epsilon}\right) \\ &= -\mathbf{i}\eta_{\mu\nu}\left(\frac{\mu^2-\Lambda^2}{(k^2-\mu^2+\mathbf{i}\epsilon)(k^2-\Lambda^2+\mathbf{i}\epsilon)}\right) \end{aligned}$$

This goes like $\frac{1}{k^4}$ at large $k$, so the integrals are more convergent. Yay.

Notice that the contribution from the Pauli-Villars photon to tree-level amplitudes goes like $|\frac{1}{k^2-\Lambda^2}| \overset{\Lambda \gg k}{\sim} \frac{1}{\Lambda^2}$ (where $k$ is the momentum going through the photon line, determined by the external momenta), which innocuously vanishes as $\Lambda \to \infty$.

Remembering that the residue of the pole in the propagator is the probability for the field operator to create a particle from the vacuum, you might worry that this is a negative probability, and unitarity isn't manifest. This particle is a *ghost*. However, we will choose $\Lambda$ so large that the pole in the propagator at $k^2 = \Lambda^2$ will never be accessed and we'll never have external Pauli-Villars particles. We are using this as a device to define the theory in a regime of energies much less than $\Lambda$. You shouldn't take the regulated theory too seriously: for example, the wrong-sign propagator means wrong-sign kinetic terms for the PV fields. This means that very wiggly configurations will be energetically *favored* rather than suppressed by the Hamiltonian. It will not make much sense non-perturbatively.

I emphasize that this regulator is one possibility of many. They each have their drawbacks. They all break scale invariance. A nice thing about PV is that it is Lorentz invariant. A class of regulators which make perfect sense non-perturbatively is the lattice (as in the model with masses on springs). The price is that it really messes up the spacetime symmetries.

---

Applying this to the self-energy integral amounts to the replacement

$$\begin{aligned} \mathcal{J} &\rightsquigarrow \mathcal{J}_{\Delta(\mu^2)} - \mathcal{J}_{\Delta(\Lambda^2)} \\ &= \left[\left(\ln\left(\ell^2+\Delta(\mu^2)\right)-1\right) - \left(\ln\left(\ell^2+\Delta(\Lambda^2)\right)-1\right)\right]\Big|_0^\infty \\ &= \ln\frac{\ell^2+\Delta(\mu^2)}{\ell^2+\Delta(\Lambda^2)}\Bigg|_0^\infty \\ &= \ln 1/1 - \ln\frac{\Delta(\mu^2)}{\Delta(\Lambda^2)} = \ln\frac{\Delta(\Lambda^2)}{\Delta(\mu^2)}. \end{aligned}$$

[End of Lecture 22]

Notice that we can take advantage of our ignorance of the microphysics to make the cutoff as big as we like and thereby simplify our lives:

$$\Delta(\Lambda^2) = x\Lambda^2 + (1-x)m_0^2 - x(1-x)p^2 \overset{\Lambda \gg \text{everyone}}{\approx} x\Lambda^2.$$

Finally then

$$\Sigma_2(p)_{PV} = \frac{\alpha}{2\pi}\int_0^1 dx(2m_0 - x\not{p}) \ln \frac{x\Lambda^2}{x\mu^2 + (1-x)m_0^2 - x(1-x)p^2}. \tag{7.14}$$

Having arrived at this regulated expression for the self-energy we need to "impose a renormalization condition," *i.e.* introduce some observable physics in terms of which to parametrize our answers. We return to (7.12): the shift in the mass as a result of this one-loop self-energy is

$$\begin{aligned}
\delta m \equiv m - m_0 &= \Sigma_2(\not{p} = m) + \mathcal{O}(e^4) = \Sigma_2(\not{p} = m_0) + \mathcal{O}(e^4) \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \ln \frac{x\Lambda^2}{\underbrace{x\mu^2 + (1-x)m_0^2 + x(1-x)m_0^2}_{=x\mu^2+(1-x^2)m_0^2 \equiv f(x,m_0,\mu)}} \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \left( \underbrace{\ln\frac{\Lambda^2}{m_0^2}}_{\text{divergent}} + \underbrace{\ln\frac{xm_0^2}{f(x,m_0,\mu)}}_{\text{relatively small}} \right) \\
&\approx \frac{\alpha}{2\pi}\left(2 - \frac{1}{2}\right) m_0 \ln\frac{\Lambda^2}{m_0^2} = \frac{3\alpha}{4\pi} m_0 \ln\frac{\Lambda^2}{m_0^2}.
\end{aligned} \tag{7.15}$$

In the penultimate step (with the $\approx$), we've neglected the finite bit (labelled 'relatively small') compared to the logarithmically divergent bit: we've already assumed $\Lambda \gg$ all other scales in the problem.

**Mass renormalization.** Now the physics input: The mass of the electron is 511 keV (you can ask how we measure it and whether the answer we get depends on the resolution of the measurement, and indeed there is more to this story; this is a low-energy answer, for example we could make the electron go in a magnetic field and measure the radius of curvature of its orbit and set $m_e v^2/r = evB/c$), so

$$511 \text{ keV} \approx m_e = m_0 \left(1 + \frac{3\alpha}{4\pi}\ln\frac{\Lambda^2}{m_0^2}\right) + \mathcal{O}(\alpha^2).$$

In this equation, the LHS is a measured quantity. In the correction on the RHS $\alpha \approx \frac{1}{137}$ is small, but it is multiplied by $\ln\frac{\Lambda^2}{m_0}$ which is arbitrarily large. This means that the bare mass $m_0$, which is going to absorb the cutoff dependence here, must actually be really small. (Notice that actually I've lied a little here: the $\alpha$ we've been using is still the bare charge; we will need to renormalize that one, too, before we are done.) I emphasize: $m_0$ and the other fake, bare parameters in $\mathcal{L}$ depend on $\Lambda$ and the order of perturbation theory to which we are working and other theorist bookkeeping garbage; $m_e$ does not. At each order in perturbation theory, we eliminate $m_0$ and write our predictions in terms of $m_e$. It is not too surprising that the mass of the electron includes such contributions: it must be difficult to travel through space if you are constantly emitting and re-absorbing photons.

**Wavefunction renormalization.** The actual propagator for the electron, near the electron pole is

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)} \overset{p\sim m}{\simeq} \frac{\mathbf{i}Z}{\not{p} - m} + \text{regular terms}. \tag{7.16}$$

The residue of the pole at the electron mass is no longer equal to one, but rather $Z$. To see what $Z$ actually is at this order in $e^2$, Taylor expand near the pole

$$\begin{aligned}\Sigma(p) &\overset{\text{Taylor}}{=} \Sigma(\not{p} = m) + \frac{\partial\Sigma}{\partial\not{p}}|_{\not{p}=m}(\not{p} - m) + \cdots \\ &= \Sigma(\not{p} = m_0) + \frac{\partial\Sigma}{\partial\not{p}}|_{\not{p}=m_0}(\not{p} - m_0) + \cdots + \mathcal{O}(e^4)\end{aligned}$$

So then (7.16) becomes

$$\tilde{G}^{(2)}(p) \overset{p\sim m}{\sim} \frac{\mathbf{i}}{\not{p} - m - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}(\not{p} - m)} = \frac{\mathbf{i}}{(\not{p} - m)\left(1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}\right)} \tag{7.17}$$

So that

$$Z = \frac{1}{1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}} \simeq 1 + \frac{\partial\Sigma}{\partial\not{p}}|_{m_0} \equiv 1 + \delta Z$$

and at leading nontrivial order

$$\begin{aligned}\delta Z = \frac{\partial\Sigma_2}{\partial\not{p}}|_{m_0} &\overset{(7.14)}{=} \frac{\alpha}{2\pi}\int_0^1 dx\left(-x\ln\frac{x\Lambda^2}{f(x,m_0,\mu)} + (2m_0 - xm_0)\frac{-2x(1-x)}{f(x,m_0,\mu)}\right) \\ &= -\frac{\alpha}{4\pi}\left(\ln\frac{\Lambda^2}{m_0^2} + \text{finite}\right).\end{aligned} \tag{7.18}$$

Here $f = f(x, m_0, \mu)$ is the same quantity defined in the second line of (7.15). We'll see below that the cutoff-dependence in $\delta Z$ plays a crucial role in making the $S$ matrix (for example for the $e\mu \to e\mu$ process we've been discussing) cutoff-independent and finite, when written in terms of physical variables.

## 7.5 Big picture interlude

OK, I am having a hard time just pounding away at one-loop QED. Let's take a break and think about the self-energy corrections in scalar field theory. Then we will step back and think about the general structure of short-distance senstivity in (relativistic) QFT, before returning to the QED vertex correction and vacuum polarization.

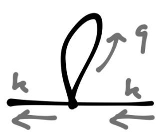
### 7.5.1 Self-energy in $\phi^4$ theory

[Zee §III.3] Let's return to the $\phi^4$ theory in $D = 3+1$ for a moment. The $\mathcal{M}_{\phi\phi\leftarrow\phi\phi}$ amplitude is not the only place where the cutoff appears.

Above we added a counterterm of the same form as the $\phi^4$ term in the Lagrangian. Now we will see that we need counterterms for everybody:

$$\mathcal{L} = -\frac{1}{2}\left(\phi\Box\phi + m^2\phi^2\right) - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 + \frac{1}{2}\delta Z\phi\Box\phi + \frac{1}{2}\delta m^2\phi^2.$$

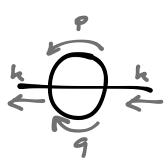
Here is a way in which $\phi^4$ theory is weird: At one loop there is no wavefunction renormalization. That is,

$$\delta\Sigma_1(k) =$$


$$= -\mathbf{i}g\int^\Lambda đ^4q\frac{\mathbf{i}}{q^2-m^2+\mathbf{i}\epsilon} = \delta\Sigma_1(k=0) \sim g\Lambda^2$$

which is certainly quadratically divergent, but totally independent of the external momentum. This means that when we Taylor expand in $k$ (as we just did in (7.17)), this diagram only contributes to the mass renormalization.

So let's see what happens if we keep going:

$$\delta\Sigma_2(k) =$$


$$= (-\mathbf{i}g)^2\int đ^4p\int đ^4q\,\mathbf{i}D_0(p)\mathbf{i}D_0(q)\mathbf{i}D_0(k-p-q) \equiv I(k^2, m, \Lambda).$$

Here $\mathbf{i}D_0(p) \equiv \frac{\mathbf{i}}{p^2-m^2+\mathbf{i}\epsilon}$ is the free propagator (the factor of $\mathbf{i}$ is for later convenience), and we've defined $I$ by this expression. The fact that $I$ depends only on $k^2$ is a consequence of Lorentz invariance. Counting powers of the loop momenta, the short-distance bit of this integral is of the schematic form $\int^\Lambda \frac{d^8P}{P^6} \sim \Lambda^2$, also quadratically divergent, but this time $k^2$-dependent, so there will be a nonzero $\delta Z \propto g^2$. As we just did for the electron self-energy, we should Taylor expand in $k$. (We'll learn more about why and when the answer is analytic in $k^2$ at $k = 0$ later.) The series expansion in $k^2$ (let's do it about $k^2 = 0 \sim m^2$ to look at the UV behavior) is

$$\delta\Sigma_2(k^2) = A_0 + k^2 A_1 + k^4 A_2 + \cdots$$

where $A_0 = I(k^2 = 0) \sim \Lambda^2$. In contrast, dimensional analysis says $A_1 = \frac{\partial}{\partial k^2} I|_{k^2=0} \sim \int \frac{d^8 P}{P^8} \sim \Lambda^{0^+} \sim \ln \Lambda$ has two fewer powers of the cutoff. After that it's clear sailing: $A_2 = \left(\frac{\partial}{\partial k^2}\right)^2 I|_{k^2=0} \sim \int^\Lambda \frac{d^8 P}{P^{10}} \sim \Lambda^{-2}$ is finite as we remove the cutoff, and so are all the later coefficients.

Putting this together, the inverse propagator is

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m^2 - \underbrace{(\delta\Sigma_1(0) + A_0)}_{\equiv a \sim \Lambda^2} - k^2 A_1 - k^4 A_2 + \cdots$$

The $\cdots$ here includes both higher orders in $g$ ($\mathcal{O}(g^3)$) and higher powers of $k^2$, *i.e.* higher derivative terms. If instead the physical pole were at a nonzero value of the mass, we should Taylor expand about $k^2 = m_P^2$ instead:

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m_0^2 - \underbrace{(\delta\Sigma_1(0) + A_0)}_{\equiv a \sim \Lambda^2} - (k^2 - m_P^2) A_1 - (k^2 - m_P^2)^2 A_2 + \cdots$$

where now $A_n \equiv \frac{1}{n!} \left(\frac{\partial}{\partial k^2}\right)^n \Sigma_2(k^2)|_{k^2 = m_P^2}$.

Therefore, the propagator is

$$D(k) = \frac{1}{(1 - A_1)(k^2 - m_P^2)} + \cdots = \frac{Z}{k^2 - m_P^2} + \cdots$$

with

$$Z = \frac{1}{1 - A_1}, \quad m_P^2 = m^2 + a.$$

Some points to notice: • $\delta Z = A_1$.

• The contributions $A_{n \geq 2}(k^2)^n$ can be reproduced by counterterms of the form $A_n \phi \Box^n \phi$. Had they been cutoff dependent we would have needed to add such (cutoff-dependent) counterterms.

• The mass-squared of the scalar field in $D = 3+1$ is quadratically divergent, while the mass of the spinor was only log divergent. This UV sensitivity of scalar fields is ubiquitous[5] (see the homework) and is the source of many headaches.

• On the term 'wavefunction renormalization': who is $\phi$? Also just a theorist's letter. Sometimes (in condensed matter) it is defined by some relation to observation (like

[5]At least for most regulators. We'll see that dim reg is special.

the height of a wave in the mattress), in high energy theory not so much. Classically, we fixed its (multiplicative) normalization by setting the coefficient of $\phi\Box\phi$ to one. If we want to restore that convention after renormalization, we can make a redefinition of the field $\phi_R \equiv Z^{-1/2}\phi$. This is the origin of the term 'wavefunction renormalization'. A slightly better name would be 'field renormalization', but even better would be just 'kinetic term renormalization'.

**Renormalized perturbation theory revisited.** The full story for the renormalized perturbation expansion in $\phi^4$ theory is

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m_P^2\phi^2 - \frac{g_P}{4!}\phi^4 + \mathcal{L}_{ct}$$

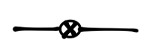
with

$$\mathcal{L}_{ct} = \frac{1}{2}\delta Z\,(\partial\phi)^2 + \frac{1}{2}\delta m^2\phi^2 + \frac{\delta_g}{4!}\phi^4.$$

Here are the instructions for using it: The Feynman rules are as before: the coupling and propagator are

 $= -\mathbf{i}g_P,$  $= \dfrac{\mathbf{i}}{k^2 - m_P^2 + \mathbf{i}\epsilon}$ (7.19)

but the terms in $\mathcal{L}_{ct}$ (the counterterms) are treated as new vertices, and treated perturbatively:

 $= \mathbf{i}\delta_g,$  $= \mathbf{i}(\delta Z k^2 + \delta m^2).$

All integrals are regulated, in the same way (whatever it is). The counterterm couplings $\delta_g, \delta Z, \delta m^2$ are determined iteratively, as follows: given the $\delta_{N-1}$s up to $\mathcal{O}(g_P^N)$, we fix each one $\delta = \delta_{N-1} + g_P^N \Delta\delta_N + \mathcal{O}(g_P^{N+1})$ by demanding that (7.19) are actually true up to $\mathcal{O}(g_P^{N+1})$. This pushes the cutoff dependence back into the muck a bit further.

I say this is the full story, but wait: we didn't try to compute amplitudes with more than four $\phi$s (such as $3 \leftarrow 3$ scattering of $\phi$ quanta). How do we know those don't require new counterterms (like a $\phi^6$ term, for example)?

### 7.5.2 Where is the UV sensitivity?

[still Zee §III.3, Peskin ch. 10. We'll follow Zee's discussion pretty closely for a bit.] Given some process in a relativistic, perturbative QFT, how do we know if it will depend on the cutoff? We'd like to be able answer this question for a theory with scalars, spinors, vectors. Here's how: First, look at each diagram $\mathcal{A}$ (order by order in the loop expansion). Define the 'superficial' degree of divergence of $\mathcal{A}$ to be $D_{\mathcal{A}}$ if

$\mathcal{A} \sim \Lambda^{D_\mathcal{A}}$ (in the limit that $\Lambda \ll$ all other scales – this is an asymptotic statement). A log divergent amplitude has $D_\mathcal{A} = 0$ (sometimes it's called $D_\mathcal{A} = 0^+$).

Let's start simple, and study the $\phi^4$ theory in $D = 4$. Consider a connected diagram $\mathcal{A}$ with $B_E$ external scalar lines. I claim that $D_\mathcal{A} = 4 - B_E$. Why didn't it depend on any other data of the diagram, such as

$$\begin{aligned} B_I &\equiv \# \text{ of internal scalar lines } (\textit{i.e.}, \text{ propagators}) \\ V &\equiv \# \text{ of } \phi^4 \text{ vertices} \\ L &\equiv \# \text{ of loops} \end{aligned}$$

$B_I = 8$
$B_E = 4$
$V = 5$
$L = 4$

? We can understand this better using two facts of graph theory and some power counting. I recommend checking my claims below with an example, such as the one at right.

**Graph theory fact #1:** These quantities are not all independent. For a connected graph,

$$L = B_I - (V - 1). \tag{7.20}$$

Math proof[6]: Imagine placing the vertices on the page and adding the propagators one at a time. You need $V - 1$ internal lines just to connect up all $V$ vertices. After that, each *internal* line you add necessarily adds one more loop. ■

Another way to think about this fact makes clear that $L = \#$ of loops $= \#$ of momentum integrals. Before imposing momentum conservation at the vertices, each internal line has a momentum which we must integrate: $\prod_{\alpha=1}^{B_I} \int đ^D q_\alpha$. We then stick a $\delta^{(D)}(\sum q)$ for each vertex, but one of these gives the overall momentum conservation $\delta^{(D)}(k_T)$, so we have $V - 1$ fewer momentum integrals. For the example above, (7.20) says $4 = 8 - (5 - 1)$.

**Graph theory fact #2:** Each external line comes out of one vertex. Each internal line connects two vertices. Altogether, the number of ends of lines sticking out of vertices is

$$B_E + 2B_I = 4V$$

where the RHS comes from noting that each vertex has four lines coming out of it (in $\phi^4$ theory). In the example, this is $4 + 2 \cdot 8 = 4 \cdot 5$. So we can eliminate

$$B_I = 2V - B_E/2. \tag{7.21}$$

[6] I learned this one from my class-mate M.B. Schulz.

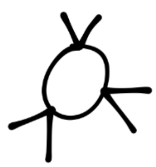
Now we count powers of momenta:

$$\mathcal{A} \sim \prod_{a=1}^{L} \int^{\Lambda} đ^D k_a \prod_{\alpha=1}^{B_I} \frac{1}{k_\alpha^2}.$$

Since we are interested in the UV structure, I've set the mass to zero, as well as all the external momenta. The only scale left in the problem is the cutoff, so the dimensions of $\mathcal{A}$ must be made up by the cutoff:

$$\begin{aligned} D_{\mathcal{A}} = [\mathcal{A}] &= DL - 2B_I \\ &\overset{(7.20)}{=} B_I(D-2) - D(V-1) \\ &\overset{(7.21)}{=} D + \frac{2-D}{2} B_E + V(D-4). \end{aligned}$$

If we set $D = 3+1 = 4$, we get $D_{\mathcal{A}} = 4 - B_E$ as claimed. Notice that with $B_E = 2$ we indeed reproduce $D_{\mathcal{A}} = 2$, the quadratic divergence in the mass renormalization, and with $B_E = 4$ we get $D_{\mathcal{A}} = 0$, the log divergence in the $2 \leftarrow 2$ scattering. This pattern continues: with more than four external legs, $D_{\mathcal{A}} = 4 - B_E < 0$, which means the cutoff dependence must go away when $\Lambda \to 0$. This is illustrated by the following diagram with $B_E = 6$:

 $\sim \int^{\Lambda} \frac{đ^4 P}{P^6} \sim \Lambda^{-2}.$

So indeed we don't need more counterterms for higher-point interactions in this theory.

[End of Lecture 23]

Why is the answer independent of $V$ in $D = 4$? This has the dramatic consequence that once we fix up the cutoff dependence in the one-loop diagrams, the higher orders have to work out, *i.e.* it strongly suggests that the theory is renormalizable. [7]

Before we answer this, let's explore the pattern a bit more. Suppose we include also a fermion field $\psi$ in our field theory, and suppose we couple it to our scalar by a

[7] Why isn't it a proof of renormalizability? Consider the following integral:

$$\mathcal{I} = \int^{\Lambda} \frac{d^4 p}{(p^2 + m^2)^5} \int^{\Lambda} d^4 k.$$

According to our method of counting, we would say $D_{\mathcal{I}} = 4 + 4 - 10 = -2$ and declare this finite and cutoff-independent. On the other hand, it certainly does depend on the physics at the cutoff. (I bet it is possible to come up with more pathological examples.) The rest of the work involving 'nested divergences' and forests is in showing that the extra structure in the problem prevents things like $\mathcal{I}$ from being Feynman amplitudes.

Yukawa interaction:

$$S_{\text{bare}}[\phi,\psi] = -\int d^D x \left( \frac{1}{2}\phi\left(\Box + m_\phi^2\right)\phi + \bar{\psi}\left(-\partial\!\!\!/ + m_\psi\right)\psi + y\phi\bar{\psi}\psi + \frac{g}{4!}\phi^4 \right).$$

To find the degree of divergence in an amplitude in this model, we have to independently keep track of the number fermion lines $F_E, F_I$, since a fermion propagator has dimension $[\frac{1}{p\!\!\!/}] = -1$, so that $D_{\mathcal{A}} = [\mathcal{A}] = DL - 2B_I - F_I$. The number of ends-of-fermion-lines is $2V_y = 2F_E + F_I$ and the number of ends-of-boson-lines is $V_y + 4V_g = B_E + 2B_I$. The number of loops is $L = B_I + F_I - (V_y + V_g - 1)$. Putting these together (I used Mathematica) we get

$$D_{\mathcal{A}} = D + (D-4)\left(V_g + \frac{1}{2}V_y\right) + B_E\left(\frac{2-D}{2}\right) + F_E\left(\frac{1-D}{2}\right). \tag{7.22}$$

Again in $D = 4$ the answer is independent of the number of vertices! Is there something special about four spacetime dimensions?

To temper your enthusiasm, consider adding a four-fermion interaction: $G(\bar{\psi}\psi)(\bar{\psi}\psi)$ (or maybe $G_V(\bar{\psi}\gamma^\mu\psi)(\bar{\psi}\gamma_\mu\psi)$ or $G_A(\bar{\psi}\gamma^\mu\gamma^5\psi)(\bar{\psi}\gamma_\mu\gamma^5\psi)$ or any other pile of gamma matrices in between, with the indices contracted). When you redo this calculation on the homework, you'll find that in $D = 4$ a diagram (for simplicity, one with no $\phi^4$ or Yukawa interactions) has

$$D_{\mathcal{A}} = 4 - (1)B_E - \left(\frac{3}{2}\right)F_E + 2V_G.$$

This dependence on the number of four-fermi vertices means that there are worse and worse divergences as we look at higher-order corrections to a given process. Even worse, it means that for any number of external lines $F_E$ no matter how big, there is a large enough order in perturbation theory in $G$ where the cutoff will appear! This means we need $\delta_n(\bar{\psi}\psi)^n$ counterterms for every $n$, which we'll need to fix with physical input. This is a bit unappetizing, and such an interaction is called "non-renormalizable". However, when we remember that we only need to make predictions to a given precision (so that we only need to go to a finite order in this process) we will see that such theories are nevertheless quite useful.

So why were those other examples independent of $V$? It's because the couplings were dimensionless. Those theories were classically scale invariant (except for the mass terms).

### 7.5.3 Naive scale invariance in field theory

[Halpern] Consider a field theory of a scalar field $\phi$ in $D$ spacetime dimensions, with an action of the form

$$S[\phi] = \int d^D x \left( \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - g\phi^p \right)$$

for some constants $p, g$. Which value of $p$ makes this scale invariant? (That is: when is $g$ dimensionless, and hence possibly the coupling for a renormalizable interaction.)

Naive dimensions:

$$[S] = [\hbar] = 0, \quad [x] \equiv -1, \quad [d^D x] = -D, \quad [\partial] = 1$$

The kinetic term tells us the engineering dimensions of $\phi$:

$$0 = [S_{\text{kinetic}}] = -D + 2 + 2[\phi] \implies [\phi] = \frac{D-2}{2}.$$

Notice that the $D = 1$ case agrees with our quantum mechanics counting from §7.1. Quantum field theory in $D = 1$ spacetime dimensions *is* quantum mechanics.

Then the self-interaction term has dimensions

$$0 = [S_{\text{interaction}}] = -D + [g] + p[\phi] \implies [g] = D - p[\phi] = D + p\frac{2-D}{2}$$

We expect scale invariance when $[g] = 0$ which happens when

$$\boxed{p = p_D \equiv \frac{2D}{D-2}},$$

*i.e.* the scale invariant scalar-field self-interaction in $D$ spacetime dimensions is $\phi^{\frac{2D}{D-2}}$.

| $D$ | 1 | 2 | 3 | 4 | 5 | 6 | ... | $\infty$ |
|---|---|---|---|---|---|---|---|---|
| $[\phi]$ | $-\frac{1}{2}$ | 0 | $\frac{1}{2}$ | 1 | 3/2 | 2 | ... | $D/2$ |
| scale-inv't $p \equiv p_D$ | $-2$ | $\infty\star$ | 6 | 4 | 10/3 | 3 | ... | 2 |

$\star$ What is happening in $D = 2$? The field is dimensionless, and so *any* power of $\phi$ is naively scale invariant, as are more complicated interactions like $g_{ij}(\phi)\partial_\mu \phi^i \partial^\mu \phi^j$, where the coupling $g(\phi)$ is a function of $\phi$. This allows for scale-invariant *non-linear sigma models*, where the fields are coordinates on a curved manifold with metric $ds^2 = g_{ij} d\phi^i d\phi^j$.

In dimensions where we get fractional powers, this isn't so nice.

Notice that the mass term $\Delta S = \int d^D x \frac{m^2}{2}\phi^2$ gives

$$0 = -D + 2[m] + 2[\phi] \quad \implies \quad [m] = 1 \ \ \forall D < \infty$$

– it's a mass, yay.

What are the consequences of this engineering dimensions calculation in QFT? For $D > 2$, an interaction of the form $g\phi^p$ has

$$[g] = D \cdot \frac{p_D - p}{p_D} \begin{cases} < 0 \text{ when } p > p_D, & \text{non-renormalizable or irrelevant} \\ = 0 \text{ when } p = p_D, & \text{renormalizable or marginal} \\ > 0 \text{ when } p < p_D, & \text{super-renormalizable or relevant.} \end{cases} \tag{7.23}$$

Consider the 'non-renormalizable' case. Suppose we calculate in QFT some quantity $f$ with $[f]$ as its naive dimension, in perturbation theory in $g$, *e.g.* by Feynman diagrams. We'll get:

$$f = \sum_{n=0}^{\infty} g^n c_n$$

with $c_n$ independent of $g$. So

$$[f] = n[g] + [c_n] \quad \implies \quad [c_n] = [f] - n[g]$$

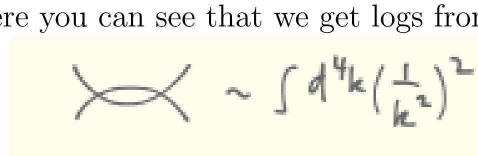
So if $[g] < 0$, $c_n$ must have more and more powers of some mass (inverse length) as $n$ increases. What dimensionful quantity makes up the difference? Sometimes it is masses or external momenta. But generically, it gets made up by UV divergences (if everything is infinite, dimensional analysis can fail, nothing is real, I am the walrus). More usefully, in a meaningful theory with a UV cutoff, $\Lambda_{UV}$, the dimensions get made up by the UV cutoff, which has $[\Lambda_{UV}] = 1$. Generically: $c_n = \tilde{c}_n \left(\Lambda_{UV}\right)^{-n[g]}$, where $\tilde{c}_n$ is dimensionless, and $n[g] < 0$ – it's higher and higher powers of the cutoff.

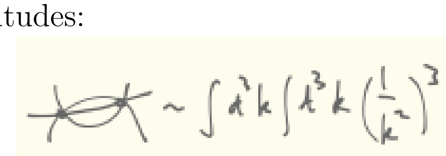
Consider the renormalizable (classically scale invariant) case: $[c_n] = [f]$, since $[g] = 0$. But in fact, what you'll get is something like

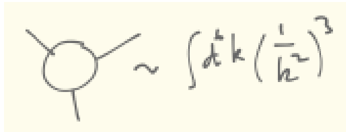
$$c_n = \tilde{c}_n \log^{\nu(n)}\left(\frac{\Lambda_{UV}}{\Lambda_{IR}}\right) ,$$

where $\Lambda_{IR}$ is an infrared cutoff or a mass, $[\Lambda_{IR}] = 1$.

Some classically scale invariant examples (so that $m = 0$ and the bare propagator is $1/k^2$) where you can see that we get logs from loop amplitudes:

$\phi^4$ in $D = 4$: 

$\phi^6$ in $D = 3$: 

$\phi^3$ in $D = 6$: 

In $D = 2$, even the propagator for a massless scalar field has logs:

$$\langle \phi(x)\phi(0)\rangle = \int đ^2k \frac{e^{-\mathbf{i}kx}}{k^2} \sim \log \frac{|x|}{\Lambda_{UV}}.$$

The terms involving 'renormalizable' in (7.23) are somewhat old-fashioned and come from a high-energy physics point of view where the short-distance physics is unknown, and we want to get as far as we can in that direction with our limited knowledge (in which case the condition 'renormalizability' lets us get away with this indefinitely – it lets us imagine we know everything). The latter terms are natural in the opposite situation (like condensed matter physics) where we know some basically correct microscopic description but want to know what happens at low energies. Then an operator like $\frac{1}{M^{24}}\phi^{28}$ whose coefficient is suppressed by some large mass scale $M$ is *irrelevant* for physics at energies far below that scale. Inversely, an operator like $m^2\phi^2$ gives a mass to the $\phi$ particles, and matters very much (is relevant) at energies $E < m$. In the marginal case, the quantum corrections have a chance to make a big difference.

## 7.6 Vertex correction in QED

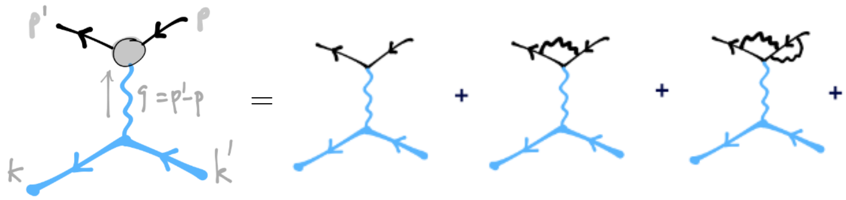
[Peskin chapter 6, Schwartz chapter 17, Zee chapter III.6] Back to work on QED. The vertex correction  has some great physics payoffs:

- We'll cancel the cutoff dependence we found in the $S$ matrix from $\delta Z$.
- We'll compute $g-2$ (the anomalous magnetic moment) of the electron, the locus of some of the most precise agreement between theory and experiment. (Actually the agreement is so good that it's used as the *definition* of the fine structure constant. But a similar calculation gives the leading anomalous magnetic moment of the muon.)
- We'll see that the exclusive differential cross section $\left(\frac{d\sigma}{d\Omega}\right)_{e\mu\leftarrow e\mu}$ that we've been considering is not really an observable. Actually it is infinity![8] The key word here is 'exclusive,' which means that we demand that the final state is *exactly* one electron and one muon and *absolutely nothing else*. Think for a moment about how you might do that measurement.

[8]More accurately, the exclusive cross section is zero; the one-loop correction is minus infinity, which is perturbation theory's clumsy attempt to correct the finite tree level answer to make it zero.

This is an example of an IR divergence. While UV divergences mean you're overstepping your bounds (by taking too seriously your Lagrangian parameters or your knowledge of short distances), IR divergences mean you are *asking the wrong question.*

To get started, consider the following class of diagrams.




$$\equiv \mathbf{i}\mathcal{M} \qquad = \mathbf{i}e^2 \left(\bar{u}(p')\Gamma^\mu(p,p')u(p)\right) \frac{1}{q^2}\bar{u}(K')\gamma_\mu u(K) \tag{7.24}$$

The shaded blob is the *vertex function* $\Gamma$. The role of the light blue factors is just to make and propagate the photon which hits our electron; let's forget about them. Denote the photon momentum by $q = p' - p$. We'll assume that the electron momenta $p, p'$ are on-shell, but $q^\mu$ is not, as in the $e\mu$ scattering process. Then $q^2 = 2m^2 - 2p' \cdot p$.

Before calculating the leading correction to the vertex $\Gamma^\mu = \gamma^\mu + \mathcal{O}(e^2)$, let's think about what the answer can be. It is a vector made from $p, p', \gamma^\mu$ and $m, e$ and numbers. It can't have any $\gamma^5$ or $\epsilon^{\mu\nu\rho\sigma}$ by parity symmetry of QED. So on general grounds we can organize it as

$$\Gamma^\mu(p,p') = A\gamma^\mu + B(p+p')^\mu + C(p-p')^\mu \tag{7.25}$$

where $A, B, C$ are Lorentz-invariant functions of $p^2 = (p')^2 = m^2$, $p \cdot p'$, $\not{p}, \not{p}'$. But, for example, $\not{p}\gamma^\mu u(p) = (m\gamma^\mu - p^\mu)u(p)$ which just mixes up the terms; really $A, B, C$ are just functions of the momentum transfer $q^2$. Gauge invariance, in the form of the Ward identity, says that contracting the photon line with the photon momentum should give zero:

$$0 \overset{\text{Ward}}{=} q_\mu \bar{u}(p')\Gamma^\mu u(p) \overset{(7.25)}{=} \bar{u}(p')\left( A \underbrace{\not{q}}_{=\not{p}'-\not{p} \overset{\bar{u}(p')\ldots u(p)}{=} m-m=0} + B \underbrace{(p+p')\cdot(p-p')}_{=m^2-m^2=0} + Cq^2 \right) u(p)$$

Therefore $0 = Cq^2\bar{u}(p')u(p)$ for general $q^2$ and general spinors, so $C = 0$. This is the moment for the Gordon identity to shine:

$$\bar{u}(p')\gamma^\mu u(p) = \bar{u}(p')\left(\frac{p^\mu + p'^\mu}{2m} + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\right)u(p)$$

(where $\sigma^{\mu\nu} \equiv \frac{\mathbf{i}}{2}[\gamma^\mu, \gamma^\nu]$) can be used to eliminate the $p+p'$ term[9]. The Gordon identity shows that the QED interaction vertex $\bar{u}(p')\gamma^\mu u(p)A_\mu$ contains a magnetic moment bit in addition to the $p+p'$ term (which is there for a charged scalar field).

It is then convenient (and conventional) to parametrize the vertex in terms of the two *form factors* $F_{1,2}$:

$$\Gamma^\mu(p,p') = \gamma^\mu F_1(q^2) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}F_2(q^2). \tag{7.26}$$

This little monstrosity has the complete information about the coupling of the electron to the electromagnetic field, such as for example a background electromagnetic field. It is a parametrization of the matrix elements of the current between two one-electron states, incorporating the fact of gauge invariance.

The first term at zero momentum $eF_1(q^2=0)$ is the electric charge of the electron (if you don't believe it, use the vertex (7.26) to calculate the Coulomb field of the electron; there are some details on page 186 of Peskin). Since the tree-level bit of $F_1$ is 1, if by the letter $e$ here we mean the actual charge, then we'd better include counterterms ($\mathcal{L}_{ct} \ni \bar{\psi}\delta_e\gamma^\mu A_\mu\psi$) to make sure it isn't corrected: $F_1(0)=1$.

On the homework last quarter you showed (or see Peskin p. 187) that the magnetic moment of the electron is

$$\vec{\mu} = g\frac{e}{2m}\vec{S},$$

where $\vec{S} \equiv \xi^\dagger\frac{\vec{\sigma}}{2}\xi$ is the electron spin. Comparing with the vertex function, this says that the $g$ factor is

$$g = 2(F_1(0)+F_2(0)) = 2+2F_2(0) = 2+\mathcal{O}(\alpha).$$

We see that the anomalous magnetic moment of the electron is $2F_2(q^2=0)$.

Now that we have some expectation about the form of the answer, and some ideas

[9] Actually this is why we didn't include a $\sigma^{\mu\nu}$ term. You could ask: what about a term like $\sigma^{\mu\nu}(p+p')^\nu$? Well, there's another Gordon identity that relates that to things we've already included:

$$\bar{u}_2\sigma_{\mu\nu}(p_1+p_2)^\nu u_1 = \mathbf{i}\bar{u}_2\left(q_\mu - (m_1-m_2)\gamma_\mu\right)u_1.$$

It is proved the same way: just use the Dirac equation $\not{p}_1 u_1 = m_1u_1, \bar{u}_2\not{p}_2 = \bar{u}_2m_2$ and the Clifford algebra. We are interested here in the case where $m_1=m_2$.

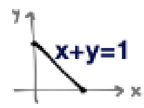
about what it's for, we sketch the evaluation of the one-loop QED vertex correction:


$$= -\mathbf{i}e^3 \int \text{đ}^4 k \; \bar{u}(p')\gamma^\nu \frac{\not{k}' + m_e}{(k')^2 - m_e^2} \gamma^\mu \frac{\not{k} + m_e}{k^2 - m_e^2} \gamma^\rho u(p) \cdot \frac{\eta_{\nu\rho}}{(p-k)^2 - m_\gamma^2}$$

with $k' \equiv k + q$.

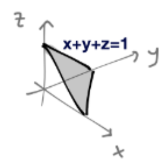
(1) Feynman parameters again. The one we showed before can be rewritten more symmetrically as:

$$\frac{1}{AB} = \int_0^1 dx \int_0^1 dy \; \delta(x + y - 1) \frac{1}{(xA + yB)^2}$$




Now how can you resist the generalization[10]:

$$\frac{1}{ABC} = \int_0^1 dx \int_0^1 dy \int_0^1 dz \; \delta(x + y + z - 1) \frac{2}{(xA + yB + zC)^3}$$




[10]Peskin outlines a proof by induction of the whole family of such identities on page 190. But here's a simpler proof using *Schwinger parameters*. You'll agree that

$$\frac{1}{A} = \int_0^\infty ds \; e^{-sA}. \tag{7.27}$$

Applying this identity to each factor gives

$$\frac{1}{A_1 A_2 \cdots A_n} = \int_0^\infty ds_1 \cdots \int_0^\infty ds_n \; e^{-\sum_{i=1}^n s_i A_i}.$$

Now use scaling to set $\tau \equiv \sum_{i=1}^n s_n$, and $x_i \equiv s_i/\tau$. Then

$$\frac{1}{A_1 A_2 \cdots A_n} = \int_0^\infty d\tau \tau^{n-1} \prod_{i=1}^n \int_0^1 dx_i \delta\left(\sum_{i=1}^n x_i - 1\right) \; e^{-\tau \sum_i x_i A_i}.$$

Now do the integral over $\tau$, using $\int_0^\infty d\tau \tau^{n-1} e^{-\tau X} = \frac{(n-1)!}{X^n}$ (differentiate (7.27) wrt $A$), to arrive at

$$\frac{1}{A_1 A_2 \cdots A_n} = \prod_{i=1}^n \int_0^1 dx_i \delta\left(\sum_{i=1}^n x_i - 1\right) \frac{(n-1)!}{\left(\sum_i x_i A_i\right)^n}.$$

So, set $A = (k')^2 - m_e^2, B = k^2 - m_e^2, C = (p-k)^2 - m_\gamma^2$ (with the appropriate $\mathbf{i}\epsilon$s), so that the integral we have to do is

$$\int \frac{đ^4 k N^\mu}{(k^2 + k \cdot (\cdots) + \cdots)^3}.$$

[End of Lecture 24]

(2) Complete the square, $\ell = k - zp + xq$ to get $\int \frac{đ^4 \ell N^\mu}{(\ell^2 - \Delta)^3}$ where

$$\Delta = -xyq^2 + (1-z)^2 m^2 + z m_\gamma^2.$$

The $\ell$-dependence in the numerator is either 1 or $\ell^\mu$ or $\ell^\mu \ell^\nu$. In the integral over $\ell$, the second averages to zero, and the third averages to $\eta^{\mu\nu}\ell^2 \frac{1}{4}$. As a result, the momentum integrals we need are just

$$\int \frac{đ^D \ell}{(\ell^2 - \Delta)^m} \text{ and } \int \frac{đ^D \ell\ \ell^2}{(\ell^2 - \Delta)^m}.$$

Right now we only need $D = 4$ and $m = 3$, but it turns out to be quite useful to think about them all at once. Like in our discussion of the electron self-energy diagram, we can evaluate them by Wick rotating (which changes the denominator to $\ell_E^2 + \Delta$) and going to polar coordinates. This gives:

$$\int \frac{đ^D \ell}{(\ell^2 - \Delta)^m} = (-1)^m \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2}\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2}}. \tag{7.28}$$

$$\int \frac{đ^D \ell\ \ell^2}{(\ell^2 - \Delta)^m} = (-1)^{m-1} \frac{D}{2} \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2} - 1\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2} - 1}. \tag{7.29}$$

Notice that these integrals are not equal to infinity when the parameter $D$ is not an integer. This is the idea behind *dimensional regularization*.

(0) But for now let's persist in using the Pauli Villars regulator. (I call this step (0) instead of (3) because it should have been there all along.) Here this means we subtract from the amplitude the same quantity with $m_\gamma$ replaced by $\Lambda^2$. The dangerous bit comes from the $\ell^2$ term we just mentioned, since $m - D/2 - 1 = 3 - 4/2 - 1 = 0$ means logs.

The numerator is

$$\begin{aligned} N^\mu &= \bar{u}(p')\gamma^\nu \left(\not{k} + \not{q} + m_e\right) \gamma^\mu \left(\not{k} + m_e\right) \gamma_\nu u(p) \\ &= -2\left(\mathcal{A}\bar{u}(p')\gamma^\mu u(p) + \mathcal{B}\bar{u}(p')\sigma^{\mu\nu} q_\nu u(p) + \mathcal{C}\bar{u}(p') q^\mu u(p)\right) \end{aligned} \tag{7.30}$$

where

$$\mathcal{A} = -\frac{1}{2}\ell^2 + (1-x)(1-y)q^2 + (1 - 4z + z^2)m^2$$

$$
\begin{aligned}
\mathcal{B} &= \mathbf{i} m z(1-z) \\
\mathcal{C} &= m(z-2)(y-x) \ . \qquad (7.31)
\end{aligned}
$$

The blood of many men was spilled to arrive at these simple expressions (actually most of the algebra is done explicitly on page 319 of Schwartz). Now you say: but you promised there would be no term like $\mathcal{C}$ because of the Ward identity. Indeed I did and indeed there isn't because $\mathcal{C}$ is odd in $x \leftrightarrow y$ while everything else is even, so this term integrates to zero.

The first term (with $\mathcal{A}$) is a correction to the charge of the electron and will be UV divergent. More explicitly, we get, using Pauli-Villars,

$$\int đ^4\ell \left( \frac{\ell^2}{\left(\ell^2 - \Delta_{m_\gamma}\right)^3} - \frac{\ell^2}{\left(\ell^2 - \Delta_\Lambda\right)^3} \right) = \frac{\mathbf{i}}{(4\pi)^2} \ln \frac{\Delta_\Lambda}{\Delta_{m_\gamma}}.$$

The other bits are finite, and we ignore the terms that go like negative powers of $\Lambda$. More on this cutoff dependence soon. But first something wonderful:

### 7.6.1 Anomalous magnetic moment

The second term $\mathcal{B}$ contains the anomalous magnetic moment:

$$
\begin{aligned}
F_2(q^2) &= \frac{2m}{e} \cdot (\text{the term with } \mathcal{B}\ ) \\
&= \frac{2m}{e} 4e^3 m \int dx dy dz \delta(x+y+z-1) z(1-z) \underbrace{\int \frac{đ^4\ell}{\left(\ell^2 - \Delta\right)^3}}_{=\frac{-\mathbf{i}}{32\pi^2\Delta}} \\
&= \frac{\alpha}{\pi} m^2 \int dx dy dz \delta(x+y+z-1) \frac{z(1-z)}{(1-z)^2 m^2 - xyq^2}. \qquad (7.32)
\end{aligned}
$$

The magnetic moment is the long-wavelength bit of this:

$$F_2(q^2 = 0) = \frac{\alpha}{\pi} m^2 \int_0^1 dz \int_0^{1-z} dy \frac{z}{(1-z)m^2} = \frac{\alpha}{2\pi}.$$

$$\boxed{g = 2 + \frac{\alpha}{\pi} + \mathcal{O}(\alpha^2).}$$

A rare opportunity for me to plug in numbers: $g = 2.00232$.

### 7.6.2 IR divergences mean wrong questions.

There is a term in the numerator from the $\mathcal{A}\gamma^\mu$ bit

$$\int \frac{đ^4\ell}{(\ell^2 - \Delta)^3} = c \frac{1}{\Delta}$$

(with $c = -\frac{\mathbf{i}}{32\pi^2}$ again), but without the factor of $z(1-z)$ we had in the magnetic moment calculation. It looks like we've gotten away without having to introduce a UV regulator here, too (so far). But now look at what happens when we try to do the Feynman parameter integrals. For example, at $q^2 = 0$, we get (if we had set $m_\gamma = 0$)

$$\int dxdydz\delta(x+y+z-1)\frac{m^2(1-4z+z^2)}{\Delta} = m^2 \int^1 dz \int_0^{1-z} dy \frac{-2+2(1-z)+(1-z)^2}{(1-z)^2 m^2}$$
$$= \int^1 dz \frac{-2}{(1-z)} + \text{finite}, \tag{7.33}$$

which diverges at the upper limit of integration. In fact it's divergent even when $q^2 \neq 0$. This is a place where we actually need to include the photon mass, $m_\gamma$, for our own safety.

The (IR singular bit of the) vertex (to $\mathcal{O}(\alpha)$) is of the form

$$\Gamma^\mu = \gamma^\mu \left(1 - \frac{\alpha}{2\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{m_\gamma{}^2}\right)\right) + \text{ stuff which is finite as } m_\gamma \to 0. \tag{7.34}$$

Notice that the IR divergent stuff depends on the electron momenta $p, p'$ only through $q$, the momentum of the photon. So it looks like we are led to conclude

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mu e \leftarrow \mu e} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \left(1 - \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{m_\gamma{}^2}\right)\right) + \mathcal{O}\left(\alpha^2\right)$$

which blows up when we remove the fake photon mass $m_\gamma \to 0$.

[Schwartz §20.1] I wanted to just quote the above result for (7.34) but I lost my nerve, so here is a bit more detail leading to it. The IR dangerous bit comes from the second term in $\mathcal{A}$ above. That is,

$$F_1(q^2) = 1 + f(q^2) + \delta_1 + \mathcal{O}(\alpha^2)$$

with

$$f(q^2) = \frac{e^2}{8\pi^2} \int_0^1 dxdydz\delta(x+y+z-1) \left(\ln \frac{z\Lambda^2}{\Delta} + \frac{q^2(1-x)(1-y) + m_e^2(1-4z+z^2)}{\Delta}\right).$$

$\delta_1$ here is a counterterm for the $\Psi\gamma^\mu A_\mu \Psi$ vertex.

We can be more explicit if we consider $-q^2 \gg m_e^2$ so that we can ignore the electron mass everywhere. Then we would choose the counterterm $\delta_1$ so that

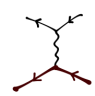
$$1 = F_1(0) \implies \delta_1 = -f(0) \overset{m_e/q \to 0}{\to} -\frac{e^2}{8\pi^2}\frac{1}{2}\ln\frac{\Lambda^2}{m_\gamma^2}.$$

And the form of $f(q^2)$ is

$$f(q^2)|_{m_e=0} = \frac{e^2}{8\pi^2}\int dxdydz\delta(x+y+z-1)\left(\underbrace{\ln\frac{(1-x-y)\Lambda^2}{\Delta}}_{\text{IR finite}} + \frac{q^2(1-x)(1-y)}{-xyq^2+(1-x-y)m_\gamma^2}\right)$$

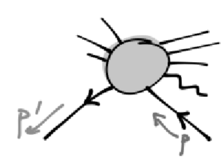
$$F_1(q^2)|_{m_e=0} = 1 - \frac{e^2}{16\pi^2}\left(\ln^2\frac{-q^2}{m_\gamma^2} + 3\ln\frac{-q^2}{m_\gamma^2}\right) + \text{finite}.$$

In doing the integrals, we had to remember the $\mathbf{i}\epsilon$ in the propagators, which can be reproduced by the replacement $q^2 \to q^2 + \mathbf{i}\epsilon$. This $\ln^2(q^2/m_\gamma)$ is called a *Sudakov double logarithm.* Notice that taking differences of these at different $q^2$ will not make it finite.

**Diversity and inclusion to the rescue.** Before you throw up your hands in despair, I would like to bring to your attention another consequence of the masslessness of the photon: It means real (as opposed to virtual) photons can be made with arbitrarily low energy. But a detector has a minimum triggering energy: the detector works by particles doing some physical something to stuff in the detector, and it has a finite energy resolution – it takes a finite amount of energy for those particles to do stuff. This means that a process with exactly one $e$ and one $\mu$ in the final state  cannot be distinguished from a process ending in $e\mu$ plus a photon of arbitrarily small energy, such as would result from  (final-state radiation) or  (initial-state radiation). This ambiguity is present for *any* process with external charged particles.

Being more inclusive, then, we cannot distinguish amplitudes of the form

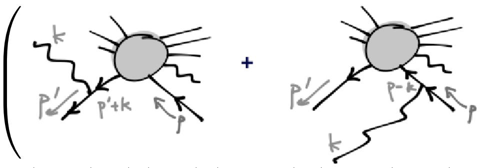
$$\bar{u}(p')\mathcal{M}_0(p',p)u(p) \equiv -\mathbf{i}$$




,

from more inclusive amplitudes like

$$-\mathbf{i}$$




$$= \bar{u}(p')\gamma^\mu \frac{e}{\not{p}'+\not{k}-m_e}\mathcal{M}_0(p',p)u(p)\epsilon^\star_\mu(k) + \bar{u}(p')\mathcal{M}_0(p',p)\frac{e}{\not{p}-\not{k}-m_e}\gamma^\mu u(p)\epsilon^\star_\mu(k)\ .$$

Now, by assumption the photon is real ($k^2$=0) and it is *soft*, in the sense that $k^0 < E_c$, the detector cutoff. So we can approximate the numerator of the second term as

$$\left(\not{p} - \not{k} + m_e\right)\gamma^\mu u(p) \simeq \left(\not{p} + m_e\right)\gamma^\mu u(p) = (2p^\mu + \gamma^\mu \underbrace{\left(-\not{p} + m_e\right))u(p)}_{=0} = 2p^\mu u(p).$$

In the denominator we have *e.g.* $(p-k)^2 - m_e^2 = p^2 - m_e^2 - 2p\cdot k + k^2 \sim -2p\cdot k$ since the electron is on shell and $k \ll p$. Therefore

$$\mathcal{M}\left(e\mu + \text{one soft } \gamma \leftarrow e\mu\right) = e\bar{u}(p')\mathcal{M}_0(p',p)u(p)\left(\frac{p'\cdot\epsilon^\star}{p'\cdot k + \mathbf{i}\epsilon} - \frac{p\cdot\epsilon^\star}{p\cdot k - \mathbf{i}\epsilon}\right) \quad (7.35)$$

---

**This is bremsstrahlung.** Before we continue this calculation to find the inclusive amplitude which a real detector actually measures, let's pause to relate the previous expression to some physics we know. Where have we seen this kind of expression

$$\frac{p'^\mu}{p'\cdot k + \mathbf{i}\epsilon} - \frac{p^\mu}{p\cdot k - \mathbf{i}\epsilon} \equiv \frac{1}{\mathbf{i}e}\tilde{j}^\mu(k)$$

before? Notice that the $\mathbf{i}\epsilon$ are different because one comes from final state and one from initial. Well, this object is the Fourier transform $\tilde{j}^\mu(k) = \int d^4x\, e^{+\mathbf{i}kx} j^\mu(x)$ of the current

$$j^\mu(x) = e\int d\tau \frac{dy^\mu}{d\tau}\delta^{(4)}(x - y(\tau))$$

associated with a particle which executes a piecewise linear motion [11]

$$y(\tau) = \begin{cases} \frac{p^\mu}{m}\tau, & \tau < 0 \\ \frac{p'^\mu}{m}\tau, & \tau > 0 \end{cases}.$$

This is a good approximation to the motion a free particle which experiences a sudden acceleration; sudden means that the duration of the pulse is short compared to $\omega^{-1}$ for any frequency we're going to measure. The electromagnetic radiation that such an accelerating charge produces is given classically by Maxwell's equation: $\tilde{A}^\mu(k) = -\frac{1}{k^2}\tilde{j}^\mu(k)$.

I claim further that the factor $f_{IR}(q^2) = \frac{\alpha}{\pi}\ln\left(\frac{-q^2}{m^2}\right)$ (which entered our lives in (7.34)) arises classically as the *number* of soft photons produced by such a process in

---

[11] Check it:

$$\int d^4x j^\mu(x)e^{+\mathbf{i}kx} = e\int d\tau\frac{dy^\mu(\tau)}{d\tau}e^{\mathbf{i}k\cdot y(\tau)} = e\int_0^\infty d\tau\frac{p'^\mu}{m}e^{\mathbf{i}\left(\frac{k\cdot p'}{m} + \mathbf{i}\epsilon\right)\tau} + e\int_{-\infty}^0 d\tau\frac{p^\mu}{m}e^{\mathbf{i}\left(\frac{k\cdot p}{m} - \mathbf{i}\epsilon\right)\tau} = \tilde{j}^\mu(k).$$

Notice that the $\mathbf{i}\epsilon$ are convergence factors in the Fourier transforms.

each decade of wavenumber. You can figure this out by plugging $\tilde{A}^\mu(k) = -\frac{1}{k^2}\tilde{j}^\mu(k)$ into the electromagnetic energy $\frac{1}{2}\int d^3x\,(E^2+B^2)$. See Peskin §6.1 for help.

---

$$\left(\frac{d\sigma}{d\Omega}\right)^{E_\gamma<E_c}_{\mu e\gamma_{\text{soft}}\leftarrow\mu e} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} e^2 \underbrace{\int_0^{E_c} \frac{đ^3k}{2E_k}}_{\gamma\text{ phase space}} \left|\frac{2p\cdot\epsilon^\star}{2p\cdot k} - \frac{2p'\cdot\epsilon^\star}{2p'\cdot k}\right|^2 \overset{E_k=|\vec{k}|}{\sim} \int_0 \frac{d^3k}{k^3} = \infty.$$

This is *another* IR divergence. (One divergence is bad news, but two is an opportunity for hope.) Just like we must stick to our UV regulators like religious zealots, we must cleave tightly to the consistency of our IR regulators: we need to put back the photon mass:

$$E_k = \sqrt{\vec{k}^2 + m_\gamma{}^2}$$

which means that the lower limit of the $k$ integral gets cut off at $m_\gamma$:

$$\int_0^{E_c}\frac{dk}{E_k} = \left(\int_0^{m_\gamma} + \int_{m_\gamma}^{E_c}\right)\frac{dk}{\sqrt{k^2+m_\gamma{}^2}} \sim \underbrace{\int_0^{m_\gamma}\frac{dk}{m_\gamma}}_{=1} + \underbrace{\int_{m_\gamma}^{E_c}\frac{dk}{k}}_{\ln\frac{E_c}{m_\gamma}}.$$

Being careful about the factors, the actual cross section measured by a detector with energy resolution $E_c$ is[12]

$$\begin{aligned}
\left(\frac{d\sigma}{d\Omega}\right)^{\text{observed}} &= \left(\frac{d\sigma}{d\Omega}\right)_{e\mu\leftarrow\mu e} + \left(\frac{d\sigma}{d\Omega}\right)^{E_\gamma<E_c}_{\mu e\gamma_{\text{soft}}\leftarrow\mu e} + \mathcal{O}(\alpha^3) \\
&= \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}}\left(\underbrace{1 - \frac{\alpha}{\pi}f_{IR}(q^2)\ln\left(\frac{-q^2}{m_\gamma{}^2}\right)}_{\text{vertex correction}} + \underbrace{\frac{\alpha}{\pi}f_{IR}(q^2)\ln\left(\frac{E_c^2}{m_\gamma{}^2}\right)}_{\text{soft photons}}\right) \\
&= \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}}\left(1 - \frac{\alpha}{\pi}f_{IR}(q^2)\ln\left(\frac{-q^2}{E_c^2}\right)\right)
\end{aligned}$$

The thing we can actually measure is independent of the IR regulator photon mass $m_\gamma$, and finite when we remove it. On the other hand, it depends on the detector resolution. Like in the plot of some kind of Disney movie, an apparently minor character whom you may have been tempted to regard as an ugly detail has saved the day.

---

[12] Notice that we add the cross-sections, not the amplitudes, for these processes with different final states. Here's why: even though we don't measure the existence of the photon, *something* does: it gets absorbed by some part of the apparatus or the rest of the world and therefore becomes entangled some of its degrees of freedom; when we fail to distinguish between those states, we trace over them, and this erases the interference terms we would get if we summed the amplitudes.

I didn't show explicitly that the coefficient of the log is the same function $f_{IR}(q^2)$. In fact this function is $f_{IR}(q^2) = \frac{1}{2}\log(-q^2/m^2)$, so the product $f_{IR}\ln q^2 \sim \ln^2 q^2$ is the Sudakov double logarithm. A benefit of the calculation which shows that the same $f_{IR}$ appears in both places (Peskin chapter 6.5) is that it also shows that this pattern persists at higher order in $\alpha$: there is a $\ln^2(q^2/{m_\gamma}^2)$ dependence in the two-loop vertex correction, and a matching $-\ln^2(E_c^2/{m_\gamma}^2)$ term in the amplitude to emit *two* soft photons. There is a $\frac{1}{2!}$ from Bose statistics of these photons. The result exponentiates, and we get

$$e^{-\frac{\alpha}{\pi}f\ln(-q^2/{m_\gamma}^2)}e^{-\frac{\alpha}{\pi}f(E_c^2/{m_\gamma}^2)} = e^{-\frac{\alpha}{\pi}f\ln(-q^2/E_c^2)}.$$

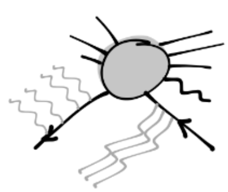
You may be bothered that I've made all this discussion about the corrections from the electron line, but said nothing about the muon line. But the theory should make sense even if the electron and muon charges $Q_e, Q_m$ were different, so the calculation should make sense term-by-term in an expansion in $Q_m$.

Some relevant names for future reference: The name for the guarantee that this always works in QED is the *Bloch-Nordsieck theorem*. Closely-related but more serious issues arise in QCD, the theory of quarks and gluons; this is the beginning of the story of *jets* (a jet is some IR-cutoff dependent notion of a QCD-charged particle plus the cloud of stuff it carries with it) and *parton distribution functions*.

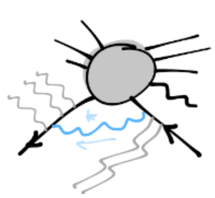
[End of Lecture 25]

---

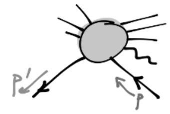
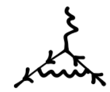
**Sketch of exponentiation of soft photons.** [Peskin §6.5] Consider a diagram with $n$ soft external photons, summed over ways of distributing them on an initial and final electron line:

$$\sum_{n_f=1}^{n}$$

$$= \bar{u}(p')\mathbf{i}\mathcal{M}_0 u(p) e^n \prod_{\alpha=1}^{n}\left(\frac{p'^{\mu_\alpha}}{p'\cdot k_\alpha} - \frac{p^{\mu_\alpha}}{p\cdot k_\alpha}\right) \equiv \mathcal{A}_n.$$

Here the difference in each factor is just as in (7.35), one term from initial and one from final-state emission; expanding the product gives the sum over $n_f = 1 - n_i$, the number coming from the final-state line. From this expression, we can make a diagram with a soft-photon *loop* by picking an initial line $\alpha$ and a final line $\beta$ setting $k_\alpha = -k_\beta \equiv k$ and tying them together with a propagator and summing over $k$:


$$= \mathcal{A}_{n-2}\frac{e^2}{2}\int đ^4k \frac{-\mathbf{i}\eta_{\rho\sigma}}{k^2}\left(\frac{p'}{p'\cdot k} - \frac{p}{p\cdot k}\right)^\rho \left(\frac{p'}{-p'\cdot k} - \frac{p}{-p\cdot k}\right)^\sigma$$

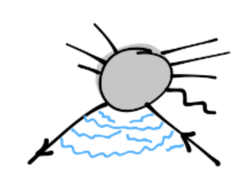
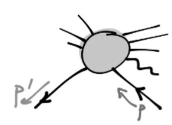
The factor of $\frac{1}{2}$ accounts for the symmetry under exchange of $\alpha \leftrightarrow \beta$. For the case of $n = 2$, this is the whole story, and this is

$\bar{u}\mathbf{i}\mathcal{M}_0 u \cdot \mathbf{X} = \Big($   $\Big) \cdot \Big($  $\Big)_{\text{soft part}}$

(where here 'soft part' means the part which is singular in $m_\gamma$) from which we conclude that

$$\mathbf{X} = -\frac{\alpha}{2\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{m_\gamma^2}\right) + \text{finite}.$$

Taking the most IR-divergent bit with $m$ virtual soft photons (order $\alpha^m$) for each $m$ gives

$\mathcal{M}_{\text{virtual soft}} = \sum_{m=0}^{\infty} \Big($  $\Big) = \underbrace{\Big(}_{}$   $\underbrace{\Big)}_{\mathbf{i}\mathcal{M}_0} \underbrace{\sum_m \frac{1}{m!}\mathbf{X}^m}_{e^{\mathbf{X}}}$

where the $1/m!$ is a symmetry factor from interchanging the virtual soft photons.

Now consider the case of one real external soft ($E \in [m_\gamma, E_c]$) photon in the final state. The cross section is

$$\begin{aligned}
d\sigma_{1\gamma} &= \int d\Pi \underbrace{\sum_{\text{pols}} \epsilon^\mu \epsilon^{\star\nu}}_{=-\eta^{\mu\nu}} \mathcal{M}_\mu \mathcal{M}_\nu^\star \\
&= |\bar{u}(p')\mathcal{M}_0 u(p)|^2 \int \frac{đ^3 k}{2E_k} \left(-\eta_{\mu\nu}\right) e^2 \left(\frac{p'}{p'\cdot k} - \frac{p}{p\cdot k}\right)^\mu \left(\frac{p'}{-p'\cdot k} - \frac{p}{-p\cdot k}\right)^\nu \\
&\equiv d\sigma_0 \mathbf{Y}, \\
\mathbf{Y} &= \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{E_c^2}{m_\gamma^2}\right).
\end{aligned}$$

(The integral is done in Peskin, page 201.) Therefore, the exclusive cross section, including contributions of soft real photons gives

$$\sum_{n=0}^{\infty} d\sigma_{n\gamma} = d\sigma_0 \sum_n \frac{1}{n!}\mathbf{Y}^n = d\sigma_0 e^{\mathbf{Y}}.$$

Here the $n!$ is because the final state contains $n$ identical bosons.

Putting the two effects together gives the promised cancellation of $m_\gamma$ dependence to all orders in $\alpha$:

$$d\sigma = d\sigma_0 e^{2\mathbf{X}} e^{\mathbf{Y}}$$

$$
\begin{aligned}
&= d\sigma_0 \exp\left(-\frac{\alpha}{\pi} f_{IR}(q^2) \ln\frac{-q^2}{m_\gamma^2} + \frac{\alpha}{\pi} f_{IR}(q^2)\ln\frac{E_c^2}{m_\gamma^2}\right) \\
&= d\sigma_0 \exp\left(-\frac{\alpha}{\pi} f_{IR}(q^2)\ln\frac{-q^2}{E_c^2}\right)
\end{aligned}
$$

This might seem pretty fancy, but unpacking the sum we did, the basic statement is that the probability of finding $n$ photons with energy in a given (low-energy) range $[E_-, E_+]$ is

$$
P_{[E_-,E_+]} = \frac{1}{n!}\lambda^n e^{-\lambda}, \qquad \lambda = \frac{\alpha}{\pi} f_{IR}(q^2)\ln\frac{E_+}{E_-} = \langle n\rangle = \left\langle n^2\right\rangle - \langle n\rangle^2
$$

a Poisson distribution. This is just what one finds in a coherent state of the radiation field.

---

### 7.6.3 Some magic from gauge invariance of QED

We found that the self-energy of the electron gave a wavefunction renormalization factor

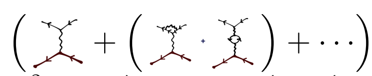
$$
Z_2 = 1 + \frac{\partial\Sigma}{\partial\not{p}}|_{\not{p}=m_0} + \mathcal{O}(e^4) = 1 - \frac{\alpha}{4\pi}\ln\frac{\Lambda^2}{m^2} + \text{finite} + \mathcal{O}(\alpha^2).
$$

We care about this because there is a factor of $Z_2$ in the LSZ formula for an $S$-matrix element with two external electrons. On the other hand, we found a cutoff-dependent correction to the vertex $e\gamma^\mu F_1(q^2)$ of the form

$$
F_1(q^2) = 1 + \frac{\alpha}{4\pi}\ln\frac{\Lambda^2}{m^2} + \text{finite} + \mathcal{O}(\alpha^2).
$$

Combining these together

$$
S_{e\mu\leftarrow e\mu} = \left(\sqrt{Z_2(e)}\right)^2
$$



$$
= \left(1 - \frac{\alpha}{4\pi}\ln\frac{\Lambda^2}{m^2} + \cdots\right) e^2\bar{u}(p')\left(\gamma^\mu\left(1 + \frac{\alpha}{4\pi}\ln\frac{\Lambda^2}{m^2} + \cdots\right) + \alpha\frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\right)u(p)
$$

the UV divergence from the vertex cancels the one in the self-energy. Why did this have to happen? During our discussion of the IR divergences, I mentioned a counterterm $\delta_1$ for the vertex. But how many counterterms do we get here? Is there a point of view which makes this cancellation obvious? Notice that the $\cdots$ multiplying the $\gamma^\mu$ term still contain the vacuum polarization diagram, which is our next subject, and which may be (is) cutoff dependent. Read on.

## 7.7 Vacuum polarization

[Zee, III.7] We've been writing the QED lagrangian as

$$\mathcal{L} = \bar\psi\left(\partial\!\!\!/ + \mathbf{i}e\tilde{A}\!\!\!/ - m\right)\psi - \frac{1}{4}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}.$$

I've put tildes on the photon field because of what's about to happen: Suppose we rescale the definition of the photon field $e\tilde{A}_\mu \equiv A_\mu, e\tilde{F}_{\mu\nu} \equiv F_{\mu\nu}$. Then the coupling $e$ moves to the photon kinetic term:

$$\mathcal{L} = \bar\psi\left(\partial\!\!\!/ + \mathbf{i}A\!\!\!/ - m\right)\psi - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}.$$

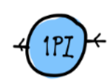
With this normalization, instead of measuring the coupling between electrons and photons, the coupling constant $e$ measures the difficulty a photon has propagating through space:

$$\langle A_\mu A_\nu\rangle \sim \frac{-\mathbf{i}\eta_{\mu\nu}e^2}{q^2}.$$

None of the physics is different, since each internal photon line still has two ends on a $\bar\psi A\!\!\!/\psi$ vertex.

But from this point of view it is clear that the magic of the previous subsection is a consequence of gauge invariance, here's why: the demand of gauge invariance relates the coefficients of the $\bar\psi\partial\!\!\!/\psi$ and $\bar\psi A\!\!\!/\psi$ terms[13]. Therefore, any counterterm we need for the $\bar\psi\partial\!\!\!/\psi$ term (which comes from the electron self-energy correction and is traditionally called $\delta Z_2$) must be the *same* as the counterterm for the $\bar\psi A\!\!\!/\psi$ term (which comes from the vertex correction and is called $\delta Z_1$). No magic, just gauge invariance.

A further virtue of this reshuffling of the factors of $e$ (emphasized by Zee on page 205) arises when we couple more than one species of charged particle to the electromagnetic field, *e.g.* electrons and muons or, more numerously, protons: once we recognize that charge renormalization is a property of the photon itself, it makes clear that quantum corrections cannot mess with the ratio of the charges. A deviation from $-1$ of the ratio of the charges of electron and proton as a result of interactions might seem plausible given what a mess the proton is, and would be a big deal for atoms. Gauge invariance forbids it.

Just as we defined the electron self-energy (amputated 2-point function) as 

$=$ $-\mathbf{i}\Sigma(p\!\!\!/)$ (with two spinor indices implied), we define the photon self-energy as

$$+\mathbf{i}\Pi_{\mu\nu}(q^2) \equiv$$ 

$=$  $+\ \mathcal{O}(e^4)$

[13] Notice that the gauge transformation of the rescaled $A_\mu$ is $A_\mu \to A_\mu + \partial_\mu\lambda(x), \psi(x) \to e^{\mathbf{i}q\lambda(x)}\psi(x)$ so that $D_\mu\psi \equiv (\partial + q\mathbf{i}A)_\mu\,\psi \to e^{\mathbf{i}q\lambda}D_\mu\psi$ where $q$ is the charge of the field ($q = -1$ for the electron). This is to be contrasted with the transformation of $\tilde{A}_\mu \to \tilde{A}_\mu - \partial_\mu\lambda(x)/e$.

(the diagrams on the RHS are amputated). It is a function of $q^2$ by Lorentz symmetry. (The reason for the difference in sign is that the electron propagator is $\frac{+\mathbf{i}}{\not p - m}$ while the photon propagator is $\frac{-\mathbf{i}\eta_{\mu\nu}}{q^2}$.) We can parametrize the answer as

$$\Pi^{\mu\nu}(q^2) = A(q^2)\eta^{\mu\nu} + B(q^2)q^\mu q^\nu .$$

The Ward identity says

$$0 = q_\mu \Pi^{\mu\nu}(q^2) \implies 0 = Aq^\nu + Bq^2 q^\nu \implies B = -A/q^2 .$$

Let $A \equiv \Pi q^2$ so that

$$\Pi^{\mu\nu}(q^2) = \Pi(q^2)q^2 \underbrace{\left(\eta^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)}_{=\Delta_T^{\mu\nu}} .$$

This object $\Delta_T^{\mu\nu}$ is a projector

$$\Delta^\mu_{T\rho}\Delta_T{}^\rho_\nu = \Delta^\mu_{T\nu} \tag{7.36}$$

onto modes transverse to $q^\mu$. Recall that we can take the bare propagator to be

 $= \dfrac{-\mathbf{i}\Delta_T}{q^2}$

without changing any gauge-invariant physics. This is useful because then

$\tilde G^{(2)}(q) =$  $+\cdots$

$$\overset{(7.36)}{=} \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right)\mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \cdots\right)$$

$$\overset{\Delta_T^2=\Delta_T}{=} \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \Pi\Delta_T + \Pi^2\Delta_T + \cdots\right) = \frac{-\mathbf{i}\Delta_T}{q^2}\frac{1}{1-\Pi(q^2)} . \tag{7.37}$$

Does the photon get a mass? If the thing I called $A$ above $q^2\Pi(q^2) \overset{q^2\to 0}{\to} A_0 \neq 0$ (that is, if $\Pi(q^2) \sim \frac{A_0}{q^2}$ or worse), then $\tilde G \overset{q^2\to 0}{\sim} \frac{1}{q^2 - A_0}$ does not have a pole at $q^2 = 0$. If $\Pi(q^2)$ is regular at $q^2 = 0$, then the photon remains massless. In order to get such a singularity in the photon self energy $\Pi(q^2) \sim \frac{A_0}{q^2}$ we need a process like $\delta\Pi \sim$ , where the intermediate state is a massless boson with propagator $\sim \frac{A_0}{q^2}$. As I will explain below, this is the Higgs mechanism (not the easiest way to understand it).

The Ward identity played an important role here. Why does it work for the vacuum polarization?

$q_\mu \Pi_2^{\mu\nu}(q^2) = q_\mu$  $\propto e^2 \int đ^4p \,\mathrm{tr} \frac{1}{\not p + \not q - m}\not q \frac{1}{\not p - m}\gamma^\nu .$

But here is an identity:

$$\frac{1}{\not p + \not q - m} \not q \frac{1}{\not p - m} = \frac{1}{\not p - m} - \frac{1}{\not p + \not q - m}. \tag{7.38}$$

Now, if we shift the integration variable $p \to p + q$ in the second term, the two terms cancel.

Why do I say 'if'? If the integral depends on the UV limit, this shift is not innocuous. So we have to address the cutoff dependence.

In addition to the (lack of) mass renormalization, we've figured out that the electromagnetic field strength renormalization is

$$Z_\gamma \equiv Z_3 = \frac{1}{1 - \Pi(0)} \sim 1 + \Pi(0) + \mathcal{O}(e^4).$$

We need $Z_\gamma$ for example for the $S$-matrix for processes with external photons, like Compton scattering.

Claim: If we do it right[14], the cutoff dependence looks like[15]:

$$\Pi_2(q^2) = \frac{\alpha_0}{4\pi} \left( -\frac{2}{3} \ln \Lambda^2 + \underbrace{2D(q^2)}_{\text{finite}} \right)$$

where $\Lambda$ is the UV scale of ignorance. The photon propagator gets corrected to

$$\frac{e_0^2 \Delta_T}{q^2} \rightsquigarrow \frac{Z_3 e_0^2 \Delta_T}{q^2},$$

and $Z_3 = \frac{1}{1-\Pi(0)}$ blows up logarithmically if we try to remove the cutoff. You see that the fine structure constant $\alpha_0 = \frac{e_0^2}{4\pi}$ has acquired the subscript of deprecation: we can make the photon propagator sensible while removing the cutoff if we are willing to recognize that the letter $e_0$ we've been carrying around is a fiction, and write everything

[14]What I mean here is: if we do it in a way which respects the gauge invariance and hence the Ward identity. The simple PV regulator we've been using does not quite do that. However, an only slightly more involved implementation, explained in Zee page 202-204, does. Alternatively, we could use dimensional regularization everywhere.

[15]The factor in front of the $\ln \Lambda$ can be made to look like it does in other textbooks using $\alpha = \frac{e^2}{4\pi}$, so that

$$\frac{\alpha_0}{4\pi} \left( \frac{2}{3} \ln \Lambda^2 \right) = \frac{e_0^2}{12\pi^2} \ln \Lambda.$$

in terms of $e \equiv \sqrt{Z_3} e_0$ where $\frac{e^2}{4\pi} = \frac{1}{137}$ is the measured fine structure constant. To this order, then, we write

$$e_0^2 = e^2 \left(1 + \frac{\alpha_0}{4\pi}\frac{2}{3}\ln\Lambda^2\right) + \mathcal{O}(\alpha^2). \tag{7.39}$$

$$m_0 = m + \mathcal{O}(\alpha_0) = m + \mathcal{O}(\alpha). \tag{7.40}$$

Since the difference between $\alpha_0$ and $\alpha$ is higher order (in either), our book-keeping is unchanged. Inverting the relationship perturbatively, the renormalized charge is

$$e^2 = e_0^2 \left(1 - \frac{\alpha_0}{4\pi}\frac{2}{3}\ln\Lambda^2 + \mathcal{O}(\alpha^2)\right)$$

– in QED, the quantum fluctuations *reduce* the charge, as you might expect from the interpretation of this phenomenon as dielectric screening.

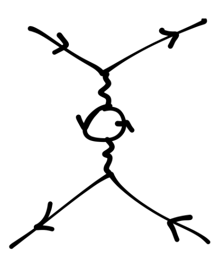
[End of Lecture 26]

In the example case of $e\mu \leftarrow e\mu$ scattering, the UV cutoff dependence looks like

$$\begin{aligned} S_{e\mu\leftarrow e\mu} = &\sqrt{Z_e^2}\left(1 - \frac{\alpha_0}{4\pi}\ln\Lambda^2 + \frac{\alpha_0}{2\pi}A(m_0)\right)e_0^2 \\ &L_\mu \bar{u}(p')\left[\gamma^\mu\left(1 + \frac{\alpha_0}{4\pi}\ln\Lambda^2 + \frac{\alpha_0}{2\pi}(B+D) + \frac{\alpha_0}{4\pi}\left(-\frac{2}{3}\ln\Lambda^2\right)\right) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\frac{\alpha_0}{2\pi}C(q^2, m_0)\right]u(p) \\ = &e^2 L_\mu \bar{u}(p')\left[\gamma^\mu\left(1 + \frac{\alpha}{2\pi}(A+B+D)\right) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\frac{\alpha}{2\pi}C\right]u(p) + \mathcal{O}(\alpha^2) \end{aligned} \tag{7.41}$$

where $L_\mu$ is the stuff from the muon line, and $A, B, C, D$ are finite functions of $m, q^2$. In the second step, two things happened: (1) we cancelled the UV divergences from the $Z$-factor and from the vertex correction: this had to happen because there was no possible counterterm. (2) we used (7.39) and (7.40) to write everything in terms of the measured $e, m$.

Claim: this works for all processes to order $\alpha^2$. For example, Bhabha scattering gets a contribution of the form

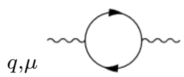


$$\propto e_0 \frac{1}{1 - \Pi(0)} e_0 = e^2.$$

In order to say what are $A, B, D$ we need to specify more carefully a renormalization scheme. To do that, I need to give a bit more detail about the integral.

### 7.7.1 Under the hood

The vacuum-polarization contribution of a fermion of mass $m$ and charge $e$ at one loop is



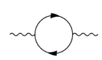
$$= -\int đ^D k \mathrm{tr}\left( (\mathbf{i}e\gamma^\mu) \frac{\mathbf{i}\left(\not{k} + m\right)}{k^2 - m^2} (\mathbf{i}e\gamma^\nu) \frac{\mathbf{i}\left(\not{q} + \not{k} + m\right)}{(q+k)^2 - m^2} \right)$$

The minus sign out front is from the fermion loop. Some boiling, which you can find in Peskin (page 247) or Zee (§III.7), reduces this to something manageable. The steps involved are: (1) a trick to combine the denominators, like the Feynman trick $\frac{1}{AB} = \int_0^1 dx \left(\frac{1}{(1-x)A + xB}\right)^2$. (2) some Dirac algebra, to turn the numerator into a polynomial in $k, q$. As Zee says, our job in this course is not to train to be professional integrators. The result of this boiling can be written

$$\mathbf{i}\Pi_2^{\mu\nu}(q) = -e^2 \int đ^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{\left(\ell^2 - \Delta\right)^2}$$

with $\ell = k + xq$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)q^2$, and the numerator is

$$N^{\mu\nu} = 2\ell^\mu\ell^\nu - \eta^{\mu\nu}\ell^2 - 2x(1-x)q^\mu q^\nu + \eta^{\mu\nu}\left(m^2 + x(1-x)q^2\right) + \text{terms linear in } \ell^\mu \quad .$$

At this point I have to point out a problem with applying the *regulator* we've been using (I emphasize that this is a distinct issue from the choice of RG scheme). With a euclidean momentum cutoff, the diagram  gives something of the form

$$\mathbf{i}\Pi_2^{\mu\nu} \propto e^2 \int^\Lambda d^4\ell_E \frac{\ell_E^2 \eta^{\mu\nu}}{\left(\ell_E^2 - \Delta\right)^2} + \ldots \propto e^2\Lambda^2\eta^{\mu\nu}$$

This is *NOT* of the form $\Pi^{\mu\nu} = \Delta_T^{\mu\nu}\Pi(p^2)$; rather it produces a correction to the photon mass proportional to the cutoff. What happened? Our cutoff was not gauge invariant. Oops.[16]

**Fancier PV regularization.** [Zee page 202] We can fix the problem by adding also heavy Pauli-Villars electron ghosts. Suppose we add a bunch of them with masses

[16]Two points: How could we have predicted that the cutoff on euclidean momentum $\ell_E^2 < \Lambda^2$ would break gauge invariance? Its violation of the Ward identity here is a proof, but involved some work. The idea is that the momentum of a charged field shifts under a gauge transformation. Second: it *is* possible to construct a gauge invariant regulator with an explicit UV cutoff, using a lattice. The price, however, is that the gauge field enters only via the link variables $U(x, \hat{e}) = e^{\mathbf{i}\int_x^{x+\hat{e}} A}$ where $x$ is a site in the lattice and $\hat{e}$ is the direction to a neighboring site in the lattice. For more, look up 'lattice gauge theory' in Zee's index. More on this later.

$m_a$ and couplings $\sqrt{c_a}e$ to the photon. Then the vacuum polarization is that of the electron itself plus

$$-\sum_a c_a \int đ^D k \operatorname{tr}\left( (\mathbf{i}e\gamma^\mu) \frac{\mathbf{i}}{\not{q}+\not{k}-m_a} (\mathbf{i}e\gamma^\nu) \frac{\mathbf{i}}{\not{q}-m_a} \right) \sim \int^\Lambda đ^4 k \left( \frac{\sum_a c_a}{k^2} + \frac{\sum_a c_a m_a^2}{p^4} + \cdots \right).$$

So, if we take $\sum_a c_a = -1$ we cancel the $\Lambda^2$ term, and if we take $\sum_a c_a m_a^2 = -m^2$, we also cancel the $\ln \Lambda$ term. This requires at least two PV electron fields, but so what? Once we do this, the momentum integral converges, and the Ward identity applies, so the answer will be of the promised form $\Pi^{\mu\nu} = q^2 \Pi \Delta_T^{\mu\nu}$. After some more boiling, the answer is

$$\Pi_2(q^2) = \frac{1}{2\pi^2} \int dx x(1-x) \ln \frac{M^2}{m^2 - x(1-x)q^2}$$

where $\ln M^2 \equiv -\sum_a c_a \ln m_a^2$. This $M$ plays the role of the UV scale of ignorance thenceforth.

Notice that this is perfectly consistent with our other two one-loop PV calculations: in those, the extra PV electrons never get a chance to run. At higher loops, we would have to make sure to be consistent.

**Dimensional regularization.** A regulator which is more automatically gauge invariant is dimensional regularization (dim reg). I have already been writing many of the integrals in $D$ dimensions. One small difference when we are considering this as a regulator for an integral of fixed dimension is that we don't want to violate dimensional analysis, so we should really replace

$$\int d^4\ell \longrightarrow \int \frac{d^{4-\epsilon}\ell}{\bar{\mu}^{-\epsilon}}$$

where $D = 4 - \epsilon$ and $\bar{\mu}$ is an arbitrary mass scale which will appear in the regulated answers, which we put here to preserve dim'l analysis – *i.e.* the couplings in dim reg will have the same engineering dimensions they had in the unregulated theory (dimensionless couplings remain dimensionless). $\bar{\mu}$ will parametrize our RG, *i.e.* play the role of the RG scale. (It is often called $\mu$ at this step and then suddenly replaced by something also called $\mu$; I will instead call this $\bar{\mu}$ and relate it to the thing that ends up being called $\mu$.)

[Zinn-Justin 4th ed page 233] Dimensionally regularized integrals can be defined systematically with a few axioms indicating how the $D$-dimensional integrals behave under

1. translations $\int đ^D p f(p+q) = \int đ^D p f(p)$ [17]

[17] Note that this rule fails for the euclidean momentum cutoff. Also note that this is the property we needed to demonstrate the Ward identity for the vertex correction using (7.38).

2. scaling $\int đ^D p f(sp) = |s|^{-D} \int đ^D p f(p)$

3. factorization $\int đ^D p \int đ^D q f(p) g(q) = \int đ^D p f(p) \int đ^D q g(q)$

The (obvious?) third axiom implies the following formula for the sphere volume as a continuous function of $D$:

$$\left(\frac{\pi}{a}\right)^{D/2} = \int d^D x e^{-a\vec{x}^2} = \Omega_{D-1} \int_0^\infty x^{D-1} dx e^{-ax^2} = \frac{1}{2} a^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right) \Omega_{D-1} \ . \tag{7.42}$$

This defines $\Omega_{D-1}$ for general $D$.

In dim reg, the one-loop vacuum polarization correction *does* satisfy the gauge-invariance Ward identity $\Pi^{\mu\nu} = \Delta_T^{\mu\nu} \delta\Pi_2$. A peek at the tables of dim reg integrals shows that $\Pi_2$ is:

$$\begin{aligned} \Pi_2(p^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\overset{D \to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right)\right) \end{aligned} \tag{7.43}$$

where we have introduced the heralded $\mu$:

$$\mu^2 \equiv 4\pi \bar{\mu}^2 e^{-\gamma_E}$$

where $\gamma_E$ is the Euler-Mascheroni constant; we define $\mu$ in this way so that, like Rosencrantz and Guildenstern, $\gamma_E$ both appears and disappears from the discussion at this point.

In the second line of (11.10), we expanded the $\Gamma$-function about $D = 4$. Notice that what was a log divergence, becomes a $\frac{1}{\epsilon}$ pole in dim reg. There are other singularities of this function at other integer dimensions. It is an interesting question to ponder why the integrals have such nice behavior as a function of $D$. That is: they only have simple poles. A partial answer is that in order to have worse (*e.g.* essential) singularities at some $D$, the perturbative field theory would have to somehow fail to make sense at larger $D$.

Now we are in a position to choose a renormalization condition (also known as a *renormalization scheme*), which will specify how much of the finite bit of $\Pi$ gets subtracted by the counterterm. One possibility is to demand that the photon propagator is not corrected at $q = 0$, *i.e.* demand $Z_\gamma = 1$. Then the resulting one-loop shift is

$$\delta\Pi_2(q^2) \equiv \Pi_2(q^2) - \Pi_2(0) = \frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \log\left(\frac{m^2 - x(1-x)q^2}{m^2}\right).$$

We'll use this choice below.

Another popular choice, about which more later, is called the $\overline{\text{MS}}$ scheme, in which $\Pi$ is defined by the rule that we *subtract the* $1/\epsilon$ *pole*. This means that the counterterm is

$$\delta_{F^2}^{(\overline{\text{MS}})} = -\frac{e^2}{2\pi^2}\frac{2}{\epsilon}\underbrace{\int_0^1 dx x(1-x)}_{=1/6}.$$

(Confession: I don't know how to state this in terms of a simple renormalization condition on $\Pi_2$. Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and $\mu$.) The resulting vacuum polarization function is

$$\delta\Pi_2^{(\overline{\text{MS}})}(p^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{\mu^2}\right).$$

### 7.7.2 Physics from vacuum polarization

One class of physical effects of vacuum polarization arise from attaching the corrected photon propagator to a static delta-function charge source. The resulting effective Coulomb potential is the fourier transform of

$$\tilde{V}(q) = \frac{1}{q^2}\frac{e^2}{1-\Pi(q^2)} \equiv \frac{e^2_{\text{eff}}(q)}{q^2}. \tag{7.44}$$

This has consequences in both IR and UV.

$\boxed{\text{IR:}}$ In the IR ($q^2 \ll m^2$), it affects the spectra of atoms. The leading correction is

$$\delta\Pi_2(q) = \int dx x(1-x)\ln\left(1 - \frac{q^2}{m^2}x(1-x)\right) \overset{q\ll m}{\simeq} \int dx x(1-x)\left(-\frac{q^2}{m^2}x(1-x)\right) = -\frac{q^2}{30m^2}$$

which means

$$\tilde{V}(q) \overset{q\ll m}{\simeq} \frac{e^2}{q^2} + \frac{e^2}{q^2}\left(-\frac{q^2}{30m^2}\right) + \cdots$$

and hence

$$V(r) = -\frac{e^2}{4\pi r^2} - \frac{e^4}{60\pi^2 m^2}\delta(r) + \cdots \equiv V + \Delta V.$$

This shifts the energy levels of hydrogen $s$-orbitals (the ones with support at the origin) by $\Delta E_s = \langle s|\Delta V|s\rangle$ which contributes to lowering the $2S$ state relative to the $2P$ state (the Lamb shift).

This delta function is actually a long-wavelength approximation to what is called the Uehling potential; its actual range is $1/m_e$, which is the scale on which $\Pi_2$ varies . The delta function approximation is a good idea for atomic physics, since $\frac{1}{m_e} \ll a_0 = \frac{1}{\alpha m_e}$, the Bohr radius. See Schwartz p. 311 for a bit more on this.

$\boxed{\text{UV:}}$ In the UV limit ($q^2 \gg m^2$), we can approximate $\ln\left(1 - \frac{q^2}{m^2}x(1-x)\right) \simeq \ln\left(-\frac{q^2}{m^2}x(1-x)\right) \simeq \ln\left(-\frac{q^2}{m^2}\right)$ to get[18]

$$\Pi_2(q^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x) \ln\left(1 - \frac{q^2}{m^2}x(1-x)\right) \simeq \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x) \ln\left(-\frac{q^2}{m^2}\right) = \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right).$$

Therefore, the effective charge in (7.44) at high momentum exchange is

$$e^2_{\text{eff}}(q^2) \overset{q^2 \gg m_e^2}{\simeq} \frac{e^2}{1 - \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right)}. \tag{7.45}$$

(Remember that $q^2 < 0$ for t-channel exchange, as in the static potential, so the argument of the log is positive and this is real.)

Two things: if we make $q^2$ big enough, we can make the loop correction as big as the 1. This requires $|q| \sim 10^{286}$ eV. Good luck with that. This is called a *Landau pole*. The second thing is: this perspective of a scale-dependent coupling is very valuable, and is a crucial ingredient in the renormalization group.

[18]The last step is safe since the $x(1-x)$ suppresses the contributions of the endpoints of the $x$ integral, so we can treat $x(1-x)$ as finite.

# 8 Consequences of unitarity

Next I would like to fulfill my promise to show that conservation of probability guarantees that some things are positive (for example, $Z$ and $1 - Z$, where $Z$ is the wavefunction renormalization factor). We will show that amplitudes develop an imaginary part when the virtual particles become real. (Someone should have put an extra factor of $\mathbf{i}$ in the definition to resolve this infelicity.) We will discuss the notion of *density of states* in QFT (this should be a positive number!), and in particular the notion of the *density of states contributing to a correlation function* $G = \langle \mathcal{O}\mathcal{O} \rangle$, also known as the *spectral density* of $G$ (or of the operator $\mathcal{O}$). In high-energy physics this idea is associated with the names Källen-Lehmann and is part of a program of trying to use complex analysis to make progress in QFT. These quantities are also ubiquitous in the theory of condensed matter physics and participate in various *sum rules*. This discussion will be a break from perturbation theory; we will say things that are true with a capital 't'.

## 8.1 Spectral density

[Zee III.8, Appendix 2, Xi Yin's notes for Harvard Physics 253b] In the following we will consider a (time-ordered) two-point function of an operator $\mathcal{O}$. We will make hardly any assumptions about this operator. We will assume it is a scalar under rotations, and will assume translation invariance in time and space. But we need not assume that $\mathcal{O}$ is 'elementary'. This is an extremely loaded term, a useful definition for which is: a field governed by a nearly-quadratic action. Also: try to keep an eye out for where (if anywhere) we assume Lorentz invariance.

So, let

$$-\mathbf{i}\mathcal{D}(x) \equiv \langle 0| \mathcal{T}\mathcal{O}(x)\mathcal{O}^\dagger(0) |0\rangle .$$

Notice that we do not assume that $\mathcal{O}$ is hermitian. Use translation invariance to move the left operator to the origin: $\mathcal{O}(x) = e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}$. This follows from the statement that $\mathbf{P}$ generates translations [19]

$$\partial_\mu \mathcal{O}(x) = \mathbf{i}[\mathbf{P}_\mu, \mathcal{O}(x)] .$$

[19]Note that $\mathbf{P}$ here is a $D$-component vector of operators

$$\mathbf{P}_\mu = (\mathbf{H}, \vec{\mathbf{P}})_\mu$$

which includes the Hamiltonian – we are using relativistic notation – but we haven't actually required any assumption about the action of boosts.

And let's unpack the time-ordering symbol:

$$-\mathbf{i}\mathcal{D}(x) = \theta(t)\langle 0| e^{\mathbf{iP}x}\mathcal{O}(0)e^{-\mathbf{iP}x}\mathcal{O}^\dagger(0)|0\rangle + \theta(-t)\langle 0|\mathcal{O}^\dagger(0)e^{\mathbf{iP}x}\mathcal{O}(0)e^{-\mathbf{iP}x}|0\rangle\,. \quad (8.1)$$

Now we need a resolution of the identity operator on the *entire* QFT $\mathcal{H}$:

$$\mathbb{1} = \sum_n |n\rangle\langle n|\,.$$

This innocent-looking $n$ summation variable is hiding an enormous sum! Let's also assume that the groundstate $|0\rangle$ is translation invariant:

$$\mathbf{P}|0\rangle = 0.$$

We can label each state $|n\rangle$ by its total momentum:

$$\mathbf{P}|n\rangle = p_n|n\rangle\,.$$

Let's examine the first term in (8.1); sticking the $\mathbb{1}$ in a suitable place:

$$\langle 0| e^{\mathbf{iP}x}\mathcal{O}(0)\mathbb{1}e^{-\mathbf{iP}x}\mathcal{O}^\dagger(0)|0\rangle = \sum_n \langle 0|\mathcal{O}(0)|n\rangle\langle n|e^{-\mathbf{iP}x}\mathcal{O}^\dagger(0)|0\rangle = \sum_n e^{-\mathbf{i}p_n x}\|\,\mathcal{O}_{0n}\|^2\,,$$

with $\mathcal{O}_{0n} \equiv \langle 0|\mathcal{O}(0)|n\rangle$ the matrix element of our operator between the vacuum and the state $|n\rangle$. Notice the absolute value: unitarity of our QFT requires this to be positive and this will have valuable consequences. [End of Lecture 27]

Next we work on the time-ordering symbol. I claim that :

$$\theta(x^0) = \theta(t) = -\mathbf{i}\int đ\omega \frac{e^{+\mathbf{i}\omega t}}{\omega - \mathbf{i}\epsilon} \quad ; \qquad \theta(-t) = +\mathbf{i}\int đ\omega\frac{e^{+\mathbf{i}\omega t}}{\omega + \mathbf{i}\epsilon}\,.$$

Just like in our discussion of the Feynman contour, the point of the $\mathbf{i}\epsilon$ is to push the pole inside or outside the integration contour. The half-plane in which we must close the contour depends on the sign of $t$. There is an important sign related to the orientation with which we circumnavigate the pole. Here is a check that we got the signs and factors right:

$$\frac{d\theta(t)}{dt} = -\mathbf{i}\partial_t\int đ\omega\frac{e^{\mathbf{i}\omega t}}{\omega - \mathbf{i}\epsilon} = \int đ\omega e^{\mathbf{i}\omega t} = \delta(t).$$

Consider now the fourier transform of $\mathcal{D}(x)$ (for simplicity, I've assumed $\mathcal{O} = \mathcal{O}^\dagger$ here):

$$-\mathbf{i}\mathcal{D}(q) = \int d^D x e^{\mathbf{i}qx}\mathbf{i}\mathcal{D}(x) = \mathbf{i}(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\|^2\left(\frac{\delta^{(D-1)}(\vec{q}-\vec{p}_n)}{q^0 - p_n^0 + \mathbf{i}\epsilon} - \frac{\delta^{(D-1)}(\vec{q}+\vec{p}_n)}{q^0 + p_n^0 - \mathbf{i}\epsilon}\right)\,. \quad (8.2)$$

With this expression in hand, you could imagine measuring the $\mathcal{O}_{0n}$s and using that to determine $\mathcal{D}$.

Now suppose that our operator $\mathcal{O}$ is capable of creating a single particle (for example, suppose, if you must, that $\mathcal{O} = \phi$, a perturbative quantum field). Such a state is labelled only by its spatial momentum: $\left|\vec{k}\right\rangle$. The statement that $\mathcal{O}$ can create this state from the vacuum means

$$\left\langle\vec{k}\right| \mathcal{O}(0)\left|0\right\rangle = \frac{Z^{\frac{1}{2}}}{\sqrt{(2\pi)^{D-1} 2\omega_{\vec{k}}}} \tag{8.3}$$

where $Z \neq 0$ and $\omega_{\vec{k}}$ is the energy of the particle as a function of $\vec{k}$. For a Lorentz invariant theory, we can parametrize this as

$$\omega_{\vec{k}} \overset{\text{Lorentz!}}{\equiv} \sqrt{\vec{k}^2 + m^2}$$

in terms of $m$, the *mass* of the particle. [20] What is $Z$? From (8.3) and the axioms of QM, you can see that it's the probability that $\mathcal{O}$ creates this 1-particle state from the vacuum. In the free field theory it's 1, and it's positive because it's a probability. $1-Z$ measures the extent to which $\mathcal{O}$ does anything *besides* create this 1-particle state.

The identity of the one-particle Hilbert space (relatively tiny!) $\mathcal{H}_1$ is

$$\mathbb{1}_1 = \int đ^{D-1}\vec{k} \left|\vec{k}\right\rangle \left\langle\vec{k}\right|, \quad \left\langle\vec{k}|\vec{k}'\right\rangle = \delta^{(D-1)}(\vec{k}-\vec{k}').$$

This is a summand in the whole horrible resolution:

$$\mathbb{1} = \mathbb{1}_1 + \cdots .$$

[20] It's been a month or two since we spoke explicitly about free fields, so let's remind ourselves about the appearance of $\omega^{-\frac{1}{2}}$ in (8.3), recall the expansion of a free scalar field in creation an annihilation operators:

$$\phi(x) = \int \frac{đ^{D-1}\vec{p}}{\sqrt{2\omega_{\vec{p}}}} \left(\mathbf{a}_{\vec{p}} e^{-\mathbf{i}px} + \mathbf{a}^\dagger_{\vec{p}} e^{\mathbf{i}px}\right) \quad .$$

For a free field $\left|\vec{k}\right\rangle = \mathbf{a}^\dagger_{\vec{k}}\left|0\right\rangle$, and $\left\langle\vec{k}\right| \phi(0)\left|0\right\rangle = \frac{1}{\sqrt{(2\pi)^{D-1}2\omega_{\vec{k}}}}$. The factor of $\omega^{-\frac{1}{2}}$ is required by the ETCRs:

$$[\phi(\vec{x}), \pi(\vec{x}')] = \mathbf{i}\delta^{D-1}(\vec{x}-\vec{x}'), \quad [\mathbf{a}_{\vec{k}}, \mathbf{a}^\dagger_{\vec{k}'}] = \delta^{D-1}(\vec{k}-\vec{k}') ,$$

where $\pi = \partial_t\phi$ is the canonical field momentum. It is just like in the simple harmonic oscillator, where

$$\mathbf{q} = \sqrt{\frac{\hbar}{2m\omega}}\left(\mathbf{a}+\mathbf{a}^\dagger\right), \quad \mathbf{p} = \mathbf{i}\sqrt{\frac{\hbar\omega}{2}}\left(\mathbf{a}-\mathbf{a}^\dagger\right).$$

I mention this because it lets us define the part of the horrible $\sum_n$ in (8.2) which comes from 1-particle states:

$$\begin{aligned}
\implies \quad -\mathbf{i}\mathcal{D}(q) &= ... + \mathbf{i}(2\pi)^{D-1}\int đ^{D-1}\vec{k}\frac{Z}{(2\pi)^{D-1}2\omega_k}\left(\frac{\delta^{D-1}(\vec{q}-\vec{k})}{q^0-\omega_{\vec{k}}+\mathbf{i}\epsilon} - (\omega_k \to -\omega_k)\right) \\
&= ... + \mathbf{i}\frac{Z}{2\omega_q}\left(\frac{1}{q^0-\omega_q+\mathbf{i}\epsilon} - \frac{1}{q^0+\omega_q+\mathbf{i}\epsilon}\right) \\
&\overset{\text{Lorentz}}{=} ... + \mathbf{i}\frac{Z}{q^2-m^2+\mathbf{i}\epsilon}
\end{aligned}$$

(Here again ... is contributions from states involving something else, *e.g.* more than one particle.) The big conclusion here is that even in the interacting theory, even if $\mathcal{O}$ is composite and complicated, if $\mathcal{O}$ can create a 1-particle state with mass $m$ with probability $Z$, then its 2-point function has a pole at the right mass, and the residue of that pole is $Z$. (This result was promised last quarter when we discussed LSZ.)[21]

---

The imaginary part of $\mathcal{D}$ is called the *spectral density* $\boldsymbol{\rho}$ (beware that different physicists have different conventions for the factor of $\mathbf{i}$ in front of the Green's function; the spectral density is not always the imaginary part, but it's always positive (in unitary theories)!

Using

$$\mathrm{Im}\frac{1}{Q\mp\mathbf{i}\epsilon} = \pm\pi\delta(Q), \qquad \text{(for } Q \text{ real)}. \tag{8.4}$$

we have

$$\mathrm{Im}\,\mathcal{D}(q) = \pi\,(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\|^2\left(\delta^D(q-p_n)+\delta^D(q+p_n)\right).$$

More explicitly (for real operators):

$$\mathrm{Im}\,\mathbf{i}\int d^Dx\, e^{\mathbf{i}qx}\langle 0|\,\mathcal{T}\mathcal{O}(x)\mathcal{O}(0)\,|0\rangle = \pi\,(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\|^2\left(\delta^D(q-p_n)+\underbrace{\delta^D(q+p_n)}_{=0\text{ for }q^0>0\text{ since }p_n^0>0}\right).$$

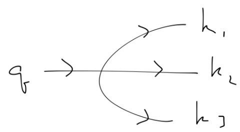
The second term on the RHS vanishes when $q^0 > 0$, since states in $\mathcal{H}$ have energy bigger than the energy of the groundstate. Therefore, the contribution of a 1-particle state to the spectral density is:

$$\mathrm{Im}\,\mathcal{D}(q) = ... + \pi Z\delta(q^2-m^2).$$

[21] If we hadn't assumed Lorentz invariance, this would be replaced by the statement: if the operator $\mathcal{O}$ can create a state with energy $\omega$ from the vacuum with probability $Z$, then its Green's function has a pole at that frequency, with residue $Z$.

This quantity $\mathrm{Im}\,\mathcal{D}(q)$ (the spectral density of $\mathcal{O}$) is positive because it is the *number* of states (with $D$-momentum in an infinitesimal neighborhood of $q$), weighted by the modulus of their overlap with the state engendered by the operator on the groundstate.

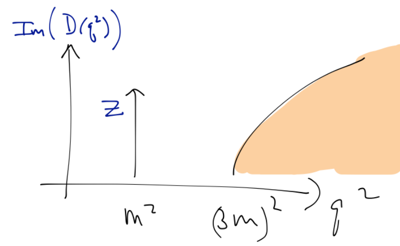
---

Now what about multiparticle states? The associated sum over such states involves multiple (spatial) momentum integrals, not fixed by the total momentum *e.g.* in $\phi^4$ theory: 

the three particles must share the momentum $q$. In this case the sum over all 3-particle states is

$$\sum_{n,\text{ 3-particle states with momentum } q} \propto \int d\vec{k}_1 d\vec{k}_2 d\vec{k}_3 \delta^D(k_1 + k_2 + k_3 - q)$$

Now instead of an isolated pole, we have a whole collection of poles right next to each other. This is a branch cut. In this example, the branch cut begins at $q^2 = (3m)^2$. $3m$ is the lowest energy $q^0$ at which we can produce three particles of mass $m$ (they have to be at rest).




Note that in $\phi^3$ theory, we would instead find that the particle can decay into two particles, and the sum over two particle states would look like

$$\sum_{n,\text{ 2-particle states with momentum } q} \propto \int d\vec{k}_1 d\vec{k}_2 \delta^D(k_1 + k_2 - q)$$

so the branch cut would start at $q^2 = (2m)^2$.

---

Now we recall some complex analysis, in the form of the *Kramers-Kronig (or dispersion) relations*:

$$\mathrm{Re}\,G(z) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega \frac{\mathrm{Im}\,G(\omega)}{\omega - z}$$

(valid if $\mathrm{Im}\,G(\omega)$ is analytic in the UHP of $\omega$ and falls off faster than $1/\omega$). These equations, which I think we were supposed to learn in E&M but no one seems to, and which relate the real and imaginary parts of an analytic function by an integral equation, can be interpreted as the statement that the imaginary part of a complex integral comes from the singularities of the integrand, and conversely that those singularities completely determine the function.

An even more dramatic version of these relations (whose imaginary part is the previous eqn) is

$$f(z) = \frac{1}{\pi} \int dw \frac{\rho(w)}{w - z} , \qquad \rho(w) \equiv \mathrm{Im} f(w + \mathbf{i}\epsilon).$$

The imaginary part determines the whole function.

---

Comments:

- The spectral density $\mathrm{Im}\,\mathcal{D}(q)$ *determines* $\mathcal{D}(q)$. When people get excited about this it is called the "S-matrix program" or something like that.
- The result we've shown protects physics from our caprices in choosing field variables. If someone else uses a different field variable $\eta \equiv Z^{\frac{1}{2}}\phi + \alpha\phi^3$, the result above with $\mathcal{O} = \eta$ shows that

$$\int d^D x e^{\mathbf{i}qx} \langle \mathcal{T} \eta(x) \eta(0) \rangle$$

  still has a pole at $q^2 = m^2$ and a cut starting at the three-particle threshold, $q^2 = (3m)^2$.
- A sometimes useful fact which we've basically already shown (for real operators):

$$\mathrm{Im}\,\mathcal{D}(q) = (2\pi)^D \sum_n \| \mathcal{O}_{0n} \|^2 \left( \delta^D(q - p_n) + \delta^D(q + p_n) \right) = \frac{1}{2} \int d^D x e^{\mathbf{i}qx} \langle 0 | [\mathcal{O}(x), \mathcal{O}(0)] | 0 \rangle \quad .$$

We can summarize what we've learned in the Lorentz-invariant case as follows: In a Lorentz invariant theory, the spectral density $\rho$ for a scalar operator $\phi$ is a scalar function of $p^\mu$ with

$$\sum_s \delta^D(p - p_s) \| \langle 0 | \phi(0) | s \rangle \|^2 = \frac{\theta(p^0)}{(2\pi)^{D-1}} \rho(p^2) \ .$$

Claims:

- $\rho(s) = \mathcal{N} \mathrm{Im}\,\mathcal{D}$ for some number $\mathcal{N}$, when $s > 0$.
- $\rho(s) = 0$ for $s < 0$. There are no states for spacelike momenta.
- $\rho(s) \geq 0$ for $s > 0$. The density of states for timelike momenta is positive or zero.
- With our assumption about one-particle states, $\rho(s)$ has a delta-function singularity at $s = m^2$, with weight $Z$. More generally we have shown that

$$\mathcal{D}(k^2) = \int ds \ \rho(s) \frac{1}{k^2 - s + \mathbf{i}\epsilon}.$$

This is called the Källen-Lehmann spectral representation of the propagator; it represents it as a sum of *free* propagators with different masses, determined by the spectral density. One consequence (assuming unitarity and Lorentz symmetry) is that at large $|k^2|$, the Green's function is bigger than $\frac{1}{k^2}$, since each term in the integral goes like $\frac{1}{k^2}$ and $\rho(s) \geq 0$ means that there cannot be cancellations between each $\frac{1}{k^2-s}$ contribution. This means that if the kinetic term for your scalar field has more derivatives, something must break at short distances (Lorentz is the easiest way out, for example on a lattice).

Taking into account our assumption about single-particle states, this is

$$\mathcal{D}(k^2) = \frac{Z}{k^2 - m^2 + \mathbf{i}\epsilon} + \int_{(3m)^2}^{\infty} ds\ \rho_c(s) \frac{1}{k^2 - s + \mathbf{i}\epsilon}$$

where $\rho_c$ is just the continuum part. The pole at the particle-mass$^2$ survives interactions, with our assumption. (The value of the mass need not be the same as the bare mass!)

- Finally, suppose that the field $\phi$ in question is a canonical field, in the sense that

$$[\phi(x,t), \partial_t \phi(y,t)] = \mathbf{i}\delta^{(d)}(x-y).$$

This is a statement both about the normalization of the field, and that its canonical momentum is its time derivative. Then[22]

$$1 = \int_0^\infty ds \rho(s). \tag{8.6}$$

If we further assume that $\phi$ can create a one-particle state with mass $m$, so that $\rho(s) = Z\delta(s-m^2) + \rho_c(s)$ where $\rho_c(s) \geq 0$ is the contribution from the continuum of $\geq$ 2-particle states, then

$$1 = Z + \int_{\text{threshold}}^\infty ds \rho_c(s)$$

[22] Here's how to see this. For free fields (chapter 2) we have

$$\langle 0|[\phi(x), \phi(y)]|0\rangle_{\text{free}} = \Delta_+(x-y, m^2) - \Delta_+(y-x, m^2),$$

where $\Delta_+(x) = \int \frac{d^dp}{2\omega_{\vec{p}}} e^{-\mathbf{i}p\cdot x}|_{p^0=\omega_{\vec{p}}}$. For an interacting canonical field, we have instead a spectral representation:

$$\langle \Omega|[\phi(x), \phi(y)]|\Omega\rangle = \int d\mu^2 \rho(\mu^2) \left(\Delta_+(x-y, \mu^2) - \Delta_+(y-x, \mu^2)\right), \tag{8.5}$$

where $\rho$ is the same spectral density as above. Now take $\partial_{x^0}|_{x^0=y^0}$ of the BHS of (8.5) and use $\partial_t \Delta_+(x-y; \mu^2)|_{x^0=y^0} = -\frac{\mathbf{i}}{2}\delta^{(d)}(\vec{x}-\vec{y})$.

is a sum rule. It shows that $Z \in [0,1]$ and is just the statement that if the field doesn't create a single particle, it must do something else. The LHS is the probability that *something* happens.

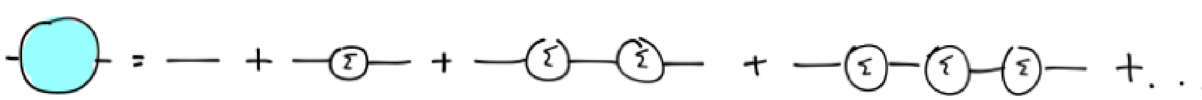
The idea of spectral representation and spectral density is more general than the Lorentz-invariant case. In particular, the spectral density of a Green's function is an important concept in the study of condensed matter. For example, the spectral density for the electron 2-point function is the thing that actually gets measured in angle-resolved photoemission experiments (ARPES).

## 8.2 Cutting rules and optical theorem

[Zee §III.8] So, that may have seemed like some math. What does this mean when we have in our hands a perturbative QFT? Consider the two point function of a relativistic scalar field $\phi$ which has a perturbative cubic interaction:

$$S = \int d^D x \left( \frac{1}{2} \left( (\partial\phi)^2 + m^2 \phi^2 \right) - \frac{g}{3!}\phi^3 \right).$$

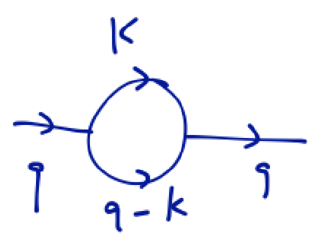
Sum the geometric series of 1PI insertions to get



$$\mathbf{i}\mathcal{D}_\phi(q) = \frac{\mathbf{i}}{q^2 - m^2 - \Sigma(q) + \mathbf{i}\epsilon}$$

where $\Sigma(q)$ is the 1PI two point vertex.

The leading contribution to $\Sigma$ comes from the one loop diagram at right and is




$$\mathbf{i}\Sigma_{1\text{ loop}}(q^2) = (\mathbf{i}g)^2 \int đ^D k \frac{\mathbf{i}}{k^2 - m^2 + \mathbf{i}\epsilon} \frac{\mathbf{i}}{(q-k)^2 - m^2 + \mathbf{i}\epsilon}.$$

Consider this function for real $q$, for which there are actual states of the scalar field – timelike $q^\mu$, with $q^0 > m$. The real part of $\Sigma$ shifts the mass. But what does it mean if this function has an imaginary part?

---

Claim: $\text{Im}\Sigma/m$ is a decay rate.

It moves the energy of the particle off of the real axis from $m$ (in its rest frame) to

$$\sqrt{m^2 + \mathbf{i}\text{Im}\Sigma(m^2)} \overset{\text{small Im}\Sigma \sim g^2}{\simeq} m + \mathbf{i}\frac{\text{Im}\Sigma(m^2)}{2m}.$$

The fourier transform to real time is an amplitude for propagation in time of a state with complex energy $\mathcal{E}$: its wavefunction evolves like $\psi(t) \sim e^{-\mathbf{i}\mathcal{E}t}$ and has norm

$$\| \psi(t) \|^2 \sim \| e^{-\mathbf{i}\left(E - \mathbf{i}\frac{1}{2}\Gamma\right)t} \|^2 = e^{-\Gamma t}.$$

In our case, we have $\Gamma \sim \operatorname{Im}\Sigma(m^2)/m$ (I'll be more precise below), and we interpret that as the rate of decay of the norm of the single-particle state. There is a nonzero probability that the state turns into something else as a result of time evolution in the QFT: the single particle must decay into some other state – generally, multiple particles. (We will see next how to figure out into what it decays.)

The absolute value of the Fourier transform of this quantity $\psi(t)$ is the kind of thing you would measure in a scattering experiment. This is

$$F(\omega) = \int dt\ e^{-\mathbf{i}\omega t}\psi(t) = \int_0^\infty dt\ e^{-\mathbf{i}\omega t} e^{\mathbf{i}\left(M - \frac{1}{2}\mathbf{i}\Gamma\right)t} = \frac{1}{\mathbf{i}\left(\omega - M\right) - \frac{1}{2}\Gamma}$$

$$\| F(\omega) \|^2 = \frac{1}{\left(\omega - M\right)^2 + \frac{1}{4}\Gamma^2}$$

is a Lorentzian in $\omega$ with width $\Gamma$. So $\Gamma$ is sometimes called a *width*.

---

So: what is $\operatorname{Im}\Sigma_{1\text{ loop}}$ in this example?

We will use

$$\frac{1}{k^2 - m^2 + \mathbf{i}\epsilon} = \mathcal{P}\frac{1}{k^2 - m^2} - \mathbf{i}\pi\delta(k^2 - m^2) \equiv \mathcal{P} - \mathbf{i}\Delta$$

where $\mathcal{P}$ denotes 'principal part'. Then

$$\operatorname{Im}\Sigma_{1\text{ loop}}(q) = -g^2 \int d\Phi \left(\mathcal{P}_1\mathcal{P}_2 - \Delta_1\Delta_2\right)$$

with $d\Phi = đ^D k_1 đ^D k_2 (2\pi)^D \delta^D(k_1 + k_2 - q)$.

This next trick, to get rid of the principal part bit, is from Zee's book (the second edition on p.214; he also does the calculation by brute force in the appendix to that section). We can find a representation for the 1-loop self-energy in terms of real-space propagators: it's the fourier transform of the amplitude to create two $\phi$ excitations at the origin at time zero with a single $\phi$ field (this is $-ig$), to propagate them both from 0 to $x$ (this is $(\mathbf{i}\mathcal{D}(x))^2$) and then destroy them both with a single $\phi$ field (this is $-ig$ again). Altogether:

$$\mathbf{i}\Sigma(q) = \int d^dx\ e^{\mathbf{i}qx} \left(\mathbf{i}g\right)^2 \mathbf{i}\mathcal{D}(x)\mathbf{i}\mathcal{D}(x)$$

$$= g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 + \mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \mathbf{i}\epsilon} \tag{8.7}$$

In the bottom expression, the $\mathbf{i}\epsilon$s are designed to produce the *time-ordered* $\mathcal{D}(x)$s. Consider instead the strange combination

$$\begin{aligned} 0 &= \int d^d x \; e^{\mathbf{i}qx} \, (\mathbf{i}g)^2 \, \mathbf{i}\mathcal{D}_{\text{adv}}(x) \mathbf{i}\mathcal{D}_{\text{ret}}(x) \\ &= g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 - \sigma_1 \mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \sigma_2 \mathbf{i}\epsilon} \end{aligned} \tag{8.8}$$

where $\sigma_{1,2} \equiv \text{sign}(k^0_{1,2})$. This expression vanishes because the integrand is identically zero: there is no value of $t$ for which both the advanced and retarded propagators are nonzero (one has a $\theta(t)$ and the other has a $\theta(-t)$, and this is what's accomplished by the red $\sigma$s). Therefore, we can add the imaginary part of zero

$$-\text{Im}\,(0) = -g^2 \int d\Phi \left( \mathcal{P}_1 \mathcal{P}_2 + \sigma_1 \sigma_2 \Delta_1 \Delta_2 \right)$$

to our expression for $\text{Im}\,\Sigma_{\text{1-loop}}$ to cancel the annoying principal part bits:

$$\text{Im}\,\Sigma_{\text{1-loop}} = g^2 \int d\Phi \left( (1 + \sigma_1 \sigma_2) \, \Delta_1 \Delta_2 \right).$$

The quantity $(1 + \sigma_1\sigma_2)$ is only nonzero when $k_1^0$ and $k_2^0$ have the same sign; but in $d\Phi$ is a delta function which sets $q^0 = k_1^0 + k_2^0$. WLOG we can take $q^0 > 0$ since we only care about the propagation of positive-energy states. Therefore both $k_1^0$ and $k_2^0$ must be positive.

[End of Lecture 28]

The result is that the only values of $k$ on the RHS that contribute are ones with *positive* energy, which satisfy all the momentum conservation constraints:

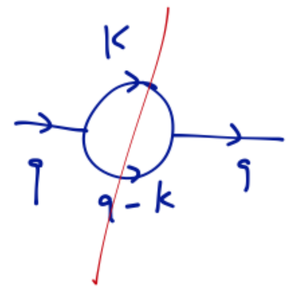
$$\text{Im}\,\Sigma = g^2 \int d\Phi \theta(k_1^0) \theta(k_2^0) \Delta_1 \Delta_2$$

$$= \frac{g^2}{2} \int \frac{đ^{D-1}\vec{k}_1}{2\omega_{\vec{k}_1}} \frac{đ^{D-1}\vec{k}_2}{2\omega_{\vec{k}_2}} (2\pi)^D \delta^D(k_1 + k_2 - q) \; .$$

In the last step we used the identity $\theta(k^0)\delta(k^2 - m^2) = \theta(k^0)\frac{\delta(k^0 - \omega_k)}{2\omega_k}$. But this is exactly the density of actual final states into which the thing can decay! In summary:

$$\operatorname{Im}\Sigma = \sum_{\substack{\text{actual states } n \text{ of 2 particles} \\ \text{into which } \phi \text{ can decay}}} \| \mathcal{A}_{\phi \to n} \|^2 \tag{8.9}$$

In this example the decay amplitude $\mathcal{A}$ is just $\mathbf{i}g$.




This result is generalized by the *Cutkosky cutting rules* for finding the imaginary part of a feynman diagram describing a physical process. The rough rules are the following. Assume the diagram is amputated – leave out the external propagators. Then any line drawn through the diagram which separates initial and final states (as at right) will 'cut' through some number of internal propagators; replace each of the cut propagators by $\theta(p^0)\pi\delta(p^2 - m^2) = \theta(p^0)\frac{\pi\delta(p_0 - \epsilon_p)}{2\epsilon_p}$. As Tony Zee says: the amplitude becomes imaginary when the intermediate particles become real (as opposed to virtual), aka 'go on-shell'. This is a place where the $\mathbf{i}\epsilon$s are crucial.

The general form of (8.9) is a general consequence of unitarity. Recall that the S-matrix is

$$\mathcal{S}_{fi} = \langle f | e^{-\mathbf{i}\mathbf{H}T} | i \rangle \equiv (\mathbb{1} + \mathbf{i}\mathcal{T})_{fi} .$$

$$\mathbf{H} = \mathbf{H}^\dagger \implies \mathbb{1} = \mathcal{S}\mathcal{S}^\dagger \implies 2\mathrm{Im}\mathcal{T} \equiv \mathbf{i}\left(\mathcal{T}^\dagger - \mathcal{T}\right) \overset{\mathbb{1}=\mathcal{S}\mathcal{S}^\dagger}{=} \mathcal{T}^\dagger\mathcal{T}.$$

This is called the optical theorem and it is the same as the one taught in some QM classes. In terms of matrix elements:

$$2\mathrm{Im}\mathcal{T}_{fi} = \sum_n \mathcal{T}^\dagger_{fn}\mathcal{T}_{ni}$$

Here we've inserted a resolution of the identity (again on the QFT Hilbert space, the same scary sum) in between the two $\mathcal{T}$ operators. In the one-loop approximation, in the $\phi^3$ theory here, the intermediate states which can contribute to $\sum_n$ are two-particle states, so that $\sum_n$ will turn into $\int \frac{đ\vec{k}_1}{2\omega_{k_1}} \frac{đ\vec{k}_2}{2\omega_{k_2}}$, the two-particle density of states.

A bit more explicitly, introducing a basis of scattering states

$$\langle f | \mathcal{T} | i \rangle = \mathcal{T}_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}_{fi}, \quad \mathcal{T}^\dagger_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}^\star_{if}, \qquad (\text{recall that } \not{\delta}^d \equiv (2\pi)^d\delta^d)$$

we have

$$\langle F | \mathcal{T}^\dagger \mathbb{1} \mathcal{T} | I \rangle = \sum_n \langle F | \mathcal{T}^\dagger \sum_n \prod_{f=1}^n \int \frac{đ^3 q_f}{2E_f} |\{q_f\}\rangle \langle \{q_f\} | \mathcal{T} | I \rangle$$

$$= \sum_n \prod_{f=1}^n \int \frac{đ^3 q_f}{2E_f} \not\delta^4(p_F - \sum_f q_f) \mathcal{M}^\star_{\{q_f\}F} \not\delta^4(p_I - \sum_f q_f) \mathcal{M}_{\{q_f\}I}$$

Now notice that we have a $\not\delta^4(p_F - p_I)$ on both sides, and

$$\prod_{f=1}^n \int \frac{đ^3 q_f}{2E_f} \not\delta^4(p_F - \sum_f q_f) = \int d\Pi_n$$

is the final-state phase space of the $n$ particles. Therefore, the optical theorem says

$$\mathbf{i}\left(\mathcal{M}^\star_{IF} - \mathcal{M}_{FI}\right) = \sum_n \int d\Pi_n \mathcal{M}^\star_{\{q_f\}F} \mathcal{M}_{\{q_f\}I}.$$

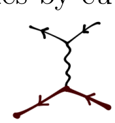
Now consider *forward scattering*, $I = F$ (notice that here it is crucial that $\mathcal{M}$ is the transition matrix, $S = \mathbb{1} + \mathbf{i}\mathcal{T} = \mathbb{1} + \mathbf{i}\not\delta(p_T)\mathcal{M}$):

$$2\text{Im}\,\mathcal{M}_{II} = \sum_n \int d\Pi_n |\mathcal{M}_{\{q_f\}I}|^2.$$

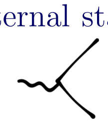
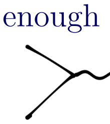
For the special case of 2-particle scattering, we can relate the RHS to the total cross section for $2 \to$ anything:

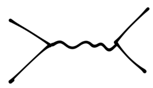
$$\text{Im}\,\mathcal{M}(k_1, k_2 \leftarrow k_1, k_2) = 2E_{cm} p_{cm} \sigma(\text{anything} \leftarrow k_1, k_2).$$

Recall that for real $x$ the imaginary part of a function of one variable with a branch cut, (like $\text{Im}\,(x + \mathbf{i}\epsilon)^\nu = \frac{1}{2}\left((x + \mathbf{i}\epsilon)^\nu - (x - \mathbf{i}\epsilon)^\nu\right)$) is equal to (half) the *discontinuity* of the function $((x)^\nu)$ across the branch cut.

In more complicated examples (such as a box diagram contributing to 2-2 scattering), there can be more than one way to cut the diagram. Different ways of cutting the diagram correspond to discontinuities in different kinematical variables. To get the whole imaginary part, we have to add these up. A physical cut is a way of separating all initial-state particles from all final-state particles by cutting only internal lines. So for example, a $t$-channel tree-level diagram (like  ) never has any imaginary part; this makes sense because the momentum of the exchanged particle is spacelike.

---

**Resonances.** A place where this technology is useful is when we want to study short-lived particles. In our formula for transition rates and cross sections we assumed plane waves for our external states. Some particles don't live long enough for separately producing them:  and then watching them decay:  ;

instead we must find them as resonances in scattering amplitudes of other particles: Im $\Bigg($  $\Bigg)$.

So, consider the case $\mathbf{i}\mathcal{M} = \langle F| \mathbf{i}\mathcal{T} |I\rangle$ where both $I$ and $F$ are one-particle states. A special case of the LSZ formula says

$$\mathcal{M} = -\left(\sqrt{Z}\right)^2 \Sigma = -Z\Sigma \tag{8.10}$$

where $-\mathbf{i}\Sigma$ is the amputated 1-1 amplitude, that is the self-energy, sum of all connected and amputated diagrams with one particle in and one particle out. Let $\Sigma(p) = A(p^2) + \mathbf{i}B(p^2)$ (not standard notation), so that near the pole in question, the propagator looks like

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{p^2 - m_0^2 - \Sigma(p)} \simeq \frac{\mathbf{i}}{(p^2 - m^2)\underbrace{\left(1 - \partial_{p^2} A|_{m^2}\right)}_{=Z^{-1}} - \mathbf{i}B} = \frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}BZ}. \tag{8.11}$$

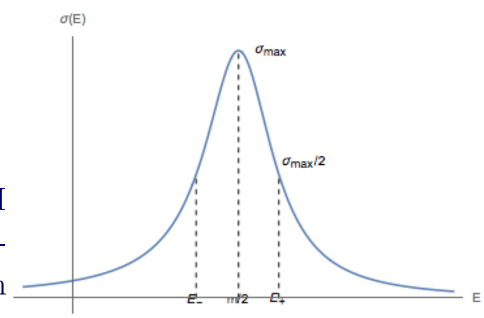
In terms of the *particle width* $\Gamma_w \equiv -ZB(m^2)/m$, this is

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}m\Gamma_w}.$$

So, if we can make the particle whose propagator we're discussing in the $s$-channel, the cross-section will be proportional to

$$\left|\tilde{G}^{(2)}(p)\right|^2 = \left|\frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}m\Gamma_w}\right|^2 = \frac{Z^2}{(p^2 - m^2)^2 + m^2\Gamma_w^2}$$

a Lorentzian or Breit-Wigner distribution: In the COM frame, $p^2 = 4E^2$, and the cross section $\sigma(E)$ has a resonance peak at $2E = m$, with width $\Gamma_w$. It is the width in the sense that the function is half its maximum when $E = E_\pm = \sqrt{\frac{m(m\pm\Gamma_w)}{4}} \simeq \frac{m}{2} \pm \frac{\Gamma}{4}$.




This width is the same as the decay rate, because of the optical theorem:

$$\Gamma_w = -\frac{BZ}{m} \overset{(8.10)}{=} -\frac{1}{m}\left(-\mathrm{Im}\mathcal{M}_{1\to 1}\right) \overset{\text{optical}}{=} \frac{1}{m}\frac{1}{2}\sum_n \int_f d\Pi_n |\mathcal{M}_{\{q_f\}1}|^2 = \Gamma$$

the last equation of which is exactly our formula for the decay rate. If it is not the case that $\Gamma \ll m$, *i.e.* if the resonance is too broad, the Taylor expansion of the inverse propagator we did in (8.11) may not be such a good idea.

**Unitarity and high-energy physics.** Two comments: (1) there had better not be any cutoff dependence in the imaginary part. If there is, we'll have trouble cancelling it by adding counterterms – an imaginary part of the action will destroy unitarity. This is elaborated a bit in Zee's discussion.

(2) Being bounded by 1, probabilities can't get too big. Cross sections are also bounded: there exist precise bounds from unitarity on the growth of cross sections with energy, such as the Froissart bound, $\sigma_{\text{total}}(s) \leq C \ln^2 s$ for a constant $C$. Xi Yin's notes describe a proof.

On the other hand, consider an interaction whose coupling $G$ has mass dimension $k$. The cross section to which $G$ contributes has dimensions of area, and comes from squaring an amplitude proportional to $G$, so comes with at least two powers of $G$. At $E \gg$ anything else, these dimensions must be made up with powers of $E$:

$$\sigma(E \gg ...) \sim G^2 E^{-2-2k}. \tag{8.12}$$

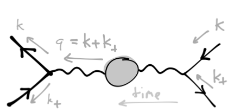
This means that if $k \leq -1$, the cross section grows at high energy. In such a case, something else must happen to 'restore unitarity'. One example is Fermi's theory of Weak interactions, which involves a 4-fermion coupling $G_F \sim M_W^{-2}$. Here we know what happens, namely the electroweak theory, about which more soon. In gravity, $G_N \sim M_{\text{Pl}}^{-2}$, we can't say yet.

## 8.3 How to study hadrons with perturbative QCD

[Peskin §18.4] Here is a powerful physics application of both the optical theorem and the spectral representation. Consider the total inclusive cross section for $e^+e^-$ scattering at energies $s = (k + k_+)^2 \gg m_e^2$:

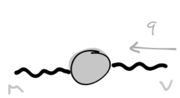
$$\sigma^{\text{anything}\leftarrow e^+e^-} \overset{\text{optical thm}}{=} \frac{1}{2s} \text{Im} \mathcal{M}(e^+e^- \leftarrow e^+e^-) \tag{8.13}$$

where on the RHS, $\mathcal{M}$ is the forward scattering amplitude (meaning that the initial and final electrons have the same momenta). We've learned a bit about the contributions of electrons and muons to the BHS of this expression, what about QCD? To leading order in $\alpha$ (small), but to all orders in the strong coupling $\alpha_s$ (big at low energies), the contributions of QCD look like

$\mathbf{i}\mathcal{M}_h =$ 

$= (-\mathbf{i}e)^2 \bar{u}(k)\gamma_\mu v(k_+) \frac{-\mathbf{i}}{s} \mathbf{i}\Pi_h^{\mu\nu}(q) \frac{-\mathbf{i}}{s} \bar{v}(k_+)\gamma_\nu u(k)$

with


$= \mathbf{i}\Pi_h^{\mu\nu}(q) \overset{\text{Ward}}{=} \mathbf{i}(q^2\eta^{\mu\nu} - q^\mu q^\nu)\Pi_h(q^2)$

the hadronic contribution to the vacuum polarization. We can pick out the contribution of the strong interactions by just keeping these bits on the BHS of (8.13):

$$\sigma^{\text{hadrons}\leftarrow e^+e^-} = \frac{1}{4}\sum_{\text{spins}} \frac{\text{Im}\,\mathcal{M}_h}{2s} = -\frac{4\pi\alpha}{s}\text{Im}\,\Pi_h(s).$$

(The initial and final spins are equal and we average over initial spins. We can ignore the longitudinal term $q^\mu q^\nu$ by the Ward identity. The spinor trace is $\sum_{\text{spins}} \bar{u}(k)\gamma_\mu v(k_+)\bar{v}(k_+)\gamma^\mu u(k) = -2k\cdot k_+ = -s$.) As a reality check, consider the contribution from one loop of a heavy lepton of mass $M^2 \gg m_e^2$:

$$\text{Im}\,\Pi_L(s+\mathbf{i}\epsilon) = -\frac{\alpha}{3}F(M^2/s)$$

and

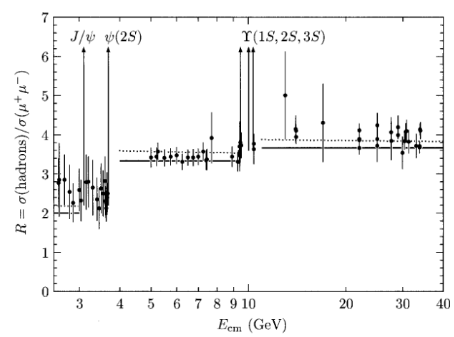
$$\sigma^{L^+L^-\leftarrow e^+e^-} = \frac{4\pi}{3}\frac{\alpha^2}{s}F(M^2/s)$$

with

$$F(M^2/s) = \begin{cases} 0, & s < (2M)^2 \\ \sqrt{1-\frac{4M^2}{s}}\left(1+\frac{2M^2}{s}\right) = 1 + \mathcal{O}\left(M^2/s\right), & s > (2M)^2 \end{cases}.$$

In perturbative QCD, the leading order result is the same from each quark with small enough mass:

$$\sigma_0^{\text{quarks}\leftarrow e^+e^-} = \underbrace{3}_{\text{colors}} \sum_{\text{flavors, f}} Q_f^2 \frac{4\pi}{3}\frac{\alpha^2}{s}F(m_f^2/s).$$




This actually does remarkably well as a crude approximation to the measured $\sigma(\text{hadrons} \leftarrow e^+e^-)$ – see Fig. 5.3 of Peskin, at right. (This figure does not appear in the paper Peskin cites, I'm not sure of the correct provenance. The key point is that the ratio of the hadronic cross section to that for muons in the final state jumps at $E = 2m_f$ for each new quark flavor (you can see $m_c \sim 1.3$ GeV and $m_b \sim 4.5$ GeV in the figure). See Peskin pp 139-141 for more.

But Q: why is a perturbative analysis of QCD relevant here? You might think asymptotic freedom means QCD perturbation theory is good at high energy or short distances, and that seems to be borne out by noticing that $\Pi_h$ is a two-point function of the quark contributions to the EM current:

$$\mathbf{i}\Pi_h^{\mu\nu}(q) = -e^2\int d^4x\; e^{-\mathbf{i}q\cdot x}\left\langle\Omega\right|\mathcal{T}J^\mu(x)J^\nu(0)\left|\Omega\right\rangle, \quad J^\mu(x) \equiv \sum_f Q_f\bar{q}_f(x)\gamma^\mu q_f(x).$$

(Here, the quark fields $q_f$ are Dirac spinors, with Lagrangian $L_q = \sum_f \bar{q}_f \left(\mathbf{i}\not{D} - m_f\right) q_f$, $D_\mu = \partial_\mu - \mathbf{i}Q_f A_\mu + \ldots$, where the ... is the coupling to the gluon field which we'll discuss next chapter. They have a color index which runs from 1 to 3 which I've suppressed.) Maybe it looks like we are taking $x \to 0$ and therefore studying short distances. But if we are interested in large *timelike* $q^\mu$ here, that means that dominant contributions to the $x$ integral are when the two points are timelike separated, and in the resolution of the identity in between the two $J$s includes physical states of QCD with lots of *real* hadrons. In contrast, the limit where we can do (maybe later we will learn how) perturbative QCD is when $q^2 = -Q_0^2 > 0$ is spacelike. (Preview: We can use the operator product expansion of the two currents.)

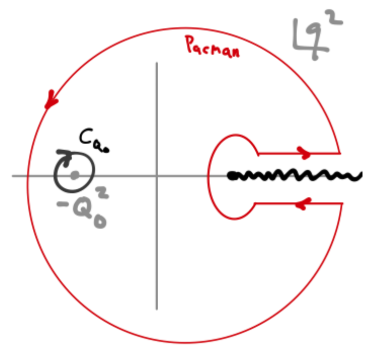
How can we use this knowledge to find the answer in the physical regime of $q^2 > 0$? The fact that $\Pi_h$ is a two-point function means that it has a spectral representation. It is analytic in the complex $q^2$ plane except for a branch cut on the positive real axis coming from production of real intermediate states, exactly where we want to know the answer. One way to encode the information we know is to package it into moments:

$$I_n \equiv -4\pi\alpha \oint_{C_{Q_0}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}} = -\frac{4\pi\alpha}{n!} \left(\partial_{q^2}\right)^n \Pi_h|_{q^2=-Q_0^2}.$$

The idea here is that the RHS can be computed by perturbative QCD. On the other hand, we know from the (appropriate generalization to currents of the) spectral representation sum rule (8.6) that $\Pi_h(q^2) \overset{|q|\gg\ldots}{\lesssim} \log(q^2)$, so for $n \geq 1$, the contour at infinity can be ignored.

Therefore

$$\begin{aligned} I_n &= -4\pi\alpha \oint_{\text{Pacman}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}} \\ &= -4\pi\alpha \int \frac{dq^2}{4\pi\mathbf{i}} \frac{\mathrm{Disc}\Pi_h}{(q^2+Q_0^2)^{n+1}} \\ &= \frac{1}{\pi}\int_{s_{\text{threshhold}}}^\infty ds \frac{s}{(s+Q_0^2)^{n+1}} \sigma^{\text{hadrons}\leftarrow e^+e^-}(s). \end{aligned}$$




On the RHS is (moments of) the measurable (indeed, measured) cross-section, and on the LHS is things we can calculate (later). If the convergence of these integrals were uniform in $n$, we could invert this relation and directly try to predict the cross section to hadrons. But it is not, and the correct cross section varies about the leading QCD answer more and more at lower energies, culminating at various Breit-Wigner resonance peaks at $\bar{q}q$ boundstates.

[End of Lecture 29]

# 9 Gauge theory

## 9.1 Massive vector fields as gauge fields

Consider a massive vector field $B_\mu$ with Lagrangian density

$$\mathcal{L}_B = -\frac{1}{4e^2}(dB)_{\mu\nu}(dB)^{\mu\nu} + \frac{1}{2}m^2 B_\mu B^\mu$$

where $(dB)_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$. (Note the funny-looking sign of the mass term which comes from $B^\mu B_\mu = B_0^2 - B_i^2$.) The mass term is not invariant under $B_\mu \to B_\mu + \partial_\mu \lambda$, the would-be gauge transformation. We can understand the connection between massive vector fields and gauge theory by the 'Stueckelberg trick' of pretending that the gauge parameter is a field: Let $B_\mu \equiv A_\mu - \partial_\mu \theta$ where $\theta$ is a new degree of freedom. Since $B$ is invariant under the transformation

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \lambda(x), \ \theta(x) \to \theta(x) + \lambda(x),$$

so is any functional of $B$. Notice that the fake new field $\theta$ transforms non-linearly (*i.e.*its transformation is affine). This was just a book-keeping step, but something nice happens:

$$(dB)_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}$$

is the field strength of $A$. The mass term becomes

$$B_\mu B^\mu = (A_\mu - \partial_\mu \theta)(A^\mu - \partial^\mu \theta).$$

This contains a kinetic term for $\theta$. We can think of this term as (energetically) setting $\theta$ equal to the longitudinal bit of the gauge field. One nice thing about this reshuffling is that the $m \to 0$ limit decouples the longitudinal bits. Furthermore, if we couple a conserved current ($\partial^\mu j_\mu = 0$) to $B$, then

$$\int d^D x \ j_\mu B^\mu = \int d^D x j_\mu A^\mu$$

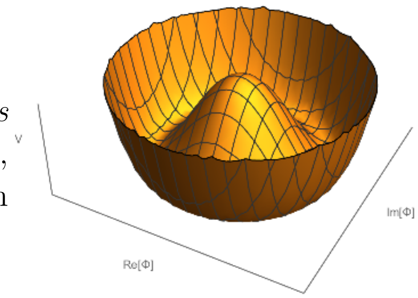
it is the same as coupling to $A_\mu$.

Who is $\theta$? Our previous point of view was that it is fake and we can just choose the gauge parameter $\lambda(x)$ to get rid of it, and set $\theta(x) = 0$. This is called *unitary gauge*, and gives us back the Proca theory of $B = A$. Alternatively, consider the following slightly bigger (more dofs) theory:

$$\mathcal{L}_h \equiv -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}|D_\mu \Phi|^2 + V(|\Phi|)$$

where $\Phi$ is a complex, charged scalar field whose covariant derivative is $D_\mu \Phi = (\partial_\mu - \mathbf{i} A_\mu)\, \Phi$, and let's take

$$V(|\Phi|) = \kappa (|\Phi|^2 - v^2)^2$$




for some couplings $\kappa, v$. This is called an *Abelian Higgs model*. This potential has a $\mathsf{U}(1)$ symmetry $\Phi \to e^{\mathbf{i}\alpha}\Phi$, and a circle of minima at $|\Phi|^2 = v^2$ (if $v^2 > 0$, which we'll assume).

In polar coordinates in field space, $\Phi \equiv \rho e^{\mathbf{i}\theta}$, the Lagrangian is

$$\mathcal{L}_h = -\frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\rho^2 (A_\mu - \partial_\mu \theta)^2 + (\partial \rho)^2 + V(\rho).$$

This differs from the action for $B$ written in terms of $A, \theta$ only in the addition of the *Higgs mode* $\rho$. Again we can go to unitary gauge and set $\theta = 0$. We find a massive gauge field $A$, plus a massive scalar $\rho$ whose mass (expanding $V(\rho)$ about $\rho = v$) is

$$\partial_\rho^2 V|_{\rho = v} = m_\rho^2 = 8\kappa v^2 \overset{\kappa \gg 1}{\gg} m_A^2 = \langle \rho \rangle^2 = v^2.$$

That is: in the limit of large $\kappa$, the excitations of $\rho$ are hard to make, and we get back $\mathcal{L}_B$. For any value of $\kappa$, we can say that the gauge field eats the would-be Goldstone boson $\theta$ and becomes heavy, in a manner consistent with gauge invariance[23]. This is the Higgs mechanism.

The description of massive gauge fields in terms of $\mathcal{L}_h$ via the Higgs mechanism is more useful than $\mathcal{L}_B$ for thinking about the renormalization of massive gauge fields: for example it is renormalizible, even if we couple $A$ to other charged fields (*e.g.* Dirac fermions). This mechanism also works in the case of non-Abelian gauge fields and is an important ingredient in the (electroweak sector of the) Standard Model.

It is also a description of what happens to the EM field in a superconductor: the photon gets a mass; the resulting expulsion of magnetic flux is called the Meissner effect. For example, if we immerse a region $x > 0$ with $\Phi = v$ in an external constant magnetic field $B_0$, $0 = \partial_\mu F^{\mu\nu} - m^2 A^\nu \implies B(x) = Be^{-x/m}$. Another consequence of the mass is that if we do manage to sneak some magnetic flux into a superconductor, the flux lines will bunch up into a localized string, as you'll show on the homework. This is called a vortex (or vortex string in $3d$) because of what $\Phi$ does in this configuration: its phase winds around the defect. In a superconductor, the role of $\Phi$ is played by the Cooper pair field (which has electric charge two). On the homework, you'll see a

[23] You can check that the mixing with $\theta$ is exactly what's required to make $\Pi(q)$ singular enough at $q = 0$ to give $A$ a mass consistent with the Ward identity, as in our discussion at (7.37).

consequence of the charge of $\Phi$ for the flux quantization of vortices. I hope to say more about its origins in terms of electrons later in §11.8.1.

I mention here the Meissner effect and the resulting collimation of flux lines partly because it will be helpful for developing a picture of confinement. In particular: think about the energetics of a magnetic monopole (suppose we had one available[24]) in a superconductor. If we try to insert it into a superconductor, it will trail behind it a vortex string along which all of its exiting magnetic flux is localized. This string has a finite tension (energy per unit length), as you'll study on the homework. If we make the superconducting region larger and larger, the energy of the monopole configuration grows linearly in the size – it is not a finite energy object in the thermodynamic limit. If monopoles were dynamical excitations of rest mass $M_m$, it would eventually become energetically favorable to pop an antimonopole out of the vacuum, so that the flux string connects the monopole to the antimonopole – this object can have finite energy inside the superconductor.

## 9.2 Festival of gauge invariance

Consider a collection of $N$ complex scalar fields (we could just as well consider spinors) with, for definiteness, an action of the form

$$\mathcal{L} = \sum_{\alpha=1}^{N} \partial_\mu \Phi^\star_\alpha \partial^\mu \Phi_\alpha - V(\Phi^\star_\alpha \Phi_\alpha) \tag{9.1}$$

(or $\mathcal{L} = \bar{\Psi}_\alpha \partial_\mu \Psi_\alpha$). The model actually has an $\mathsf{O}(2N)$ symmetry except that for kicks I grouped the scalars into pairs, and made the potential out of the combination $\sum_{\alpha=1}^{N} \Phi^\star_\alpha \Phi_\alpha$.

**Lighting review of Lie groups and Lie algebras.** (9.1) is invariant under the $\mathsf{U}(N)$ transformation

$$\Phi_\alpha \mapsto \Lambda_{\alpha\beta}\Phi_\beta, \quad \Lambda^\dagger \Lambda = \mathbb{1}. \tag{9.2}$$

Any such $\mathsf{U}(N)$ matrix $\Lambda$ can be parametrized as

$$\Lambda = \Lambda(\lambda) = e^{\mathbf{i}\sum_{A=1}^{N^2-1} \lambda^A T^A} e^{\mathbf{i}\lambda^0}$$

as we saw on the homework last quarter. $\lambda^0$ parametrizes a $\mathsf{U}(1)$ factor which commutes with everyone; we already know something about $\mathsf{U}(1)$ gauge theory from QED, so we won't focus on that. We'll focus on the non-abelian part: the $T^A$ are the generators of $\mathsf{SU}(N)$, and are traceless, so $\mathsf{SU}(N) \ni \Lambda(\lambda^0 = 0)$ has $\det \Lambda(\lambda^0 = 0) = 1$. Here the

[24] Here is the paper about the only one that's been by humans so far.

index $A = 1 : N^2 - 1 = \dim(\mathsf{SU}(N))$; the matrices $T^A$ (and hence also $\Lambda$) are $N \times N$, and satisfy the Lie algebra relations

$$[T^A, T^B] = \mathbf{i} f_{ABC} T^C \tag{9.3}$$

where $f_{ABC}$ are the structure constants of the algebra. For the case of $\mathsf{SU}(2)$, $T^A = \frac{1}{2}\sigma^A, A = 1, 2, 3$, and $f_{ABC} = \epsilon_{ABC}$. The infinitesimal version of (9.2), with $\Lambda$ close to the identity, is

$$\Phi_\alpha \mapsto \Phi_\alpha + \mathbf{i}\lambda^A T^A_{ab} \Phi_B.$$

The $N \times N$ representation is called the *fundamental representation* of $\mathsf{SU}(N)$. Other representations of the group come from other sets of $T^A_R$s which satisfy the same algebra (9.3), but can have other dimensions. For example, the structure constants themselves $\left(T^B_{\text{adj}}\right)_{AC} \equiv -\mathbf{i} f_{ABC}$ furnish the representation matrices for the adjoint representation.

**Local invariance.** The transformation above was *global* in the sense that the parameter $\lambda$ was independent of spacetime. This is an actual symmetry of the physical system associated with (9.1). Let's consider how we might change the model in (9.1) to make it invariant under a *local* transformation, with $\lambda = \lambda(x)$. In the Abelian case, we have learned

$$\Phi \mapsto e^{\mathbf{i}\lambda(x)}\Phi(x), A_\mu \mapsto A_\mu + \partial_\mu \lambda, \quad \partial_\mu \Phi \rightsquigarrow D_\mu \Phi = (\partial_\mu - \mathbf{i}A_\mu)\Phi \mapsto e^{\mathbf{i}\lambda(x)} D_\mu \Phi.$$

In words: by replacing partial derivatives with covariant derivatives, we can make gauge-invariant Lagrangians. The same thing works in the non-abelian case:

$$(D_\mu \Phi)_\alpha \equiv \partial_\mu \Phi_\alpha - \mathbf{i} A^A_\mu T^A_{\alpha\beta} \Phi_\beta$$

$$\Phi \mapsto \Phi + \mathbf{i}\lambda^A(x) T^A \Phi, \quad A^A_\mu \mapsto A^A_\mu + \partial_\mu \lambda^A - f_{ABC}\lambda^B A^C_\mu(x). \tag{9.4}$$

The difference is that there is a term depending on $A$ in the shift of the gauge field $A$. The following Yang-Mills Lagrangian density is a natural generalization of Maxwell:

$$\mathcal{L}_{YM} = -\frac{1}{4g^2} \sum_A \left( \underbrace{\partial_\mu A^A_\nu - \partial_\nu A^A_\mu + f_{ABC} A^B_\mu A^C_\nu}_{=F^A_{\mu\nu} = -F^A_{\nu\mu}} \right)^2 = -\frac{1}{4g^2} \mathrm{tr} F_{\mu\nu} F^{\mu\nu}. \tag{9.5}$$

The field strength

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} + f_{ABC}\lambda^B F^C_{\mu\nu} = F^A_{\mu\nu} + \mathbf{i}\lambda^B \left(T^B_{\text{adj}}\right)_{AC} F^C_{\mu\nu} \tag{9.6}$$

is designed so that it transforms in the adjoint representation, and therefore $S_{YM}$ is gauge-invariant. (Regarding $F$ as an $N \times N$ matrix $F = F^A T^A$, the finite version of (9.6) is $F \mapsto \Lambda F \Lambda^{-1}$, which makes it manifest that $\mathrm{tr}F^2$ is invariant.)

**Gauge fields as connections.** The preceding formulae are not too hard to verify, but where did they come from? Suppose we wanted to attach an $N$-dimensional complex vector space to each point in spacetime; on each vector space we have an action of $\mathsf{SU}(N)$, by $\Phi_\alpha(x) \mapsto \Lambda_{\alpha\beta}(x)\Phi(x)$. Suppose we would like to do physics in a way which is independent of the choice of basis for this space, at each point. We would like to be able to compare $\Phi(x)$ and $\Phi(y)$ (for example to make kinetic energy terms) in a way which respects these independent rotations. To do this, we need more structure: we need a *connection* (or *comparator*) $W_{xy}$, an object which transforms like $W_{xy} \mapsto \Lambda(x) W_{xy} \Lambda^{-1}(y)$, so that $\Phi^\dagger(x) W_{xy} \Phi(y)$ is invariant. The connection between two points $W_{xy}$ may depend on how we get from $x$ to $y$. We demand that $W(\emptyset) = \mathbb{1}$, $W(C_2 \circ C_1) = W(C_2)W(C_1)$ and $W(-C) = W^{-1}(C)$, where $-C$ is the path $C$ taken in the opposite direction.

But if we have a $W_{xy}$ for any two points, you can't stop me from considering nearby points and defining

$$D_\mu \Phi(x) \equiv \lim_{\Delta x \to 0} \frac{W(x, x+\Delta x)\Phi(x+\Delta x) - \Phi(x)}{\Delta x^\mu} \mapsto \Lambda(x) D_\mu \Phi(x) \ . \qquad (9.7)$$

Expanding near $\Delta x \to 0$, we can let

$$W(x, x+\Delta x) = \mathbb{1} - \mathbf{i} e \Delta x^\mu A_\mu(x) + \mathcal{O}(\Delta x^2) \qquad (9.8)$$

this defines the gauge field $A_\mu$ (sometimes also called the connection). To make the gauge transformation of the non-abelian connection field $A \mapsto A^\Lambda$ obvious, just remember that $D_\mu \Phi \mapsto D_\mu^{A^\Lambda}(\Lambda\Phi) \stackrel{!}{=} \Lambda\left(D_\mu^A \Phi\right)$ which means $A_\mu^\Lambda = \Lambda A_\mu \Lambda^{-1} - (\partial_\mu \Lambda)\Lambda^{-1}$. (This formula also works in the abelian case $\Lambda = e^{\mathbf{i}\lambda}$, and knows about the global structure of the group $\lambda \simeq \lambda + 2\pi$.)

The equation (9.8) can be integrated: $W_{xy} = e^{-\mathbf{i}e \int_{C_{xy}} A_\mu(\tilde{x}) d\tilde{x}^\mu}$ where $C_{xy}$ is a path in spacetime from $x$ to $y$. What if $\mathsf{G}$ is not abelian? Then I need to tell you the ordering in the exponent. We know from Dyson's equation that the solution is

$$W_{xy} = \mathcal{P} e^{-\mathbf{i}e \int_{C_{xy}} A_\mu(\tilde{x}) d\tilde{x}^\mu}$$

where $\mathcal{P}$ indicates *path-ordering* along the path $C_{xy}$, just like the time-ordered exponential we encountered in interaction-picture perturbation theory.

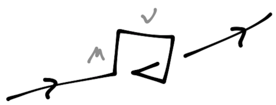

[25]

To what extent *does* $W_{xy}$ depend on the path? In the abelian case,

$$W_C = W_{C'} e^{\mathbf{i}e \oint_{C-C'} A} \stackrel{\text{Stokes}}{=} W_{C'} e^{\mathbf{i}e \int_R F_{\mu\nu} dx^\mu dx^\nu}$$

where $\partial R = C - C'$ is a 2d surface whose boundary is the difference of paths.

[25] Which 2d surface? Let me speak about the abelian case for the rest of this footnote. The difference



Imagine inserting an infinitesimal rectangle to the path which moves by $dx^\mu$ then by $dx^\nu$ and then back and back. The difference in the action on $\Phi$ is

$$dx^\mu dx^\nu [D_\mu, D_\nu]\Phi = -\mathbf{i}e dx^\mu dx^\nu F_{\mu\nu}\Phi.$$

The commutator of covariant derivatives is not an operator, but a function $[D_\mu, D_\nu] = -\mathbf{i}eF_{\mu\nu}$. (Note that this same maneuver defines the Riemann tensor in terms of derivatives covariant with respect to coordinate changes.) This same relation holds in the non-abelian case:

$$F_{\mu\nu} = \frac{\mathbf{i}}{e}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - \mathbf{i}e[A_\mu, A_\nu].$$

This object is Lie-algebra-valued, so can be expanded in a basis: $F_{\mu\nu} = F^A_{\mu\nu}T^A$, so more explicitly,

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - \mathbf{i}e f_{ABC} A^B_\mu A^C_\nu.$$

Since it is made from products of covariant derivatives, $[D, D]\Phi \mapsto \Lambda[D, D]\Phi$, it must transform in the adjoint representation, $F \mapsto \Lambda F \Lambda^{-1}$, which in infinitesimal form returns us to (9.6)

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} - f^{ABC}\lambda^B F^C_{\mu\nu}.$$

[End of Lecture 30]

## 9.3 Interlude on differential forms and algebraic topology

[Zee section IV.4] We interrupt this physics discussion with a message from our mathematical underpinnings. This is nothing fancy, mostly just some book-keeping. It's some notation that we'll find useful, which I found it rather inhibiting not to be able to use in the last lecture. As a small payoff we can define some simple topological invariants of smooth manifolds.

Suppose we are given a smooth manifold $X$ on which we can do calculus. For now, we don't even need a metric on $X$. Suppose $x^\mu$ are some local coordinates on $X$.

A $p$-form on $X$ is a completely antisymmetric $p$-index tensor,

$$A \equiv \frac{1}{p!} A_{m_1 \dots m_p} \mathrm{d}x^{m_1} \wedge \dots \wedge \mathrm{d}x^{m_p}.$$

in phase between two possible choices is $e^{\mathbf{i}e\int_{R-R'} F} \overset{\text{Stokes}}{=} e^{\mathbf{i}e\int_V dF}$ where $\partial V = R - R'$ is the 3-volume whose boundary is the difference of the two regions. The integrand vanishes by the Bianchi identity, which is actually an identity if $F = dA$. You might think this prevents magnetic sources, which appear on the RHS of the Maxwell equation $dF = \star j_m$. But actually $\int_V dF$ only appears in the combination $e^{\mathbf{i}e\int_V dF}$, so magnetic sources are perfectly consistent with independence of the choice of $R$, as long as their charge $q \equiv \int_V dF = \oint_{\partial V} F$ is quantized $ge \in 2\pi\mathbb{Z}$. This is Dirac quantization.

The coordinate one-forms are fermionic objects in the sense that $\mathrm{d}x^{m_1} \wedge \mathrm{d}x^{m_2} = -\mathrm{d}x^{m_2} \wedge \mathrm{d}x^{m_1}$ and $(\mathrm{d}x)^2 = 0$. The point in life of a $p$-form is that it can be integrated over a $p$-dimensional space. The order of its indices keeps track of the orientation (and it saves us the trouble of writing them). It is a geometric object, in the sense that it is something that can be (wants to be) integrated over a p-dimensional subspace of $X$, and its integral will only depend on the subspace, not on the coordinates we use to describe it.

Familiar examples include the gauge potential $A = A_\mu \mathrm{d}x^\mu$, and its field strength $F = \frac{1}{2} F_{\mu\nu} \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$. Given a curve $C$ in $X$ parameterized as $x^\mu(s)$, we have

$$\int_C A \equiv \int_C dx^\mu A_\mu(x) = \int ds \frac{dx^\mu}{ds} A_\mu(x(s))$$

and this would be the same if we chose some other parameterization or some other local coordinates.

The *wedge product* of a $p$-form $A$ and a $q$-form $B$ is a $p+q$ form

$$A \wedge B = A_{m_1..m_p} B_{m_{p+1}...m_{p+q}} \mathrm{d}x^{m_1} \wedge ... \wedge \mathrm{d}x^{m_{p+q}} \ ,$$

[26] The space of $p$-forms on a manifold $X$ is sometimes denoted $\Omega^p(X)$, especially when it is to be regarded as a vector space (let's say over $\mathbb{R}$).

The *exterior derivative* d acts on forms as

$$\begin{aligned} \mathrm{d} : \Omega^p(X) &\to \Omega^{p+1}(X) \\ A &\mapsto \mathrm{d}A \end{aligned}$$

by

$$\mathrm{d}A = \frac{1}{p!} \partial_{m_1} (A)_{m_2...m_{p+1}} \mathrm{d}x^{m_1} \wedge ... \wedge \mathrm{d}x^{m_{p+1}}.$$

You can check that

$$\mathrm{d}^2 = 0$$

basically because derivatives commute. Notice that $F = \mathrm{d}A$ in the example above. Denoting the boundary of a region $D$ by $\partial D$, Stokes' theorem is

$$\int_D \mathrm{d}\alpha = \int_{\partial D} \alpha.$$

[26]The components of $A \wedge B$ are then

$$(A \wedge B)_{m_1...m_{p+q}} = \frac{(p+q)!}{p!q!} A_{[m_1...m_p} B_{m_{p+1}...m_{p+q}]}$$

where [..] means sum over permutations with a $-1$ for odd permutations. Try not to get caught up in the numerical prefactors. In my expression below for the exterior derivative also there is annoying combinatoric prefactor.

And notice that $\Omega^{p>\dim(X)}(X) = 0$ – there are no forms of rank larger than the dimension of the space.

A form $\omega_p$ is *closed* if it is killed by d: $\mathrm{d}\omega_p = 0$.

A form $\omega_p$ is *exact* if it is d of something: $\omega_p = \mathrm{d}\alpha_{p-1}$. That something must be a $(p-1)$-form.

---

Because of the property $\mathrm{d}^2 = 0$, it is possible to define *cohomology* – the image of one $\mathrm{d} : \Omega^p \to \Omega^{p+1}$ is in the kernel of the next $\mathrm{d} : \Omega^{p+1} \to \Omega^{p+2}$ (*i.e.* the $\Omega^p$s form a *chain complex*). The $p$th de Rham cohomology group of the space $X$ is defined to be

$$H^p(X) \equiv \frac{\text{closed } p\text{-forms on } X}{\text{exact } p\text{-forms on } X} = \frac{\ker(\mathrm{d}) \in \Omega^p}{\mathrm{Im}\,(\mathrm{d}) \in \Omega^p}.$$

That is, two closed $p$-forms are equivalent in cohomology if they differ by an exact form:

$$[\omega_p] - [\omega_p + \mathrm{d}\alpha_{p-1}] = 0 \in H^p(X),$$

where $[\omega_p]$ denotes the equivalence class. The dimension of this group is $b^p \equiv \dim H^p(X)$ called the $p$th betti number and is a topological invariant of $X$. The euler characteristic of $X$, which you can also get by triangulating $X$ and counting edges and faces and stuff, is

$$\chi(X) = \sum_{p=0}^{d=\dim(X)} (-1)^p b^p(X).$$

Here's a very simple example, where $X = S^1$ is a circle. $x \simeq x + 2\pi$ is a coordinate; the radius will not matter since it can be varied continuously. An element of $\Omega^0(S^1)$ is a smooth *periodic* function of $x$. An element of $\Omega^1(S^1)$ is of the form $A_1(x)dx$ where $A_1$ is a smooth periodic function. Every such element is closed because there are no 2-forms on a 1d space. The exterior derivative on a 0-form is

$$\mathrm{d}A_0(x) = A_0' dx$$

Which 0-forms are closed? $A_0' = 0$ means $A_0$ is a constant. Which 1-forms can we make this way? The only one we can't make is $dx$ itself, because $x$ is not a periodic function. Therefore $b^0(S^1) = b^1(S^1) = 1$.

---

Now suppose we have a volume element on $X$, *i.e.* a way of integrating $d$-forms. This is guaranteed if we have a metric, since then we can integrate $\int \sqrt{\det g}...$, but is less structure. Given a volume form, we can define the Hodge star operation $\star$ which maps a $p$-form into a $(d-p)$-form:

$$\star : \Omega^p \to \Omega^{d-p}$$

by

$$\left(\star A^{(p)}\right)_{\mu_1 \cdots \mu_{d-p}} \equiv \epsilon_{\mu_1 \cdots \mu_d} A^{(p)\ \mu_{d-p+1} \cdots \mu_d}$$

An application: consider the Maxwell action, $\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$. You can show that this is the same as $S[A] = \int F \wedge \star F$. (Don't trust my numerical prefactor.) You can derive the Maxwell EOM by $0 = \frac{\delta S}{\delta A}$. $\int F \wedge F$ is the $\theta$ term. The magnetic dual field strength is $\tilde{F} = \star F$. Many generalizations of duality can be written naturally using the Hodge $\star$ operation.

---

As you can see from the Maxwell example, the Hodge star gives an inner product on $\Omega^p$: for two $p$-forms $\alpha, \beta$ $(\alpha, \beta) = \int \alpha \wedge \star \beta)$, $(\alpha, \alpha) \geq 0$. We can define the adjoint of d with respect to the inner product by

$$\int \mathrm{d}^\dagger \alpha \wedge \star \beta = (\mathrm{d}^\dagger \alpha, \beta) \equiv (\alpha, \mathrm{d}\beta) = \int \alpha \wedge \star \mathrm{d}\beta$$

Combining this relation with integration by parts, we find $\mathrm{d}^\dagger = \pm \star \mathrm{d} \star$.

We can make a Laplacian on forms by

$$\Delta = \mathrm{d}\mathrm{d}^\dagger + \mathrm{d}^\dagger \mathrm{d}.$$

This is a supersymmetry algebra, in the sense that $\mathrm{d}, \mathrm{d}^\dagger$ are grassmann operators.

Any cohomology class $[\omega]$ has a *harmonic* representative, $[\omega] = [\tilde{\omega}]$ where in addition to being closed $\mathrm{d}\omega = \mathrm{d}\tilde{\omega} = 0$, it is *co-closed*, $0 = \mathrm{d}^\dagger \tilde{\omega}$, and hence *harmonic* $\Delta \tilde{\omega} = 0$.

An application of this is Poincare duality: $b^p(X) = b^{d-p}(X)$ if $X$ has a volume form. This follows because the map $H^p \to H^{d-p}$ $[\omega_p] \mapsto [\star \omega_p]$ is an isomorphism. (Choose the harmonic representative, it has $\mathrm{d} \star \tilde{\omega}_p = 0$.)

The de Rham complex of $X$ can be realized as the groundstates of a physical system, namely the supersymmetric nonlinear sigma model with target space $X$. The fermions play the role of the $dx^\mu$s. The states are of the form

$$|A\rangle = \sum_{p=1}^{d} A_{\mu_1 \cdots \mu_p}(x) \psi^{\mu_1} \psi^{\mu_2} \cdots \psi^{\mu_p} |0\rangle$$

where $\psi$ are some fermion creation operators. This shows that the hilbert space is the space of forms on $X$, that is $\mathcal{H} \simeq \Omega(X) = \oplus_p \Omega^p(X)$. The supercharges act like d and

$\mathrm{d}^\dagger$ and therefore the supersymmetric groundstates are (harmonic representatives of) cohomology classes.

---

The machinery of differential forms is very useful.

## 9.4 Actions for gauge fields

The Yang-Mills (YM) action (9.5) is a gauge invariant and Lorentz invariant local functional of $A$. If the gauge field is to appear in $D = \partial + A$ it must have the same dimension as $\partial$, so $\mathcal{L}_{YM}$ has naive scaling dimension 4, like the Maxwell term, so it is marginal in $D = 4$. Notice that unlike the Maxwell term, $\mathcal{L}_{YM}$ is not quadratic in $A$: it contains cubic and quartic terms in $A$, whose form is determined by the gauge algebra $f_{ABC}$. Non-abelian gauge fields interact with themselves in a very definite way.

In even spacetime dimensions, another gauge invariant, Lorentz invariant local functional of $A$ is the total-derivative term $S_\theta = \theta \int \mathrm{tr} \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $D/2$ factors of $F$. This doesn't affect the equations of motion or perturbation theory (*e.g.* in $D = 4$, in the abelian case, $\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = 2\partial_\mu \left(\epsilon^{\mu\nu\rho\sigma} A_\nu F_{\rho\sigma}\right)$) but does matter non-perturbatively. We'll see (when we study anomalies) that for smooth gauge field configurations in a closed spacetime, this functional is an integer. This coupling of the gluons is constrained to be very small because it would give an electric dipole moment to the neutron, which the neutron doesn't seem to have; this mystery is called the strong CP problem because this coupling $\theta$ violates CP symmetry (notice that $F \wedge F$ has one time derivative and three spatial derivatives).

In odd spacetime dimensions, we should consider the Chern-Simons term (the $D = 2 + 1$ version of which we just encountered) $\int A \wedge \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $(D-1)/2$ factors of $F$. (In the non-Abelian case, there is an extra term: in 3d, $S_{CS}[A] \propto \int \mathrm{tr} \left(A \wedge F + \frac{2}{3} A \wedge A \wedge A\right)$.) This term *does* affect the equations of motion. It breaks parity symmetry. It is important in quantum Hall physics in $D = 2+1$, where it gives the gauge field fluctuations a mass.

In general dimension, we can make more couplings out of just $A$ if we take more derivatives, but they will have higher dimension.

We can couple YM gauge fields to matter by returning to our starting point: *e.g.* if $\psi(x) \mapsto \Lambda_R \psi(x)$ is a Dirac field transforming in some representation $R$ of the gauge group, then $D_\mu \psi = \left(\partial_\mu - \mathbf{i} T_R^A A_\mu^A\right) \psi$ also transforms in representation $R$, so

$$\bar{\psi} \gamma^\mu D_\mu \psi + V(\bar{\psi}\psi)$$

is a gauge-invariant lagrangian density. The lowest-dimension couplings of $A$ to matter are determined by the representation matrices $T_R^A$, which generalize the electric charge.

You might expect that we would starting doing perturbation theory in $g$ now. There is lots of physics there, but it takes a little while to get there. Given how limited our time is this quarter, we will instead think about how we might define the thing non-perturbatively and see what we learn from that.

## 9.5 Lattice gauge theory

The following beautiful construction was found by Wegner and Wilson and Polyakov; a good review is this one by Kogut.

Consider discretizing euclidean spacetime into a hypercubic lattice (for simplicity). On each *link* $xy$ of the lattice we place a $\mathsf{G}$-valued matrix $U^{ab}_{xy}$. We demand that $U_{yx} = U^{-1}_{xy}$. Three good examples to keep in mind (in decreasing order of difficulty) are:

1. $\mathsf{G} = \mathsf{U}(N)$, in which case each $U$ is a complex $N \times N$ matrix with $UU^\dagger = 1$.
2. $\mathsf{G} = \mathsf{U}(1)$, in which case $U$ is a phase, $U_{xy} = e^{\mathbf{i}\theta_{xy}}$, $\theta_{xy} \in [0, 2\pi)$.
3. $\mathsf{G} = \mathbb{Z}_n$, in which case $U = e^{2\pi \mathbf{i}\ell/n}, \ell = 1, \cdots n$, is a phase with $U^n = 1$. For $n = 2$, this is a classical spin.

Please think of $U_{xy} = \mathcal{P}e^{\mathbf{i}\int_x^y A_\mu(r)dr^\mu}$ as the Wilson line along the link (except that there is no such thing as $A_\mu(r)$ at other values of $r$). As such, we impose the gauge equivalence relation $U_{xy} \mapsto g^\dagger_x U_{xy} g_y$, where $g_x \in \mathsf{G}$ for each $x$. We will accomplish this by two steps: by writing an action $S[U]$ which has this invariance, and by integrating over $\{U\}$ with an invariant measure:

$$Z = \int \prod_\ell dU_\ell e^{-S[U]}.$$

Here $\int dU$ is the $\mathsf{G}$-invariant (Haar) measure on $\mathsf{G}$, which can be defined by the desiderata

$$\int_{\mathsf{G}} dU = 1, \quad \int_{\mathsf{G}} dU f(U) = \int_{\mathsf{G}} dU f(VU) = \int_{\mathsf{G}} dU f(UV), \forall V \in \mathsf{G}\ .$$

For $\mathsf{G} = \mathsf{U}(1)$, it is just $\int_0^{2\pi} đ\varphi$; for $\mathsf{G} = \mathbb{Z}_n$, it just $\sum_{\ell=1}^n$. You can figure out what it is for $\mathsf{SU}(2)$ (locally, it's the round measure on $S^3$). Notice the following lovely advantage of these conditions: there is no need to gauge fix anything.

This is a statistical mechanics problem of the thermodynamics of a bunch of classical rotors (slightly fancy ones in the $\mathsf{SU}(N)$ case). The review by Kogut does a great job of highlighting the fact that this class of problems is susceptible to all the tools of statistical mechanics.

What action should we use? Here is a good way to make something invariant under the gauge group: Consider the comparator for a *closed* path $C_{xx}$ which starts at $x$ and ends at $x$:

$$W(C_{xx}) = \mathcal{P}e^{\mathbf{i}\int_{C_{xx}} A}.$$

How does this transform? $W(C_{xx}) \mapsto g_x^{-1} W(C_{xx}) g_x$, but, for non-abelian $\mathsf{G}$, it's still a matrix! A gauge-*in*variant object is

$$W(C) \equiv \mathrm{tr} W(C_{xx}) = \mathrm{tr}\mathcal{P}e^{\mathbf{i}\int_{C_{xx}} A}$$

where the $g_x$ and $g_x^{-1}$ can eat each other by cylicity of the trace. We can make something gauge invariant and as local as possible by considering a path $C$ which goes around a single plaquette of the lattice: $C = \partial\Box$. This is Wilson's action:

$$S[U] = \frac{1}{2f^2}\sum_{\Box} S_{\Box}, \quad S_{\Box} \equiv \mathrm{Re} W(\partial\Box) = \mathrm{Re}\,\mathrm{tr}\prod_{\ell\in\partial\Box} U_\ell = \mathrm{Re}\,\mathrm{tr}\left(U_{x,x+dx}U_{x+dx,x+dx+dy}U_{x+dx+dy,x+dy}U_{x+dy,x}\right).$$

Now let's focus on he $\mathsf{G} = \mathsf{SU}(N)$ case, and take seriously the idea that $U_{x,x+dx} = e^{-\mathbf{i}\int_x^{x+dx} A_\mu dx^\mu}$, where $A_\mu(x)$ is an element of the Lie algebra $\mathsf{su}(N)$. An application of the CBH[27] formula $e^{sA}e^{sB} = e^{sA+sB+\frac{s^2}{2}[A,B]+\mathcal{O}(s^3)}$ shows that for a plaquette oriented in the $\mu\nu$ plane $\Box_{\mu\nu}$, with lattice spacing $a$,

$$\begin{aligned}
S_{\Box_{\mu\nu}} &\overset{\mathrm{CBH}}{=} \mathrm{Re}\,\mathrm{tr}\left(e^{\mathbf{i}\frac{a^2}{\sqrt{2}g^2}F_{\mu\nu}} + \mathcal{O}(a^3)\right) \\
&= \mathrm{Re}\,\mathrm{tr}\left(\mathbb{1} + \mathbf{i}\frac{a^2}{\sqrt{2}g^2}F_{\mu\nu} - \frac{1}{2}\frac{a^4}{4g^2}F_{\mu\nu}F_{\mu\nu} + \mathcal{O}(a^5)\right) \\
&= \mathrm{tr}\mathbb{1} - \frac{a^4}{4g^4}\mathrm{tr}F_{\mu\nu}F_{\mu\nu} + \ldots = \mathcal{L}_{YM}(\Box) + \mathrm{const.}
\end{aligned}$$

The coupling $g$ is related to $f$ in some way that can be figured out. So it is plausible that this model has a continuum limit governed by the Yang-Mills action. Realizing this possibility requires that the model defined by $Z$ have a correlation length much larger than the lattice spacing, which is a physics question.

Before examining the partition sum, how would we add charged matter? If we place fundamentals $q_x \mapsto g_x q_x$ at each site, we can make gauge invariants of the form $q_x^\dagger U_{xy}U_{yz}U_{zw}q_w$, or most simply, we can make a kinetic term for $q$ by

$$S_q = \frac{1}{a^{\#}}\sum_{x,\ell} q_x^\dagger U_{x,x+\ell} q_{x+\ell} \simeq \int d^Dx\; q^\dagger(x)\left(\not{D} - m\right) q(x) + \ldots$$

[27]Charlie-Baker-Hotel? Campbell-Baker-Hausdorff.

where $D_\mu = \partial_\mu - \mathbf{i}A_\mu$ is the covariant derivative, and we used its definition (9.7). The expression I've written is for a grassmann, spinor field; for bosonic fields the second-order terms are the leading terms which aren't a total derivative. There is some drama about the number of components of the spinor field one gets. It is not hard to get a massive Dirac fermion charged under a $\mathsf{U}(1)$ gauge field, like in QED. It is impossible to get a chiral spectrum, like a single Weyl fermion, from a gaussian, local lattice action; this is called the Nielsen-Ninomiya theorem. You might think 'oh that's not a problem, because in the Standard Model there is the same number of L and R Weyl fermions,' but it is still a problem because they carry different representations under the electroweak gauge group. The word 'gaussian' is a real loophole, but not an easy one.

How do we get physics from the lattice gauge theory path integral $Z$? We need to find some gauge-invariant observables (since anything we stick in the integrand that isn't gauge-invariant will average to zero). In the pure YM theory, a good one is our friend the Wilson loop $W(C) = \mathrm{tr}\left(\prod_{\ell\in C} U_\ell\right) \simeq \mathrm{tr}\mathcal{P}e^{\mathbf{i}\oint_C A}$. What physics does it encode? Recall what happened when we added an external source to measure the force mediated by various fields, for example in the Maxwell theory:

$$\lim_{T\to\infty} Z^{-1}\int DA\; e^{\mathbf{i}S_{\text{Maxwell}}[A]+\mathbf{i}\int A_\mu J^\mu} = e^{-\mathbf{i}V(R)T}.$$

Here we took $J^\mu(x) = \eta^{\mu 0}\left(\delta^d(\vec{x}) - \delta^d(\vec{x}-(R,0,0))\right)$ for $t$ in an interval of duration $T$, and zero before and after, two charges are held at distance $R$ for a time $T$. $V(R)$ is the energy of the resulting configuration of (here, electromagnetic) fields, *i.e.* the Coulomb potential. If instead we let the charge and anticharge annihilate at $t = 0$ and $t = T$, this is a single charge moving along a rectangular loop $C_{R\times T}$ in spacetime, with sides $R$ and $T$, and the result is just the expectation value of the associated Wilson loop. Going back to Euclidean spacetime, this is

$$\langle W(C_{R\times T})\rangle = Z^{-1}\int \prod dU\; e^{-\frac{1}{2f^2}\sum_\Box \mathrm{Re}\, S_\Box} W(C_{R\times T}) \overset{T\gg R}{\simeq} e^{-V(R)T},$$

where the LHS is the expectation value of a gauge invariant operator. There can be some funny business associated with the corners and the spacelike segments, and this is the reason that we look for the bit of the free energy which is extensive in $T$.

In the case of the Maxwell theory in the continuum, this is a gaussian integral, which we can do (see the homework), and $\log\left\langle e^{\mathbf{i}\oint_{C_{R\times T}} A}\right\rangle \simeq -E(R)T - f(T)R$ with $E(R)\sim\frac{1}{R}$, goes something like the *perimeter* of the loop $C$. In the case of a short-ranged interaction, from a massive gauge field, the perimeter law would be more literally satisfied.

In contrast, a confining force between the charges would obtain if $\langle W(C_{R\times T})\rangle \overset{T\gg R}{\simeq} e^{-V(R)T}$ with instead

$$V(R) = \sigma R \quad \Longrightarrow F = -\frac{\partial V}{\partial R} = -\sigma \ .$$

This is a distance-independent attractive force between the charges. In this case $\log \langle W\rangle \sim RT$ goes like the *area* of the (inside of the) loop, so confinement is associated with an area law for Wilson loops. A constant force means a linear potential, so it is as if the charges are connected by a string of constant tension (energy per unit length) $\sigma$.

A small warning about the area law: in general, the existence of an area law may depend on the representation in which we put the external charges:

$$W(C,R) = \mathrm{tr}_R \mathcal{P} e^{\mathbf{i}\oint_C A^A T^A_R}$$

where $T^A_R$ are the generators of $\mathsf{G}$ in some representation $R$; this is the phase associated with a (very heavy and hence non-dynamical) particle in representation $R$. For some choices of $R$, it might be possible and energetically favorable for the vacuum to pop out *dynamical* charges which then screen the force between the two external charges (by forming singlets with them). $\mathsf{G} = \mathsf{SU}(N)$ has a center $\mathbb{Z}_N \subset \mathsf{SU}(N)$ under which the adjoint is neutral, so a Wilson loop in a representation carrying $\mathbb{Z}_N$ charge (such as the fundamental, in which it acts by $\mathbb{Z}_N$ phases times the identity) cannot be screened by pure glue. QCD, which has dynamical fundamentals, is more subtle.

This point, however, motivates the study of the dynamics of lattice gauge theories to address the present question: Where might such an area law come from? I'll give two hints for how to think about it.

**Hint 1: Strong coupling expansion.** In thinking about an integral of the form

$$\int DU \ e^{\beta \sum_\Box S_\Box} W(C)$$

it is hard to resist trying to expand the exponential in $\beta$.

Unlike the perturbation series we've been talking about for months, this series has a finite radius of convergence. To understand this, it is useful to recognize that this expansion is structurally identical to the high-temperature expansion of a thermal partition function. For each configuration $C$, the function $e^{-\beta h(C)}$ is analytic in $\beta$ about $\beta = 0$ (notice that $e^{-\frac{1}{T}}$ is analytic about $T = \infty$!). The only way to get a singularity at $\beta = 0$ would be if the sum over configurations (in the thermodynamic limit) did it; this would be a phase transition at $T = \infty$; that doesn't happen because the correlation length inevitably goes to zero at $T = \infty$: every site is so busy being buffeted by thermal fluctuations that it doesn't care about the other sites at all.[28]

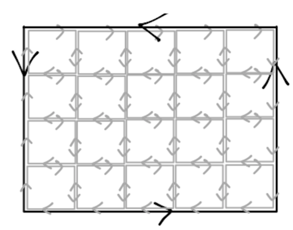
In the non-abelian case, we get to do all kinds of fun stuff with characters of the group. For simplicity, let's focus on an abelian example, which will have a similar structure (though different large-$\beta$ (weak coupling) physics). So take $U_\ell = e^{\mathbf{i}\theta_\ell} \in \mathsf{U}(1)$, in which case

$$S_{\Box_{\mu\nu}}[U] = -\left(1 - \cos\theta_{\mu\nu}\right), \quad \theta_{\mu\nu}(x) = \theta_\mu(x+\nu) - \theta_\mu(x) - \theta_\nu(x+\mu) + \theta_\nu(x) \equiv \Delta_\nu\theta_\mu - \Delta_\mu\theta_\nu(x).$$

First let's consider the case where the world is a single plaquette. Then, using the identity $\int_0^{2\pi} đ\theta \, e^{\mathbf{i}n\theta} = \delta_{n,0}$,

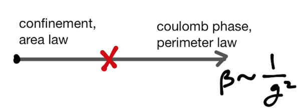
$$\begin{aligned}\langle W(\Box)\rangle &= \int \prod_\ell dU_\ell \; U_1U_2U_3U_4 \left(1 + \beta(S_\Box + S_\Box^\dagger) + \frac{1}{2}\beta^2\left(S + S^\dagger\right)^2 + \frac{1}{3!}\beta^3\left(S + S^\dagger\right)^3 + \cdots\right) \\ &= \beta \underbrace{\langle S_\Box S_{-\Box}\rangle}_{=1} + \frac{\beta^3}{2}\langle S_{2\Box}S_{-2\Box}\rangle + \mathcal{O}(\beta^5) = \beta^{A(\Box)}\left(1 + \mathcal{O}(\beta^2)\right) = e^{-f(\beta)\text{Area}} \qquad (9.9)\end{aligned}$$

with $f(\beta) = |\ln\beta|$ in this crude approximation. Here the area of the loop was just 1. I've written $S_{2\Box} = S_\Box^2$, which is only true in abelian cases.



If instead we consider a loop which encloses many plaquettes, we must pull down at least one factor of $\beta S_\Box^\dagger$ for each plaquette, in order to cancel the link factors in the integrand. We can get more factors of beta if we pull down more cancelling pairs of $\beta^n S_\Box^n S_{-\Box}^n$, but these terms are subleading at small $\beta$. The leading contribution is $\langle W(C)\rangle = e^{-f(\beta)\text{Area}}\left(1 + \mathcal{O}(\beta^2)\right)$, an area law.

Since the series converges, this conclusion can be made completely rigorous. In what sense is confinement a mystery then? Well, a hint is that our argument applies equally well (and in fact the calculation we did was) for abelian gauge theory! But QED doesn't confine – we calculated the Wilson loop at weak coupling and found a perimeter law – what gives?




The answer is that there is a phase transition in between weak and strong coupling, so weak coupling is not an analytic continuation of the strong coupling series answer. Ruling out this possibility in Yang-Mills theory would be lucrative.

[End of Lecture 31]

In fact, though, the Wilson loop expectation itself can exhibit a phase transition, even if other observables don't. I've drawn the pictures above as if the world were two-

[28]For a much more formal and, I think, less illuminating proof, see for example J-M Drouffe and J-B Zuber, Physics Reports **102** (1983) section 3.1.2. Thanks to Tarun Grover for framing the above argument.

dimensional, in which case we just cover every plaquette inside the loop. In $D > 2$, we have to choose a surface whose boundary is the loop. Rather, $\langle W \rangle$ is a statistical sum over such surfaces, weighted by $\beta^{\text{area}}$. Such surface models often exhibit a *roughening transition* as $\beta$ becomes larger and floppy surfaces are not suppressed.

By the way, the same technology can be used to study the spectrum of excitations of the gauge theory, by considering correlations like

$$\left\langle S_R(t) S_R^\dagger(0) \right\rangle_c = \sum_\alpha |c_\alpha^R|^2 e^{-m_\alpha(R)t}$$


where $S_R$ is the trace of a Wilson loop in representation $R$, around a single plaquette, and the two loops in question are separated only in time and are parallel. The subscript $c$ means connected.
The right hand side is a sum over intermediate, gauge invariant states with the right quantum numbers, and $m_\alpha(R)$ are their masses. This is obtained by inserting a complete set of energy eigenstates.

In strong coupling expansion, we get a sum over discretized tubes of plaquettes, with one boundary at each loop (the connected condition prevents disconnected surfaces), the minimal number of plaquettes for a hypercubic lattice is $4t$,

giving

$$\left\langle S_R(t) S_R^\dagger(0) \right\rangle_c \sim A\beta^{4t}\left(1 + \mathcal{O}(\beta^2)\right)$$

and the smallest glueball mass becomes $m_0 \sim 4|\ln\beta|$, similar to the scale of the string tension. Actually, the corrections exponentiate to give something of the form $m_0(R) = -4\ln\beta + \sum_k m_k(R)\beta^k$.

**Hint 2: monopole condensation and dual Meissner effect.**

[Banks' book has a very nice discussion of this.] Recall that a single magnetic monopole is not a finite energy situation inside an infinite superconductor, because it has a tensionful Abrikosov flux string attached to it. A monopole and an antimonopole are linearly confined, with a constant force equal to the string tension.

On the other hand, electric-magnetic duality is a familiar invariance of Maxwell's equations:

$$\partial^\mu F_{\mu\nu} = J_\nu^{(e)}, \partial^\mu \tilde{F}_{\mu\nu} = J_\nu^{(m)} \tag{9.10}$$

is invariant under the replacements

$$F_{\mu\nu} \to \tilde{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}, \quad J_\nu^{(e)} \to J_\nu^{(m)}.$$

In doing a weak-coupling expansion (*e.g.* as we did in QED), we make a choice (having not seen magnetic charges, they must be heavy) to *solve* the second equation of (9.10) by introducing a smooth vector potential $A_\mu$ via

$$F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu + \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\int d^4 y J^{(m)}(y)^\sigma f^\rho(x-y)$$

with $\partial_\rho f^\rho(x) = \delta^4(x)$. Here we are treating the magnetic sources as fixed, *e.g.* because they are heavy. The support of the function $f^\rho$ is called the *Dirac string*. A monopole is placed at the end of a long and infinitely thin solenoid, which carries away its magnetic flux $\int_{\text{sphere around monopole}} B = \int_{\text{cross-section of solenoid}} B = g$, and is invisible classically. Quantumly, it could be detected by Aharonov-Bohm effect of a charged particle going around it $e^{\mathbf{i}e\oint A} = e^{\mathbf{i}e\int B} = e^{\mathbf{i}eg}$ unless $eg \in 2\pi\mathbb{Z}$, Dirac quantization again. (For particles with both electric and magnetic charge (they are called dyons), the condition is $q_1 m_2 - q_2 m_1 \in 2\pi\mathbb{Z}$.)

So, the duality interchanges electric and magnetic things. So, if condensation of electric charge (meaning $\langle\Phi\rangle = v$ for some electrically charged field $\Phi$) means that $A_\mu$ is massive (Higgs effect) and that monopoles are confined by tensionful magnetic flux tubes, then we can just replace the relevant words to learn that: Condensation of magnetic charge $\langle\Phi_m\rangle \neq 0$ means that some dual photon ($\tilde{A}_\mu$ with $d\tilde{A} = \tilde{F}$) is massive, and that *electric* charges are linearly confined by tensionful *electric* flux tubes.

This was pointed out by Mandelstam and 't Hooft in 1974. In 1994 Seiberg and Witten (hep-th/9407087) showed in detail that this happens in a highly supersymmetric example. In abelian lattice models, we can actually implement the duality transformation explicitly by various path integral tricks. One path through this story (found in 1978 by Banks, Myerson, Kogut and also Peskin) is described in Banks' book. Along the way, one encounters dualities with many familiar statistical mechanical models, such as the XY model.

# 10 Anomalies and fermion path integrals

## 10.1 Coherent state path integrals for fermions

We'll need these for our discussion of anomalies, and they are extremely useful for doing perturbative QCD (which differs from Yang-Mills theory by the addition of fermionic quarks).

[Peskin, §95, Shankar, *Principles of QM*, path integrals revisited. In this chapter of his great QM textbook, Shankar sneaks in lots of insights useful for modern condensed matter physics.]

Consider the algebra of a single fermion mode operator[29]:

$$\{\mathbf{c}, \mathbf{c}\} = 0, \quad \{\mathbf{c}^\dagger, \mathbf{c}^\dagger\} = 0, \quad \{\mathbf{c}, \mathbf{c}^\dagger\} = 1 \ .$$

With a single mode, the general Hamiltonian is

$$\mathbf{H} = \mathbf{c}^\dagger \mathbf{c} \left(\omega_0 - \mu\right)$$

($\omega_0$ and $\mu$ are (redundant when there is only one mode) constants). This algebra is represented on a two-state system $|1\rangle = \mathbf{c}^\dagger |0\rangle$. We might be interested in its thermal partition function

$$Z = \mathrm{tr}\, e^{-\frac{\mathbf{H}}{T}} \ .$$

(In this example, it happens to equal $Z = 1 + e^{-\frac{\omega_0 - \mu}{T}}$, as you can see by computing the trace in the eigenbasis of $\mathbf{n} = \mathbf{c}^\dagger \mathbf{c}$. But never mind that; the one mode is a proxy for many, where it's not quite so easy to sum.) How do we write this trace as a path integral? We can do this by insertion lots of resolutions of the identity (this is sometimes called 'Trotterizing'), using any resolution of the identity on $\mathcal{H}$, so there can be many very-different-looking answers to this question.

Let's define coherent states for fermionic operators:

$$\mathbf{c} |\psi\rangle = \psi |\psi\rangle \ . \tag{10.1}$$

Here $\psi$ is a c-number (not an operator), but acting twice with $\mathbf{c}$ we see that we must have $\psi^2 = 0$. So $\psi$ is a grassmann number. These satisfy

$$\psi_1 \psi_2 = -\psi_2 \psi_1, \psi \mathbf{c} = -\mathbf{c} \psi \tag{10.2}$$

[29]For many modes,

$$\{\mathbf{c}_i, \mathbf{c}_j\} = 0, \quad \{\mathbf{c}_i^\dagger, \mathbf{c}_j^\dagger\} = 0, \quad \{\mathbf{c}_i, \mathbf{c}_j^\dagger\} = \mathbb{1}\delta_{ij} \ .$$

– they anticommute with each other and with fermionic operators, and commute with ordinary numbers and bosons. They seem weird but they are easy. We'll need to consider multiple grassmann numbers when we have more than one fermion mode, where $\{\mathbf{c}_1, \mathbf{c}_2\} = 0$ will require that they anticommute $\{\psi_1, \psi_2\} = 0$ (as in the definition (10.2)); note that we will be simultaneously diagonalizing operators which *anti*commute.

The solution to equation (10.1) is very simple:

$$|\psi\rangle = |0\rangle - \psi\,|1\rangle$$

where as above $|0\rangle$ is the empty state ($\mathbf{c}\,|0\rangle = 0$) and $|1\rangle = \mathbf{c}^\dagger\,|0\rangle$ is the filled state. (Check: $\mathbf{c}\,|\psi\rangle = \mathbf{c}\,|0\rangle - \mathbf{c}\psi\,|1\rangle = +\psi\mathbf{c}\,|1\rangle = \psi\,|0\rangle = \psi\,|\psi\rangle$ .)

Similarly, the left-eigenvector of the creation operator is

$$\langle\bar{\psi}|\,\mathbf{c}^\dagger = \langle\bar{\psi}|\,\bar{\psi}, \quad \langle\bar{\psi}| = \langle 0| - \langle 1|\,\bar{\psi} = \langle 0| + \bar{\psi}\,\langle 1|\,.$$

Notice that these states are weird in that they are elements of an enlarged hilbert space with grassmann coefficients (usually we just allow complex numbers). Also, $\bar{\psi}$ is *not* the complex conjugate of $\psi$ and $\langle\bar{\psi}|$ is not the adjoint of $|\psi\rangle$. Rather, their overlap is

$$\langle\bar{\psi}|\psi\rangle = 1 + \bar{\psi}\psi = e^{\bar{\psi}\psi}.$$

---

**Grassmann calculus summary.** In the last expression we have seen an example of the amazing simplicity of Taylor's theorem for grassmann functions:

$$f(\psi) = f_0 + f_1\psi\ .$$

Integration is just as easy and in fact is the same as taking derivatives:

$$\int \psi d\psi = 1, \qquad \int 1 d\psi = 0.$$

With more than one grassmann we have to worry about the order:

$$1 = \int \bar{\psi}\psi d\psi d\bar{\psi} = -\int \bar{\psi}\psi d\bar{\psi} d\psi.$$

The only integral, really, is the gaussian integral:

$$\int e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi = a.$$

Many of these give

$$\int e^{-\bar{\psi}\cdot A\cdot\psi} d\bar{\psi} d\psi = \det A.$$

Here $\bar{\psi} \cdot A \cdot \psi \equiv \left(\bar{\psi}_1, \cdots, \bar{\psi}_M\right) \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & \ddots & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_M \end{pmatrix}$. One way to get this expression is to change variables to diagonalize the matrix $A$.

$$\left\langle \bar{\psi}\psi \right\rangle \equiv \frac{\int \bar{\psi}\psi e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi}{\int e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi} = -\frac{1}{a} = -\left\langle \psi\bar{\psi} \right\rangle .$$

If for many grassman variables we use the action $S = \sum_i a_i \bar{\psi}_i \psi_i$ (diagonalize $A$ above) then

$$\left\langle \bar{\psi}_i \psi_j \right\rangle = \frac{\delta_{ij}}{a_i} \equiv \langle \bar{i} j \rangle \tag{10.3}$$

and Wick's theorem here is

$$\left\langle \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l \right\rangle = \langle \bar{i} l \rangle \langle \bar{j} k \rangle - \langle \bar{i} k \rangle \langle \bar{j} l \rangle .$$

---

**Back to quantum mechanics:** The resolution of $\mathbb{1}$ in this basis is

$$\mathbb{1} = \int d\bar{\psi} d\psi \; e^{-\bar{\psi}\psi} \left|\psi\right\rangle \left\langle \bar{\psi}\right| \tag{10.4}$$

And if $\mathbf{A}$ is a bosonic operator (made of an even number of grassmann operators),

$$\mathrm{tr}\mathbf{A} = \int d\bar{\psi} d\psi \; e^{-\bar{\psi}\psi} \left\langle -\bar{\psi}\right| \mathbf{A} \left|\psi\right\rangle \; .$$

(Note the minus sign; it will lead to a deep statement.) So the partition function is:

$$Z = \int d\bar{\psi}_0 d\psi_0 \; e^{-\bar{\psi}_0\psi_0} \left\langle -\bar{\psi}_0\right| \underbrace{\overbrace{e^{-\frac{\mathbf{H}}{T}}}_{=(1-\Delta\tau\mathbf{H})\cdots(1-\Delta\tau\mathbf{H})}}_{M \text{ times}} \left|\psi_0\right\rangle$$

Now insert (10.4) in between each pair of Trotter factors to get

$$Z = \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l\psi_l} \left\langle \bar{\psi}_{l+1}\right| (1 - \Delta\tau\mathbf{H}) \left|\psi_l\right\rangle \; .$$

Because of the $-\bar{\psi}$ in (10.4), to get this nice expression we had to define an extra letter

$$\bar{\psi}_M = -\bar{\psi}_0, \quad \psi_M = -\psi_0 \tag{10.5}$$

so we could replace $\left\langle -\bar{\psi}_0\right| = \left\langle \bar{\psi}_M\right|$.

Now we use the coherent state property to turn the matrix elements into grassmann-valued functions:

$$\langle \bar{\psi}_{l+1} | \left(1 - \Delta\tau H(\mathbf{c}^\dagger, \mathbf{c})\right) |\psi_l\rangle = \langle \bar{\psi}_{l+1} | \left(1 - \Delta\tau H(\bar{\psi}_{l+1}, \psi_l)\right) |\psi_l\rangle \overset{\Delta\tau \to 0}{=} e^{\bar{\psi}_{l+1}\psi_l} e^{-\Delta\tau H(\bar{\psi}_{l+1}, \psi_l)}.$$

It was important that in $\mathbf{H}$ all $\mathbf{c}$s were to the right of all $\mathbf{c}^\dagger$s, *i.e.* that $\mathbf{H}$ was normal ordered.)

So we have

$$\begin{aligned} Z &= \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l\psi_l} e^{\bar{\psi}_{l+1}\psi_l} e^{-\Delta\tau H(\bar{\psi}_{l+1}, \psi_l)} \\ &= \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l \exp\left(\Delta\tau \left(\underbrace{\frac{\bar{\psi}_{l+1} - \bar{\psi}_l}{\Delta\tau}}_{=\partial_\tau \bar{\psi}} \psi_l - H(\bar{\psi}_{l+1}, \psi_l)\right)\right) \\ &\simeq \int [D\bar{\psi}D\psi] \exp\left(\int_0^{1/T} d\tau \; \bar{\psi}(\tau)\left(-\partial_\tau - \omega_0 + \mu\right)\psi(\tau)\right) = \int [D\bar{\psi}D\psi] e^{-S[\bar{\psi},\psi]}. \end{aligned} \tag{10.6}$$

Points to note:

- In the penultimate step we defined, as usual, continuum fields

$$\psi(\tau_l = \Delta\tau l) \equiv \psi_l, \qquad \bar{\psi}(\tau_l = \Delta\tau l) \equiv \bar{\psi}_l.$$

- We elided the difference $H(\bar{\psi}_{l+1}, \psi_l) = H(\bar{\psi}_l, \psi_l) + \mathcal{O}(\Delta\tau)$ in the last expression. This difference is usually negligible and sometimes helpful (an example where it's helpful is the discussion of the number density below).

- The APBCs (10.5) on $\psi(\tau + \frac{1}{T}) = -\psi(\tau)$ mean that in its fourier representation[30]

$$\psi(\tau) = T\sum_n \psi(\omega) e^{-\mathbf{i}\omega_n\tau}, \qquad \bar{\psi}(\tau) = T\sum_n \bar{\psi}(\omega) e^{\mathbf{i}\omega_n\tau} \tag{10.7}$$

  the *Matsubara frequencies*

$$\omega_n = (2n+1)\pi T, \quad n \in \mathbb{Z}$$

  are odd multiples of $\pi T$.

- The measure $[D\bar{\psi}D\psi]$ is defined by this equation, just as in the bosonic path integral.

[30] $\bar{\psi}$ is still not the complex conjugate of $\psi$ but the relative sign is convenient.

- The derivative of a grassmann function is also defined by this equation; note that $\psi_{l+1} - \psi_l$ is not 'small' in any sense.
- In the last step we integrated by parts, *i.e.* relabeled terms in the sum, so

  $$\sum_l \left(\bar\psi_{l+1} - \bar\psi_l\right)\psi_l = \sum_l \bar\psi_{l+1}\psi_l - \sum_l \bar\psi_l\psi_l = \sum_{l'=l-1} \bar\psi_{l'}\psi_{l-1} - \sum_l \bar\psi_l\psi_l = -\sum_l \bar\psi_l\left(\psi_l - \psi_{l-1}\right).$$

  Note that no grassmanns were moved through each other in this process.

The punchline of this discussion for now is that the euclidean action is

$$\boxed{S[\bar\psi,\psi] = \int d\tau\left(\bar\psi\partial_\tau\psi + H(\bar\psi,\psi)\right)}\,.$$

The first-order kinetic term we've found $\bar\psi\partial_\tau\psi$ is sometimes called a 'Berry phase term'. Note the funny-looking sign.

---

**Continuum limit warning** (about the red $\simeq$ in (10.6)). The Berry phase term is *actually*

$$\sum_{l=0}^{N-1}\bar\psi_{l+1}\left(\psi_{l+1} - \psi_l\right) = T\sum_{\omega_n}\bar\psi(\omega_n)\left(1 - e^{\mathbf{i}\omega_n\tau}\right)\psi(\omega_n)$$

and in (10.6) we have kept only the leading nonzero term:

$$\left(1 - e^{\mathbf{i}\omega_n\tau}\right) \to \mathbf{i}\omega_n\tau.$$

Clearly this replacement is just fine if

$$\omega_n\tau \ll 1$$

for all $\omega_n$ which matter. Which $\omega_n$ contribute? I claim that if we use a reasonable $\mathbf{H} = \mathbf{H}_{\text{quadratic}} + \mathbf{H}_{\text{int}}$, reasonable quantities like $Z$, $\left\langle\mathcal{O}^\dagger\mathcal{O}\right\rangle$, are dominated by $\omega_n \ll \tau^{-1}$.

---

There's more we can learn from what we've done here that I don't want to pass up. Let's use this formalism to compute the fermion density at $T = 0$:

$$\langle\mathbf{N}\rangle = \frac{1}{Z}\text{tr}e^{-\mathbf{H}/T}\mathbf{c}^\dagger\mathbf{c}.$$

This is an example where the annoying $\Delta\tau$s in the path integral not only matter, but are extremely friendly to us.

---

**Frequency space, $T \to 0$.**

Let's change variables to frequency-space fields, which diagonalize $S$. The Jacobian is 1 (since fourier transform is unitary):

$$D\bar{\psi}(\tau)D\psi(\tau) = \prod_n d\bar{\psi}(\omega_n)d\psi(\omega_n) \overset{T\to 0}{\to} D\bar{\psi}(\omega)D\psi(\omega).$$

The partition function is

$$Z = \int D\bar{\psi}(\omega)D\psi(\omega)\ \exp\left(T\sum_{\omega_n}\bar{\psi}(\omega_n)\left(\mathbf{i}\omega_n - \omega_0 + \mu\right)\psi(\omega_n)\right).$$

Notice that in the zero-temperature limit

$$T\sum_{\omega_n} \mapsto \int \frac{d\omega}{2\pi} \equiv \int đ\omega.$$

(This is the same fact as $V\sum_k \mapsto \int đ^d k$ in the thermodynamic limit.) So the zero-temperature partition function is

$$Z \overset{T\to 0}{=} \int D\bar{\psi}(\omega)D\psi(\omega)\ \exp\left(\int_{-\infty}^{\infty} đ\omega\bar{\psi}(\omega)\left(\mathbf{i}\omega - \omega_0 + \mu\right)\psi(\omega)\right).$$

Using the gaussian-integral formula (10.3) you can see that the *propagator* for $\psi$ is

$$\left\langle \bar{\psi}(\omega_1)\psi(\omega_2)\right\rangle = \underbrace{\frac{\delta_{\omega_1,\omega_2}}{T}}_{\overset{T\to 0}{\to}\delta(\omega_1-\omega_2)} \frac{2\pi}{\mathbf{i}\omega_1 - \omega_0 + \mu}. \tag{10.8}$$

In particular $\left\langle \bar{\psi}(\omega)\psi(\omega)\right\rangle = \frac{2\pi/T}{\mathbf{i}\omega-\omega_0+\mu}$. $\delta(\omega = 0) = 1/T$ is the 'volume' of the time direction.

---

Back to the number density. Using the same strategy as above, we have

$$\langle \mathbf{N}\rangle = \frac{1}{Z}\int \prod_{l=0}^{M-1+1}\left(d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l\psi_l}\right)\prod_{l=1}^{M-1}\left\langle \bar{\psi}_{l+1}\right|(1-\Delta\tau\mathbf{H}(\mathbf{c}^\dagger\mathbf{c}))\left|\psi_l\right\rangle \underbrace{\left\langle \bar{\psi}_{N+1}\right|\mathbf{c}^\dagger\mathbf{c}\left|\psi_N\right\rangle}_{=\bar{\psi}_{N+1}\psi_N=\bar{\psi}(\tau_N+\Delta\tau)\psi(\tau_N)},$$

where $\tau_N$ is any of the time steps. This formula has a built-in point-splitting of the operators!

$$\begin{aligned}\langle \mathbf{N}\rangle &= \frac{1}{Z}\int D\bar{\psi}D\psi\ e^{-S[\bar{\psi},\psi]}\bar{\psi}(\tau_N+\Delta\tau)\psi(\tau_N)\\ &= \int_{-\infty}^{\infty} đ\omega\frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega-\omega_0+\mu} = \theta(\mu-\omega_0).\end{aligned} \tag{10.9}$$

Which is the right answer: the mode is occupied in the groundstate only if $\omega_0 < \mu$. In the last step we used the fact that $\Delta\tau > 0$ to close the contour in the UHP; so we only pick up the pole if it is in the UHP. Notice that this quantity is very *UV sensitive*: if we put a frequency cutoff on the integral, $\int^\Lambda \frac{d\omega}{\omega} \sim \log\Lambda$, the integral diverges logarithmically. For most calculations the $\Delta\tau$ can be ignored, but here it told us the right way to treat the divergence. [31]

[End of Lecture 32]

## 10.2 Anomalies

[Zee §IV.7; Polyakov, *Gauge Fields and Strings*, §6.3; K. Fujikawa, *Phys. Rev. Lett.* **42** (1979) 1195; Argyres, 1996 lectures on supersymmetry §14.3; Peskin, chapter 19]

Topology means the study of quantities which can't vary smoothly, but can only jump. Like quantities which must be integers. *Anomalies* are an example of a topological phenomenon in QFT, which is therefore robust against any change in the QFT which can be made continuously (like varying masses or couplings, or the cutoff or the resolution of our description, *i.e.* a renormalization group transformation).

Suppose we have in our hands a classical field theory in the continuum which has some symmetry. Is there a well-defined QFT whose classical limit produces this classical field theory and preserves that symmetry? The path integral construction of QFT offers some insight here. The path integral involves two ingredients: (1) an action, which is shared with the classical field theory, and (2) a path integral measure. It is possible that the action is invariant but the measure is not. This is called an anomaly. It means that the symmetry is broken, and its current conservation is violated by a known amount, and this often has many other consequences that can be understood by humans.

Notice that here I am speaking about actual, global symmetries. I am not talking about gauge redundancies. If you think that two field configurations are equivalent but the path integral tells you that they would give different contributions, you are doing something wrong. Such a 'gauge anomaly' means that the system has more degrees of freedom than you thought. (In particular, it does not mean that the world

[31] The calculation between the first and second lines of (10.9) is familiar to us – it is a single Wick contraction, and can be described as a feynman diagram with one line between the two insertions. More prosaically, it is

$$\left\langle \bar\psi(\tau_N+\Delta\tau)\psi(\tau_N)\right\rangle \overset{(10.7)}{=} T^2\sum_{nm} e^{\mathbf{i}(\omega_n-\omega_m)\tau+\mathbf{i}\omega_n\Delta\tau}\left\langle \bar\psi(\omega_n)\psi(\omega_m)\right\rangle \overset{(10.8)}{=} T\sum_m \frac{e^{\mathbf{i}\omega_n\Delta\tau}}{\mathbf{i}\omega_n-\omega_0+\mu} \overset{T\to 0}{\to} \int đ\omega \frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega-\omega_0+\mu}.$$

is inconsistent. For a clear discussion of this, please see Preskill, 1990.)

You could say that we have already seen a dramatic example of an anomaly: the violation of classical scale invariance (*e.g.* in massless $\phi^4$ theory, or in massless QED) by quantum effects.

Notice that the name 'anomaly' betrays the bias that we imagine constructing a QFT by starting with a continuum action for a classical field theory; you would never imagine that *e.g.* scale invariance was an exact symmetry if you instead started from a well-defined quantum lattice model.

The example we will focus on here is the *chiral anomaly.* This is an equation for the violation of the chiral (aka axial) current for fermions coupled to a background gauge field. The chiral anomaly was first discovered in perturbation theory, by computing a certain Feynman diagram with a triangle; the calculation was motivated by the experimental observation of the process $\pi^0 \to \gamma\gamma$, which would not happen if the chiral current were conserved. (The relationship between the chiral current and the pion is explained in §11.7.)

I will outline a derivation of this effect (using the fermionic path integral) which is more illuminating than the triangle diagram. It shows that the one-loop result is exact – there are no other corrections. It shows that the quantity on the right hand side of the continuity equation for the would-be current integrates to an integer. It gives a proof of the *index theorem*, relating numbers of solutions to the Dirac equation in a background field configuration to a certain integral of field strengths. It butters your toast.

### 10.2.1 Chiral anomaly

**Chiral symmetries.** In even-dimensional spacetimes, the Dirac representation of $\mathrm{SO}(D-1,1)$ is reducible. This is because

$$\gamma^5 \equiv \prod_{\mu=0}^{D-1} \gamma^\mu \neq 1, \quad \text{satisfies} \quad \{\gamma^5, \gamma^\mu\} = 0, \forall \mu$$

which means that $\gamma^5$ commutes with the Lorentz generators

$$[\gamma^5, \Sigma^{\mu\nu}] = 0, \qquad \Sigma^{\mu\nu} \equiv \frac{1}{2}[\gamma^\mu, \gamma^\nu].$$

A left- or right-handed Weyl spinor is an irreducible representation of $\mathrm{SO}(D-1,1)$, $\psi_{L/R} \equiv \frac{1}{2}\left(1 \pm \gamma^5\right)\psi$. This allows the possibility that the L and R spinors can transform differently under a symmetry; such a symmetry is a chiral symmetry.

Note that in $D = 4k$ dimensions, if $\psi_L$ is a left-handed spinor in representation $\mathbf{r}$ of some group $G$, then its image under CPT, $\psi_L^{CPT}(t,\vec{x}) \equiv \mathbf{i}\gamma^0 \left(\psi_L(-t,-\vec{x})\right)^\star$, is right-handed and transforms in representation $\bar{\mathbf{r}}$ of $G$. Therefore chiral symmetries arise when the Weyl fermions transform in *complex representations* of the symmetry group, where $\bar{\mathbf{r}} \neq \mathbf{r}$. (In $D = 4k+2$, CPT maps left-handed fields to left-handed fields. For more detail on discrete symmetries and Dirac fields, see Peskin §3.6.)

---

Some more explicit words (of review) about chiral fermions in $D = 3+1$, mostly notation. Recall Peskin's *Weyl* basis of gamma matrices in 3+1 dimensions, in which $\gamma^5$ is diagonal:

$$\gamma^\mu = \begin{pmatrix} 0 & \bar{\boldsymbol{\sigma}}^\mu \\ \boldsymbol{\sigma}^\mu & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}^\mu \equiv (\mathbb{1}, \vec{\boldsymbol{\sigma}})^\mu, \quad \bar{\boldsymbol{\sigma}}^\mu \equiv (\mathbb{1}, -\vec{\boldsymbol{\sigma}})^\mu, \quad \gamma^5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This makes the reducibility of the Dirac representation of SO(3,1) manifest, since the Lorentz generators are $\propto [\gamma^\mu, \gamma^\nu]$ block diagonal in this basis. The gammas are a map from the $(1, \mathbf{2_R})$ representation to the $(\mathbf{2_L}, \mathbf{1})$ representation. It is sometimes useful to denote the $\mathbf{2_R}$ indices by $\alpha, \beta = 1,2$ and the $\mathbf{2_L}$ indices by $\dot{\alpha}, \dot{\beta} = 1,2$. Then we can define two-component Weyl spinors $\psi_{L/R} = P_{L/R}\psi \equiv \frac{1}{2}(1 \pm \gamma^5)\psi$ by simply forgetting about the other two components. The conjugate of a $L$ spinor $\chi = \psi_L$ ($L$ means $\gamma^5\chi = \chi$) is right-handed:

$$\bar{\chi} = \chi^\dagger\gamma^0, \quad \bar{\chi}\gamma^5 = \chi^\dagger\gamma^0\gamma^5 = -\chi^\dagger\gamma^5\gamma^0 = -\chi^\dagger\gamma^0 = -\bar{\chi}.$$

We can represent any system of Dirac fermions in terms of a collection of twice as many Weyl fermions.

For a continuous symmetry $G$, we can be more explicit about the meaning of a complex representation. The statement that $\psi$ is in representation $\mathbf{r}$ means that its transformation law is

$$\delta\psi_a = \mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)_{ab}\psi_b$$

where $t^A, A = 1..\dim G$ are generators of $G$ in representation $\mathbf{r}$; for a compact lie group $G$, we may take the $t^A$ to be Hermitian. The conjugate representation, by definition, is one with which you can make a singlet of $G$ – it's the way $\psi^{\star T}$ transforms:

$$\delta\psi_a^{\star T} = -\mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)^T_{ab}\psi_b^{\star T}.$$

So:

$$t^A_{\bar{\mathbf{r}}} = -\left(t^A_{\mathbf{r}}\right)^T.$$

The condition for a complex representation is that this is different from $t^A_{\mathbf{r}}$ (actually we have to allow for relabelling of the generators and the basis). The simplest case is

$G = \mathrm{U}(1)$, where $t$ is just a number indicating the charge. In that case, any nonzero charge gives a complex representation.

---

Consider the effective action produced by integrating out Dirac fermions coupled to a *background* gauge field (the gauge field is just going to sit there for this whole calculation):

$$e^{\mathbf{i}S_{\text{eff}}[A]} \equiv \int [D\psi D\bar{\psi}]\ e^{\mathbf{i}S[\psi,\bar{\psi},A]}\ .$$

We must specify how the fermions coupled to the gauge field. The simplest example is if $A$ is a $U(1)$ gauge field and $\psi$ is minimally coupled:

$$S[\psi,\bar{\psi},A] = \int \mathrm{d}^D x \bar{\psi}\mathbf{i}\not{D}\psi, \qquad \not{D}\psi \equiv \gamma^\mu \left(\partial_\mu + \mathbf{i}A_\mu\right)\psi.$$

---

We will focus on this example, but you could imagine instead that $A_\mu$ is a non-Abelian gauge field for the group $G$, and $\psi$ is in a representation $R$, with gauge generators $T^A(R)$ ($A = 1...\mathrm{dim}G$), so the coupling would be

$$\bar{\psi}\mathbf{i}\not{D}\psi = \bar{\psi}_a\gamma^\mu \left(\partial_\mu\delta_{ab} + \mathbf{i}A_\mu^A T^A(R)_{ab}\right)\psi_b\ . \tag{10.10}$$

Much of the discussion below applies for any even $D$.

---

In the absence of a mass term, the action (in the Weyl basis) involves no coupling between $L$ and $R$:

$$S[\psi,\bar{\psi},A] = \int \mathrm{d}^D x \left(\psi_L^\dagger \mathbf{i}\sigma^\mu D_\mu \psi_L + \psi_R^\dagger \mathbf{i}\bar{\sigma}^\mu D_\mu \psi_R\right)$$

and therefore is invariant under the global chiral rotation

$$\psi \to e^{\mathbf{i}\alpha\gamma^5}\psi, \quad \psi^\dagger \to \psi^\dagger e^{-\mathbf{i}\alpha\gamma^5}, \quad \bar{\psi} \to \bar{\psi}e^{+\mathbf{i}\alpha\gamma^5}. \quad \text{That is: } \psi_L \to e^{\mathbf{i}\alpha}\psi_L, \quad \psi_R \to e^{-\mathbf{i}\alpha}\psi_R.$$

(The mass term couples the two components

$$L_m = \bar{\psi}\left(\mathrm{Re}m + \mathrm{Im}m\gamma^5\right)\psi = m\psi_L^\dagger\psi_R + h.c.;$$

notice that the mass parameter is complex.) The associated Noether current is $j_\mu^5 = \bar{\psi}\gamma^5\gamma_\mu\psi$, and it seems like we should have $\partial^\mu j_\mu^5 \stackrel{?}{=} 0$. This follows from the massless (classical) Dirac equation $0 = \gamma^\mu\partial_\mu\psi$. (With the mass term, we would have instead $\partial^\mu j_\mu^5 \stackrel{?}{=} 2\mathbf{i}\bar{\psi}\left(\mathrm{Re}m\gamma^5 + \mathrm{Im}m\right)\psi$. )

Notice that there is another current $j^\mu = \bar{\psi}\gamma^\mu\psi$. $j^\mu$ is the current which is coupled to the gauge field, $L \ni A_\mu j^\mu$. The conservation of this current is required for gauge invariance of the effective action

$$S_{\text{eff}}[A_\mu] \stackrel{!}{=} S_{\text{eff}}[A_\mu + \partial_\mu\lambda] \sim \log\left\langle e^{\mathbf{i}\int\lambda(x)\partial_\mu j^\mu}\right\rangle + S_{\text{eff}}[A_\mu].$$

No matter what happens we can't find an anomaly in $j^\mu$. The anomalous one is the other one, the *axial current*.

To derive the conservation law we can use the Noether method. This amounts to substituting $\psi'(x) \equiv e^{\mathbf{i}\alpha(x)\gamma^5}\psi(x)$ into the action:

$$S_F[\psi'] = \int \mathrm{d}^D x \bar{\psi} e^{+\mathbf{i}\alpha\gamma^5} \mathbf{i}\not{D} e^{\mathbf{i}\alpha\gamma^5}\psi = \int \mathrm{d}^D x \left(\bar{\psi}\mathbf{i}\not{D}\psi + \bar{\psi}\mathbf{i}\gamma^5 \left(\not{\partial}\alpha\right)\psi\right) \overset{\text{IBP}}{=} S_F[\psi] - \mathbf{i}\int \alpha(x)\partial^\mu \mathrm{tr}\bar{\psi}\gamma^5\gamma_\mu\psi.$$

Then we can completely get rid of $\alpha(x)$ if we can change integration variables, *i.e.* if $[D\psi'] \overset{?}{=} [D\psi]$. Usually this is true, but here we pick up an interesting Jacobian.

Claim:

$$e^{\mathbf{i}S_{\text{eff}}[A]} = \int [D\psi' D\bar{\psi}'] e^{\mathbf{i}S_F[\psi']} = \int [D\psi D\bar{\psi}] e^{\mathbf{i}S_F[\psi] + \int \mathrm{d}^D x \alpha(x)\left(\partial_\mu j_5^\mu - \mathcal{A}(x)\right)}$$

where

$$\mathcal{A}(x) = \sum_n \mathrm{tr}\bar{\xi}_n \gamma^5 \xi_n \tag{10.11}$$

where $\xi_n$ are a basis of eigenspinors of the Dirac operator. The contribution to $\mathcal{A}$ can be attributed to zeromodes of the Dirac operator.

The expression above is actually independent of $\alpha$, since the path integral is invariant under a change of variables. For a conserved current, $\alpha$ would multiply the divergence of the current and this demand would imply current conservation. Here this implies that instead of current conservation we have a specific violation of the current:

$$\partial^\mu j^5_\mu = \mathcal{A}(x).$$

**What is the anomaly $\mathcal{A}$?** [Polyakov §6.3] An alternative useful (perhaps more efficient) perspective is that the anomaly arises from trying to define the axial current operator, which after all is a composite operator. Thus we should try to compute

$$\langle \partial_\mu j_5^\mu \rangle = \partial_\mu \left\langle \bar{\psi}(x)\gamma^\mu\gamma^5\psi(x)\right\rangle$$

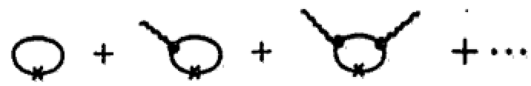
– the coincident operators on the RHS need to be regulated.

The classical (massless) Dirac equation immediately implies that the axial current is conserved

$$\partial_\mu \left(\bar{\psi}\gamma^\mu\gamma^5\psi\right) \overset{?}{=} 0.$$

Consider, on the other hand, the (Euclidean vacuum) expectation value

$$J^5_\mu \equiv \left\langle \bar{\psi}(x)\gamma_\mu\gamma^5\psi(x)\right\rangle \equiv Z^{-1}[A]\int [D\psi D\bar{\psi}] e^{-S_F[\psi]} j^5_\mu(x)$$

$=$ 

$$= -\text{Tr}\,{}_{\gamma}\gamma_\mu\gamma^5 G^{[A]}(x,x) \tag{10.12}$$

where $G$ is the Green's function of the Dirac operator in the gauge field background (and the figure is from Polyakov's book). We can construct it out of eigenfunctions of $\mathbf{i}\not{D}$:

$$\mathbf{i}\not{D}\xi_n(x) = \epsilon_n\xi_n(x), \qquad \bar{\xi}_n(x)\mathbf{i}\gamma^\mu\left(-\overleftarrow{\partial}_\mu + \mathbf{i}A_\mu\right) = \epsilon_n\bar{\xi}_n \tag{10.13}$$

in terms of which[32]

$$G(x,x') = \sum_n \frac{1}{\epsilon_n}\xi_n(x)\bar{\xi}_n(x'). \tag{10.14}$$

(I am suppressing spinor indices all over the place, note that here we are taking the outer product of the spinors.)

We want to define the coincidence limit, as $x' \to x$. The problem with this limit arises from the large $|\epsilon_n|$ eigenvalues; the contributions of such short-wavelength modes are local and most of them can be absorbed in renormalization of couplings. It should not (and does not) matter how we regulate them, but we must pick a regulator. A convenient choice here is heat-kernel regulator:

$$G_s(x,x') \equiv \sum_n e^{-s\epsilon_n^2}\frac{1}{\epsilon_n}\xi_n(x)\bar{\xi}_n(x')$$

and

$$J_\mu^5(x) = \sum_n e^{-s\epsilon_n^2}\frac{1}{\epsilon_n}\bar{\xi}_n(x)\gamma^5\gamma_\mu\xi_n(x)\ .$$

The anomaly is

$$\partial^\mu J_\mu^5 = \partial^\mu\left\langle j_\mu^5\right\rangle = \sum_n \mathbf{i}\partial^\mu\left(\bar{\xi}_n\gamma_\mu\gamma^5\xi_n\right)\frac{e^{-s\epsilon_n^2}}{\epsilon_n}.$$

The definition (10.13) says

$$\mathbf{i}\partial^\mu\left(\bar{\xi}_n\gamma^5\gamma_\mu\xi_n\right) = 2\epsilon_n\bar{\xi}_n\gamma_5\xi_n$$

using $\{\gamma^5,\gamma^\mu\} = 0$. (Notice that the story would deviate dramatically here if we were studying the vector current which lacks the $\gamma^5$.) This gives

$$\partial^\mu J_\mu^5 = 2\text{Tr}\ \gamma^5 e^{-s\left(\mathbf{i}\not{D}\right)^2}$$

[32]Actually, this step is full of danger. (Polyakov has done it to me again. Thanks to Sridip Pal for discussions of this point.) See §10.2.2 below.

with

$$(\mathbf{i}\not{D})^2 = -\left(\gamma_\mu\left(\partial_\mu + \mathbf{i}A_\mu\right)\right)^2 = -\left(\partial_\mu + A_\mu\right)^2 - \frac{\mathbf{i}}{2}\Sigma_{\mu\nu}F^{\mu\nu}$$

where $\Sigma_{\mu\nu} \equiv \frac{1}{2}[\gamma_\mu, \gamma_\nu]$ is the spin Lorentz generator. This is (10.11), now better defined by the heat kernel regulator. We've shown that in any even dimension,

$$\partial^\mu \left\langle j^5_\mu(x)\right\rangle = 2\text{Tr }_\alpha \gamma^5 e^{s\not{D}^2} \tag{10.15}$$

This can now be expanded in small $s$, which amounts to an expansion in powers of $A, F$. If there is no background field, $A = 0$, we get

$$\left\langle x|e^{-s\left(\mathbf{i}\not{\partial}\right)^2}|x\right\rangle = \int đ^D p\ e^{-sp^2} = \underbrace{K_D}_{=\frac{\Omega_{D-1}}{(2\pi)^D}\text{ as before}} \frac{1}{s^{D/2}} \overset{D=4}{=} \frac{1}{16\pi^2 s^2}. \tag{10.16}$$

This term will renormalize the charge density

$$\rho(x) = \left\langle \psi^\dagger\psi(x)\right\rangle = \text{tr}\gamma^0 G(x,x),$$

for which we must add a counterterm (in fact, it is accounted for by the counterterm for the gauge field kinetic term, *i.e.* the running of the gauge coupling). But it will not affect the axial current conservation which is proportional to

$$\text{tr}\left(\gamma^5 G(x,x)\right)|_{A=0} \propto \text{tr}\gamma^5 = 0.$$

Similarly, bringing down more powers of $(\partial + A)^2$ doesn't give something nonzero since the $\gamma^5$ remains.

In $D = 4$, the first term from expanding $\Sigma_{\mu\nu}F^{\mu\nu}$ is still zero from the spinor trace. (Not so in $D = 2$.) The first nonzero term comes from the next term:

$$\text{tr}\left(\gamma_5 e^{-s\left(\mathbf{i}\not{D}\right)^2}\right)_{xx} = \underbrace{\left\langle x|e^{-s(\mathbf{i}D)^2}|x\right\rangle}_{\overset{(10.16)}{=}\frac{1}{16\pi^2 s^2}+\mathcal{O}(s^{-1})} \cdot \frac{s^2}{8}\cdot(\mathbf{i}^2)\underbrace{\text{tr}\left(\gamma^5\Sigma^{\mu\nu}\Sigma^{\rho\lambda}\right)}_{=4\epsilon^{\mu\nu\rho\lambda}}\cdot\underbrace{\text{tr}_c}_{\text{color}}\left(F_{\mu\nu}F_{\rho\lambda}\right) + \mathcal{O}(s^1)\ .$$

In the abelian case, just ignore the trace over color indices, $\text{tr}_c$. The terms that go like positive powers of $s$ go away in the continuum limit. Therefore

$$\partial_\mu J^\mu_5 = -2\cdot\frac{1}{16\pi^2 s^2}\cdot\frac{s^2}{8}\cdot 4\epsilon^{\mu\nu\rho\lambda}\text{tr}_c F_{\mu\nu}F_{\rho\lambda} + \mathcal{O}(s^1) = -\frac{1}{8\pi^2}\text{tr}F_{\mu\nu}\left(\star F\right)^{\mu\nu}. \tag{10.17}$$

(Here $(\star F)^{\mu\nu} \equiv \frac{1}{8}\epsilon^{\mu\nu\rho\lambda}F_{\rho\lambda}$.) This is the chiral anomaly formula. It can also be usefully written as:

$$\partial_\mu J^\mu_5 = -\frac{1}{8\pi^2}\text{tr}F\wedge F = -\frac{1}{32\pi^2}\vec{E}\cdot\vec{B}.$$

- This object on the RHS is a total derivative. In the abelian case it is

$$F \wedge F = \mathrm{d}\left(A \wedge F\right) \ .$$

Its integral over spacetime is a topological (in fact $16\pi^2$ times an integer) characterizing the gauge field configuration. How do I know it is an integer? The anomaly formula! The change in the number of left-handed fermions minus the number of right-handed fermions during some time interval is:

$$\Delta Q_A \equiv \Delta\left(N_L - N_R\right) = \int \mathrm{d}t \partial_t J_0^5 = \int_{M_4} \partial^\mu J_\mu^5 = 2\int_{M_4} \frac{F \wedge F}{16\pi^2}$$

where $M_4$ is the spacetime region under consideration. If nothing is going on at the boundaries of this spacetime region (*i.e.* the fields go to the vacuum, or there is no boundary, so that no fermions are entering or leaving), we can conclude that the RHS is an integer.

- Look back at the diagrams in (10.12). Which term in that expansion gave the nonzero contribution to the axial current violation? In $D = 4$ it is the diagram with three current insertions, the ABJ triangle diagram. So in fact we did end up computing the triangle diagram. But this calculation also shows that nothing else contributes, even non-perturbatively.

- We chose a particular regulator above. The answer we got did not depend on the cutoff; in fact, whatever regulator we used, we would get this answer.

- Consider what happens if we redo this calculation in other dimensions. We only consider even dimensions because in odd dimensions there is no analog of $\gamma^5$ – the Dirac spinor representation is irreducible. In $2n$ dimensions, we need $n$ powers of $F$ to soak up the indices on the epsilon tensor. Actually there is an analogous phenomenon in odd dimensions (sometimes called parity anomaly) of an effect that is independent of the masses of the fields which you'll study on the homework. Instead of $F^n$, the thing that appears is the Chern-Simons term.

- If we had kept the non-abelian structure in (10.10) through the whole calculation, the only difference is that the trace in (10.17) would have included a trace over representations of the gauge group; and we could have considered also a non-abelian flavor transformation

$$\psi_I \to \left(e^{\mathbf{i}\gamma^5 g^a \tau^a}\right)_{IJ} \psi_J$$

for some flavor rotation generator $\tau^a$. Then we would have found:

$$\partial^\mu j_\mu^{5a} = \frac{1}{16\pi^2} \epsilon^{\mu\nu\rho\lambda} F_{\mu\nu}^A F_{\rho\lambda}^B \mathrm{tr}_{c,a}\left(T^A T^B \tau^a\right) .$$

A similar statement applies to the case of multiple species of fermion fields: their contributions to the anomaly add. Sometimes they can cancel; the Electroweak gauge interactions are an example of this.

[End of Lecture 33]

### 10.2.2 Zeromodes of the Dirac operator

Do you see now why I said that the step involving the fermion Green's function was full of danger? The danger arises because the Dirac operator (whose inverse is the Green's function) can have *zeromodes*, eigenspinors with eigenvalue $\epsilon_n = 0$. In that case, $\mathbf{i}\not{D}$ is *not* invertible, and the expression (10.14) for $G$ is ambiguous. This factor of $\epsilon_n$ is about to be cancelled when we compute the divergence of the current and arrive at (10.11). Usually this kind of thing is not a problem because we can lift the zeromodes a little and put them back at the end. But here it is actually hiding something important. The zeromodes cannot just be lifted. This is true because nonzero modes of $\mathbf{i}\not{D}$ must come in left-right pairs: this is because $\{\gamma^5, \mathbf{i}\not{D}\} = 0$, so $\mathbf{i}\not{D}$ and $\gamma^5$ cannot be simultaneously diagonalized in general. That is: if $\mathbf{i}\not{D}\xi = \epsilon\xi$ then $(\gamma^5\xi)$ is also an eigenvector of $\mathbf{i}\not{D}\xi$, with eigenvalue $-\epsilon$. Only for $\epsilon = 0$ does this fail, so zeromodes can come by themselves. So you can't just smoothly change the eigenvalue of some $\xi_0$ from zero unless it has a partner with whom to pair up. By taking linear combinations

$$\chi_n^{L/R} = \frac{1}{2}\left(1 \pm \gamma^5\right)\xi_n$$

these two partners can be arranged into a pair of simultaneous eigenvectors of $(\mathbf{i}\not{D})^2$ (with eigenvalue $\epsilon_n^2$) and of $\gamma^5$ with $\gamma^5 = \pm$ respectively.

This leads us to a deep fact, called the (Atiyah-Singer) *index theorem*: *only* zeromodes can contribute to the anomaly. Any mode $\xi_n$ with nonzero eigenvalue has a partner (with the same eigenvalue of $(\mathbf{i}\not{D})^2$) with the opposite sign of $\gamma^5$; hence they cancel exactly in

$$\mathrm{tr}\gamma^5 e^{-s(\mathbf{i}\not{D})^2} = \sum_{n, L/R} \bar{\chi}_n^{L/R}\gamma^5\chi_n^{L/R} e^{-s\epsilon_n^2} + \text{zeromodes} \quad .$$

So the anomaly equation tells us that the number of zeromodes of the Dirac operator, weighted by handedness (*i.e.* with a + for $L$ and - for $R$) is equal to

$$N_L - N_R = \int d^D x \mathcal{A}(x) = \int \frac{1}{16\pi^2} F \wedge F.$$

A practical consequence for us is that it makes manifest that the result is independent of the regulator $s$.

### 10.2.3 The physics of the anomaly

[Polyakov, page 102; Kaplan 0912.2560 §2.1; Alvarez-Gaumé] Consider non-relativistic free (*i.e.* no 4-fermion interactions) fermions in 1+1 dimensions, *e.g.* with 1-particle dispersion $\omega_k = \frac{1}{2m}\vec{k}^2$. The groundstate of $N$ such fermions is described by filling the $N$ lowest-energy single particle levels, up the Fermi momentum: $|k| \leq k_F$ are filled. We must introduce an infrared regulator so that the levels are discrete – put them in a box of length $L$, so that $k_n = \frac{2\pi n}{L}$. (In Figure 1, the red circles are possible 1-particle states, and the green ones are the occupied ones.) The lowest-energy excitations of this groundstate come from taking a fermion just below the Fermi level $k = |k_F - k_1|$ and putting it just above, $k = |k_F + k_2|$; the energy cost is

$$E_{k_1 - k_2} = \frac{1}{2m}(k_F + k_1)^2 - \frac{1}{2m}(k_F - k_2)^2 \simeq \frac{k_F}{m}(k_1 - k_2)$$

The energy – we get relativistic dispersion with velocity $v_F = \frac{k_F}{m}$. The fields near these Fermi points in $k$-space satisfy the Dirac equation[33]:

$$(\omega - v_F \delta k)\,\psi_L = 0, \quad (\omega + v_F \delta k)\,\psi_R = 0$$

where $\delta k \equiv k - k_F$.

[33] This example is worthwhile for us also because we see the relativistic Dirac equation is emerging from a non-relativistic model; in fact we could have started from an even more distant starting point – *e.g.* from a lattice model, like

$$H = -t \sum_n c_n^\dagger c_{n+1} + h.c.$$

where the dispersion would be $\omega_k = -2t(\cos ka - 1) \sim \frac{1}{2m}k^2 + \mathcal{O}(k^4)$ with $\frac{1}{2m} = ta^2$.

It would therefore seem to imply a conserved axial current – the number of left moving fermions minus the number of right moving fermions. But the fields $\psi_L$ and $\psi_R$ are not independent; with high-enough energy excitations, you reach the bottom of the band (near $k = 0$ here) and you can't tell the difference. This means that the numbers are *not* separately conserved.

We can do better in this 1+1d example and show that the amount by which the axial current is violated is given by the anomaly formula. Consider subjecting our poor 1+1d free fermions to an electric field $E_x(t)$ which is constant in space and slowly varies in time. Suppose we gradually turn it on and then turn it off; here gradually means slowly enough that the process is adiabatic. Then each particle experiences a force $\partial_t p = eE_x$ and its net change in momentum is

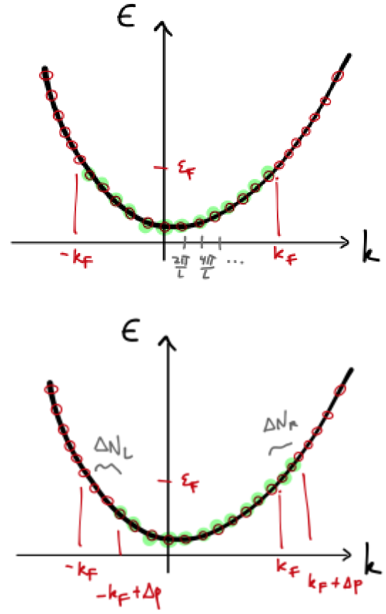



**Figure 1:** Green dots represent occupied 1-particle states. Top: In the groundstate. Bottom: After applying $E_x(t)$.

$$\Delta p = e \int \mathrm{d}t E_x(t).$$

This means that the electric field puts the fermions in a state where the Fermi surface $k = k_F$ has shifted to the right by $\Delta p$, as in the figure. Notice that the total number of fermions is of course the same – charge is conserved.

Now consider the point of view of the low-energy theory at the Fermi points. This theory has the action

$$S[\psi] = \int \mathrm{d}x\mathrm{d}t \bar{\psi} \left(\mathbf{i}\gamma^\mu \partial_\mu\right) \psi \ ,$$

where $\gamma^\mu$ are $2 \times 2$ and the upper/lower component of $\psi$ creates fermions near the left/right Fermi point. In the process above, we have added $N_R$ right-moving particles and taken away $N_L$ left-moving particles, that is *added* $N_L$ left-moving holes (aka anti-particles). The axial charge of the state has changed by

$$\Delta Q_A = \Delta(N_L - N_R) = 2\frac{\Delta p}{2\pi/L} = \frac{L}{\pi}\Delta p = \frac{L}{\pi} e \int \mathrm{d}t E_x(t) = \frac{e}{\pi}\int \mathrm{d}t\mathrm{d}x E_x = \frac{e}{2\pi}\int \epsilon_{\mu\nu}F^{\mu\nu}$$

On the other hand, the LHS is $\Delta Q_A = \int \partial^\mu J^A_\mu$. We can infer a local version of this equation by letting $E$ vary slowly in space as well, and we conclude that

$$\partial_\mu J^\mu_A = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu}.$$

This agrees exactly with the anomaly equation in $D = 1+1$ produced by the calculation above in (10.15) (see the homework).

# 11 Effective field theory

## 11.1 A parable on integrating out degrees of freedom

Here's a second parable from QM which gives some useful perspective on renormalization in QFT and on the notion of effective field theory.

[Banks p. 138] Consider a system of two coupled harmonic oscillators. We will assume one of the springs is much stiffer than the other: let's call their natural frequencies $\omega_0, \Omega$, with $\omega_0 \ll \Omega$. The euclidean-time action is

$$S[Q,q] = \int dt \left[ \frac{1}{2}\left(\dot{q}^2 + \omega_0^2 q^2\right) + \frac{1}{2}\left(\dot{Q}^2 + \Omega^2 Q^2\right) + gQq^2 \right] \equiv S_{\omega_0}[q] + S_\Omega[Q] + S_{\text{int}}[Q,q].$$

(The particular form of the $q^2Q$ coupling is chosen for convenience. Don't take too seriously the physics at negative $Q$.) We can construct physical observables in this model by studying the path integral:

$$Z = \int [dQdq] e^{-S[Q,q]}.$$

Since I put a minus sign rather than an $i$ in the exponent (and the potential terms in the action have + signs), this is a euclidean path integral.

Let's consider what happens if we do the path integral over the heavy mode $Q$, and postpone doing the path integral over $q$. This step, naturally, is called *integrating out* $Q$, and we will see below why this is a good idea. The result just depends on $q$; we can think of it as an *effective action* for $q$:

$$\begin{aligned} e^{-S_{\text{eff}}[q]} &:= \int [dQ] e^{-S[q,Q]} \\ &= e^{-S_{\omega_0}[q]} \left\langle e^{-S_{\text{int}}[Q,q]} \right\rangle_Q \end{aligned}$$

Here $\langle ... \rangle_Q$ indicates the expectation value of ... in the (free) theory of $Q$, with the action $S_\Omega[Q]$. It is a gaussian integral (because of our choice of $S_{\text{int}}$:

$$\left\langle e^{-S_{\text{int}}[Q,q]} \right\rangle_Q = \int [dQ] e^{-S_\Omega[Q] - \int ds J(s) Q(s)} = \mathcal{N} e^{\frac{1}{4}\int dsdt J(s) G(s,t) J(t)} \ .$$

This last equality is an application of the 'fundamental theorem of path integrals,' *i.e.* the gaussian integral. Here $J(s) \equiv gq(s)^2$. The normalization factor $\mathcal{N}$ is independent of $J$ and hence of $q$. And $G(s,t)$ is the inverse of the linear operator appearing in $S_\Omega$, the euclidean Green's function:

$$S_\Omega[Q] = \int dsdt Q(s) G^{-1}(s,t) Q(t).$$

More usefully, $G$ satisfies

$$\left(-\partial_s^2 + \Omega^2\right) G(s,t) = \delta(s-t)$$

The fact that our system is time-translation invariant means $G(s,t) = G(s-t)$. We can solve this equation in fourier space: $G(s) = \int \text{đ}\omega e^{-i\omega s} G_\omega$ makes it algebraic:

$$G_\omega = \frac{1}{\omega^2 + \Omega^2}$$

and we have

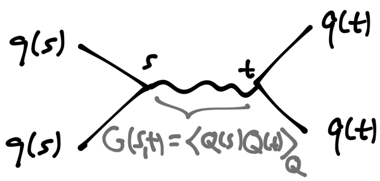
$$G(s) = \int \text{đ}\omega e^{-\mathbf{i}\omega s} \frac{1}{\omega^2 + \Omega^2}. \tag{11.1}$$

So we have:

$$e^{-S_{\text{eff}}[q]} = e^{-S_{\omega_0}[q]} e^{-\int dt ds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2}$$

or taking logs

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt ds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2 \ . \tag{11.2}$$




$Q$ mediates an interaction of four $q$s, an anharmonic term, a self-interaction of $q$. In Feynman diagrams, the leading interaction between $q$'s mediated by $Q$ comes from the diagram at left.

And the whole thing comes from exponentiating disconnected copies of this diagram. There are no other diagrams: once we make a $Q$ from two $q$s what can it do besides turn back into two $q$s? Nothing. And no internal $q$ lines are allowed, they are just sources, for the purposes of the $Q$ integral.

But it is *non-local*: we have two integrals over the time in the new quartic term. This is unfamiliar, and *bad*: *e.g.* classically we don't know how to pose an initial value problem using this action.

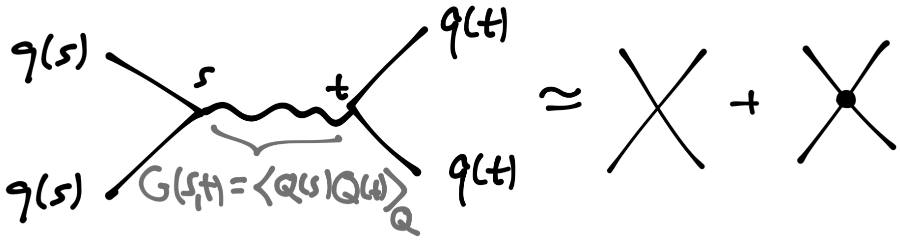
But now suppose we are interested in times much longer than $1/\Omega$, say times comparable to the period of oscillation of the less-stiff spring $2\pi/\omega$. We can accomplish this by Taylor expanding under the integrand in (11.1):

$$G(s) \overset{s \gg 1/\Omega}{\simeq} \int \text{đ}\omega e^{-\mathbf{i}\omega s} \frac{1}{\Omega^2} \underbrace{\frac{1}{1 + \frac{\omega^2}{\Omega^2}}}_{=\sum_n (-1)^n \left(\frac{\omega^2}{\Omega^2}\right)^n} \simeq \frac{1}{\Omega^2}\delta(s) + \frac{1}{\Omega^4}\partial_s^2 \delta(s) + \ldots$$

Plug this back into (11.2):

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt \frac{g^2}{2\Omega^2} q(t)^4 + \int dt \frac{g^2}{2\Omega^4} \dot{q}^2 q^2 + \ldots$$

The effects of the heavy mode $Q$ are now organized in a *derivative expansion*, with terms involving more derivatives suppressed by more powers of the high energy scale $\Omega$.




$$+\cdots \tag{11.3}$$

A useful mnemonic for integrating out the effects of the heavy field in terms of Feynman diagrams: to picture $Q$ as propagating for only a short time (compared to the external time $t-s$), we can contract its propagator to a point. The first term on the RHS shifts the $q^4$ term, the second shifts the kinetic term, the third involves four factors of $\dot{q}$...

On the RHS of this equation, we have various interactions involving four $q$s, which involve increasingly many derivatives. The first term is a quartic potential term for $q$: $\Delta V = \frac{g}{\Omega^2}q^4$; the leading effect of the fluctuations of $Q$ is to *shift* the quartic self-coupling of $q$ by a finite amount (note that we could have included a bare $\lambda_0 q^4$ potential term).

Notice that if we keep going in this expansion, we get terms with *more than two derivatives* of $q$. This is OK. We've just derived the right way to think about such terms: we treat them as a perturbation, and they are part of a never-ending series of terms which become less and less important for low-energy questions. If we want to ask questions about $x$ at energies of order $\omega$, we can get answers that are correct up to effects of order $\left(\frac{\omega}{\Omega}\right)^{2n}$ by keeping the $n$th term in this expansion.

Conversely if we are doing an experiment with precision $\Delta$ at energy $\omega$, we can measure the effects of up to the $n$th term, with

$$\left(\frac{\omega}{\Omega}\right)^{2n} \sim \Delta.$$

Another important lesson: $S_{\text{eff}}[q]$ contains couplings with negative dimensions of energy

$$\sum_n c_n \left(\partial_t^n q\right)^2 q^2, \quad \text{with } c_n \sim \frac{1}{\Omega^{2n}},$$

exactly the situation where the $S$-matrix grows too fast at high energies that we discussed at (8.12). In this case we know exactly where the probability is going: if we have enough energy to see the problem $E \sim \Omega$, we have enough energy to kick the heavy mode $Q$ out of its groundstate.

### 11.1.1 Attempt to consolidate understanding

We've just done some coarse graining: focusing on the dofs we care about ($q$), and actively ignoring the dofs we don't care about ($Q$), except to the extent that they affect those we do (e.g. the self-interactions of $q$).

Above, we did a calculation in a QM model with two SHOs. This is a paradigm of QFT in many ways. For one thing, free quantum fields are bunches of harmonic oscillators with natural frequency depending on $k$, $\Omega = \sqrt{\vec{k}^2 + m^2}$. Here we kept just two of these modes (one with large $k$, one with small $k$) for clarity. Perhaps more importantly, QM is just QFT in 0+1d. The more general QFT path integral just involves more integration variables. The idea of the Wilsonian RG (for continuum field theory) is essentially to do the integrals over the modes in descending order of wavenumber, and at each stage make the expansion described above to get a local action. And notice that basically all possible terms are generated, consistent with the symmetries (here for example, there is a $\mathbb{Z}_2$ symmetry under which $q \to -q$, so there are no odd powers of $q$). Alas, this is all I'll say about it until Physics 217 in Fall 2018.

The result of that calculation was that fluctuations of $Q$ mediate various $q^4$ interactions. It adds to the action for $q$ the following: $\Delta S_{\text{eff}}[q] \sim \int dtds q^2(t) G(t-s) q^2(s)$, as in Fig. 11.3.

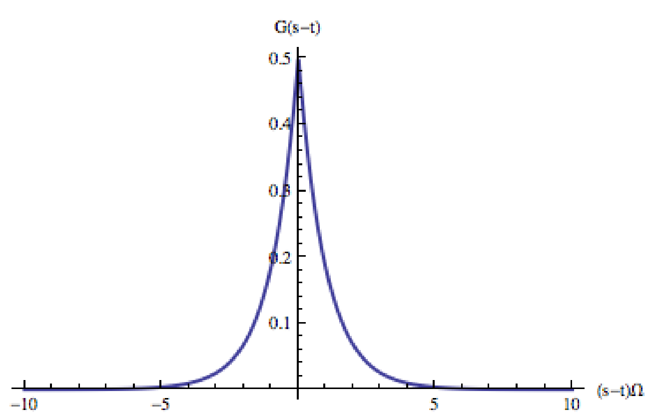
If we have the hubris to care about the exact answer, it's nonlocal in time. But if we want exact answers then we'll have to do the integral over $q$, too. On the other hand, the hierarchy of scales $\omega_0 \ll \Omega$ is useful if we ask questions about energies of order $\omega_0$, e.g.

$$\langle q(t)q(0)\rangle \text{ with } t \sim \frac{1}{\omega_0} \gg \Omega$$

Then we can Taylor expand the function $G(t-s)$, and we find a series of corrections in powers of $\frac{1}{t\Omega}$ (or more accurately, powers of $\frac{\partial_t}{\Omega}$).

(Notice that it's not so useful to integrate out light degrees of freedom to get an action for the heavy degrees of freedom; that would necessarily be nonlocal and stay nonlocal and we wouldn't be able to treat it using ordinary techniques.)

The crucial point is that the scary non-locality of the effective action that we saw only extends a distance of order $\frac{1}{\Omega}$; the kernel $G(s-t)$ looks like this:




The mechanism we've just discussed is an essential ingredient in getting *any* physics done at all. *Why* can we do physics despite the fact that we do not understand the theory of quantum gravity which governs Planckian distances? We happily do lots of physics without worrying about this! This is because the effect of those Planckian quantum gravity fluctuations – whatever they are, call them $Q$ – on the degrees of freedom we do care about (*e.g.* the Standard Model, or an atom, or the sandwich you made this morning, call them collectively $q$) are encoded in terms in the effective action of $q$ which are suppressed by powers of the high energy scale $M_{\text{Planck}}$, whose role in the toy model is played by $\Omega$. And the natural energy scale of your sandwich is much less than $M_{\text{Planck}}$.

I picked the Planck scale as the scale to ignore here for rhetorical drama, and because we really are ignorant of what physics goes on there. But this idea is equally relevant for *e.g.* being able to describe water waves by hydrodynamics (a classical field theory) without worrying about atomic physics, or to understand the physics of atoms without needing to understand nuclear physics, or to understand the nuclear interactions without knowing about the Higgs boson, and so on deeper into the onion of physics.

This wonderful situation, which makes physics possible, has a price: since physics at low energies is so insensitive to high energy physics, it makes it hard to learn about high energy physics! People have been very clever and have learned a lot in spite of this vexing property of the RG called decoupling. We can hope that will continue. (Cosmological inflation plays a similar role in hiding the physics of the early universe. It's like whoever designed this game is *trying* to hide this stuff from us.)

---

The explicit functional form of $G(s)$ (the inverse of the (euclidean) kinetic operator for $Q$) is:

$$G(s) = \int đ\omega \frac{e^{-\mathbf{i}\omega s}}{\omega^2 + \Omega^2} = e^{-\Omega|s|}\frac{1}{2\Omega}. \tag{11.4}$$

Do it by residues: the integrand has poles at $\omega = \pm\mathbf{i}\Omega$. The absolute value of $|s|$ is crucial, and comes from the fact that the contour at infinity converges in the upper (lower) half plane for $s < 0$ ($s > 0$).

---

Next, some comments about ingredients in this discussion, which provide a useful opportunity to review/introduce some important QFT technology:

- Please don't be confused by the formal similarity of the above manipulations with the construction of the generating functional of correlation functions of $Q$:

$$Z[J] \equiv \left\langle e^{\int dt Q(t)J(t)} \right\rangle_Q , \quad \langle Q(t_1)Q(t_2)...\rangle_Q = \frac{\delta}{\delta J(t_{\cdot 1})}\frac{\delta}{\delta J(t_2)}... \log Z[J]$$

[34] It's true that what we did above amounts precisely to constructing $Z[J]$, and plugging in $J = g_0 q^2$. But the motivation is different: in the above $q$ is also a dynamical variable, so we don't get to pick $q$ and differentiate with respect to it; we are merely postponing doing the path integral over $q$ until later.

- Having said that, what is this quantity $G(s)$ above? It is the (euclidean) two-point function of $Q$:

  $$G(s,t) = \langle Q(s)Q(t)\rangle_Q = \frac{\delta}{\delta J(t)}\frac{\delta}{\delta J(s)}\log Z[J].$$

  The middle expression makes it clearer that $G(s,t) = G(s-t)$ since nobody has chosen the origin of the time axis in this problem. This euclidean Green's function, the inverse of $-\partial_\tau^2 + \Omega^2$, is unique, once we demand that it falls off at large separation (unlike the real-time Green's function).

- **Adding more labels.** Quantum mechanics is quantum field theory in 0+1 dimensions. Except for our ability to do all the integrals, everything we are doing here generalizes to quantum field theory in more dimensions: quantum field theory *is* quantum mechanics (with infinitely many degrees of freedom). With more spatial dimensions, we'll want to use the variable $x$ for the spatial coordinates (which are just labels on the fields!) and it was in anticipation of this step that I used $q$ instead of $x$ for my oscillator position variables.

## 11.2 Introduction to effective field theory

[Some nice lecture notes on effective field theory can be found here: J. Polchinski, A. Manohar, I. Rothstein, D. B. Kaplan, H. Georgi. Aneesh Manohar has written an excellent and provocative new set of lectures from last year's Les Houches school which should appear on the arXiv sometime soon.]

**Diatribe about 'renormalizability'.** Taking the example of the previous subsection to its logical conclusion, we are led to the idea of an *effective field theory* (EFT). (The Wilsonian perspective on renormalization – namely that we should include all possible operators consistent with symmetries and let dimensional analysis and the dynamics decide which are important at low energies – makes this idea even more

[34] Functional derivatives are very useful! A reminder: the definition is

$$\frac{\delta J(s)}{\delta J(t)} = \delta(s-t) \tag{11.5}$$

plus the Liebniz properties (linearity, product rule).

inevitable.) There is no reason to demand that a field theory that we have found to describe physics in some regime should be a valid description of the world to arbitrarily short (or long!) distances. This is a happy statement: there can always be new physics that has been so far hidden from us. Rather, an EFT comes with a regime of validity, and with necessary cutoffs. As we will discuss, in a useful implementation of an EFT, the cutoff implies a small parameter in which we can expand (and hence compute). (In the example of $S_{\text{eff}}[q]$ of the previous subsection, the small parameter is $\omega/\Omega$.)

Caring about renormalizibility is pretending to know about physics at arbitrarily short distances. Which you don't.

Even when theories are renormalizable, this apparent victory is often false. For example, QED requires only two independent counterterms (for the mass and for the fine structure constant), and is therefore by the old-fashioned definition renormalizable, but it is superseded by the electroweak theory above 80GeV. Also: the coupling in QED actually increases logarithmically at shorter distances, and ultimately reaches a *Landau pole* at SOME RIDICULOUSLY HIGH ENERGY (of order $e^{+\frac{c}{\alpha}}$ where $\alpha \sim \frac{1}{137}$ is the fine structure constant (*e.g.* at the scale of atomic physics) and $c$ is some numerical number. Plugging in numbers gives something like $10^{330}$ GeV, which is quite a bit larger than the Planck scale). This is of course completely irrelevant for physics and even in principle because of the previous remark about electroweak unification. And if not because of that, because of the Planck scale. A heartbreaking historical fact is that Landau and many other smart people gave up on QFT as a whole because of this silly fantasy about QED in an unphysical regime.

We will see below that even in QFTs which are non-renormalizable in the strict sense, there is a more useful notion of renormalizability: effective field theories come with a small parameter (often some ratio of mass scales), in which we may expand the action. A useful EFT requires a finite number of counterterms *at each order in the expansion.*

Furthermore, I claim that this is *always* the definition of renormalizability that we are using, even if we are using a theory which is renormalizable in the traditional sense, which allows us to pretend that there is no cutoff. That is, there could always be corrections of order $\left(\frac{E}{E_{\text{new}}}\right)^n$ where $E$ is some energy scale of physics that we are doing and $E_{\text{new}}$ is some UV scale where new physics might come in; for large enough $n$, this is too small for us to have seen. The property of renormalizibility that actually matters is that we need a finite number of counterterms at each order in the expansion in $\frac{E}{E_{\text{new}}}$.

Renormalizable QFTs are in some sense *less* powerful than non-renormalizable ones – the latter have the decency to tell us when they are giving the wrong answer! That is, they tell us at what energy new physics *must* come in; with a renormalizable theory we may blithely pretend that it is valid in some ridiculously inappropriate regime like $10^{330}$ GeV.

**Notions of EFT.** There is a dichotomy in the way EFTs are used. Sometimes one knows a lot about the UV theory (*e.g.*

- electroweak gauge theory,
- QCD,
- electrons in a solid,
- water molecules

...) but it is complicated and unwieldy for the questions one wants to answer, so instead one develops an effective field theory involving just the appropriate and important dofs (*e.g.*, respectively,

- Fermi theory of weak interactions,
- chiral lagrangian (or HQET or SCET or ...),
- Landau Fermi liquid theory (or the Hubbard model or a topological field theory or ...),
- hydrodynamics (or some theory of phonons in ice or ...)

...). As you can see from the preceding lists of examples, even a single UV theory can have many different IR EFTs depending on what phase it is in, and depending on what question one wants to ask. The relationship between the pairs of theories above is always coarse-graining from the UV to the IR, though exactly what plays the role of the RG parameter can vary wildly. For example, in the example of the Fermi liquid theory, the scaling is $\omega \to 0$, and momenta scale towards the Fermi surface, not $\vec{k} = 0$.

A second situation is when one knows a description of some low-energy physics up to some UV scale, and wants to try to infer what the UV theory might be. This is a common situation in physics! Prominent examples include: the Standard Model, and quantized Einstein gravity. Occasionally we (humans) actually learn some physics and an example of an EFT from the second category moves to the first category.

**Instructions for EFT.** Answer the following questions:

1. what are the dofs?
2. what are the symmetries?
3. where is the cutoff on its validity?

Then write down all interactions between the dofs which preserve the symmetry in an expansion in derivatives, with higher-dimension operators suppressed by more powers of the UV scale.

I must also emphasize two distinct usages of the term 'effective field theory' which are common, and which the discussion above is guilty of conflating (this (often slippery) distinction is emphasized in the review article by Georgi linked at the beginning of this subsection). The Wilsonian perspective advocated above produces a low-energy description of the physics which is really just a way of solving (if you can) the original model; very reductively, it's just a physically well-motivated order for doing the integrals. If you really integrate out the high energy modes exactly, you will get a non-local action for the low energy modes. This is to be contrasted with the local actions one uses in practice, by truncating the derivative expansion. It is the latter which is really the action of the effective field theory, as opposed to the *full theory*, with some of the integrals done already. The latter will give correct answers for physics below the cutoff scale, and it will give them much more easily.

Some interesting and/or important examples where EFT has been useful (most of which we will not discuss explicitly) and where you can learn about them:

- Hydrodynamics [Kovtun]
- Fermi liquid theory [J. Polchinski, R. Shankar, *Rev. Mod. Phys.* **66** (1994) 129]
- chiral perturbation theory [D. B. Kaplan, §4]
- heavy quark effective theory [D. B. Kaplan, §1.3, Manohar and Wise, *Heavy Quark Physics*]
- random surface growth (KPZ) [Zee, chapter VI]
- color superconductors [D. B. Kaplan, §5]
- gravitational radiation from binary mergers [Goldberger, Rothstein, Porto]
- soft collinear effective theory [Becher, Stewart]

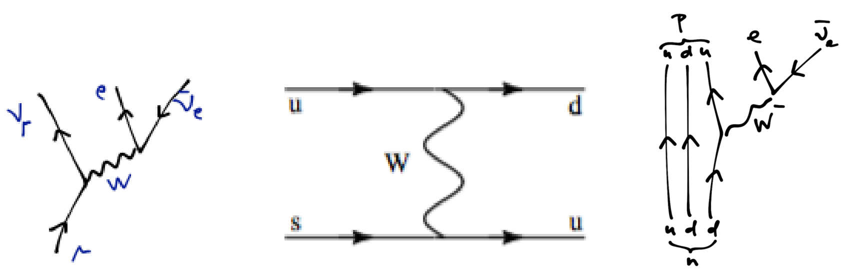
- magnets [Zee, chapter VI.5, hep-ph/9311264v1]
- effective field theory of cosmological inflation [Senatore et al, Cheung et al, Porto]
- effective field theory of dark matter direct detection [Fitzpatrick et al]

There are many others, the length of this list was limited by how long I was willing to spend digging up references. [End of Lecture 34]

## 11.3 Fermi theory of Weak Interactions

[from §5 of A. Manohar's EFT lectures] As a first example, let's think about part of the Standard Model.

$$
\begin{aligned}
L_{EW} \ni & -\frac{1}{2}\left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right)\left(\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}\right) + M_W W_\mu^+ W^{-\mu} \\
& - \frac{\mathbf{i}g}{\sqrt{2}} \bar{\psi}_i \gamma^\mu P_L \psi_j W_\mu^+ V_{ij} \quad + \text{terms involving } Z \text{ bosons}
\end{aligned}
\tag{11.6}
$$




Some things intermediate, off-shell $W$ bosons can do: $\mu$ decay, $\Delta S = 1$ processes, neutron decay

If we are asking questions with external momenta less than $M_W$, we can integrate out $W$ and make our lives simpler:

$$
\delta S_{eff} \sim \left(\frac{\mathbf{i}g}{\sqrt{2}}\right)^2 V_{ij} V_{k\ell}^\star \int đ^D p \frac{-\mathbf{i}g_{\mu\nu}}{p^2 - M_W^2} \left(\bar{\psi}_i \gamma^\mu P_L \psi_j\right)(p) \left(\bar{\psi}_k \gamma^\nu P_L \psi_\ell\right)(-p)
$$

(I am lying a little bit about the $W$ propagator in that I am not explicitly projecting out the fourth polarization with the negative residue. Also hidden in my notation is the fact that the $W$ carries electric charge, so the charges of $\bar{\psi}_i$ and $\psi_j$ in (11.6) must differ by one.) This is non-local at scales $p \gtrsim M_W$ (recall the discussion of the previous subsection). But for $p^2 \ll M_W^2$,

$$\frac{1}{p^2 - M_W^2} \overset{p^2 \ll M_W^2}{\simeq} -\frac{1}{M_W^2}\left(1 + \underbrace{\frac{p^2}{M_W^2} + \frac{p^4}{M_W^4} + \ldots}_{\text{derivative couplings}}\right) \tag{11.7}$$

$$S_F = -\frac{4G_F}{\sqrt{2}} V_{ij} V_{kl}^\star \int d^4x \left(\bar{\psi}_i \gamma^\mu P_L \psi_j\right)(x) \left(\bar{\psi}_k \gamma_\mu P_L \psi_\ell\right)(x) + \mathcal{O}\left(\frac{1}{M_W^2}\right) + \text{kinetic terms for fermions} \tag{11.8}$$

where $G_F/\sqrt{2} \equiv \frac{g^2}{8M_W^2}$ is the Fermi coupling. We can use this (Fermi's) theory to compute the amplitudes above, and it is much simpler than the full electroweak theory (for example I don't have to lie about the form of the propagator of the W-boson like I did above). It was discovered first and used quite effectively long before the existence of $W$s was suspected.

On the other hand, this theory is *not the same* as the electroweak theory; for example it is not renormalizable, while the EW theory is. Its point in life is to help facilitate the expansion in $1/M_W$. There is something about the expression (11.8) that should make you nervous, namely the big red 1 in the $1/M_W^2$ corrections: what makes up the dimensions? This becomes an issue when we ask about ...

## 11.4 Loops in EFT

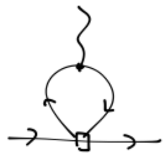
I skipped this subsection in lecture. Skip to §11.5. Suppose we try to define the Fermi theory $S_F$ with a euclidean momentum cutoff $|k_E| < \Lambda$. We expect that we'll have to set $\Lambda \sim M_W$. A simple example which shows that this is problematic is to ask about radiative corrections in the 4-Fermi theory to the coupling between the fermions and the $Z$ (or the photon).

We are just trying to estimate the magnitude of this correction, so don't worry about the factors and the gamma matrices:

$$\sim I \equiv \underbrace{\frac{1}{M_W^2}}_{\propto G_F} \underbrace{\int^\Lambda đ^4k \frac{1}{k}\frac{1}{k} \mathrm{tr}\left(\gamma \ldots\right)}_{\sim \int^\Lambda k\mathrm{d}k \sim \Lambda^2 \sim M_W^2} \sim \mathcal{O}(1).$$

Even worse, consider what happens if we use the vertex coming from the $\left(\frac{p^2}{M_W^2}\right)^\ell$

correction in (11.7)



$$\sim I_\ell \equiv \frac{1}{M_W^2}\int^\Lambda \text{đ}^4 k \frac{1}{k^2}\left(\frac{k^2}{M_W^2}\right)^\ell \sim \mathcal{O}(1)$$

– it's also unsuppressed by powers of ... well, anything. This is a problem.

**Fix:** A way to fix this is to use a "mass-independent subtraction scheme", such as dimensional regularization and minimal subtraction ($\overline{\text{MS}}$). The crucial feature is that the dimensionful cutoff parameter appears only inside logarithms ($\log \mu$), and not as free-standing powers ($\mu^2$).

With such a scheme, we'd get instead

$$I \sim \frac{m^2}{M_W^2}\log\mu \qquad I_\ell \sim \left(\frac{m^2}{M_W^2}\right)^{\ell+1}\log\mu$$

where $m$ is some mass scale *other* than the RG scale $\mu$ (like a fermion mass parameter, or an external momentum, or a dynamical scale like $\Lambda_{QCD}$).

We will give a more detailed example next. The point is that in a mass-independent scheme, the regulator doesn't produce new dimensionful things that can cancel out the factors of $M_W$ in the denominator. It respects the 'power counting': if you see $2\ell$ powers of $1/M_W$ in the coefficient of some term in the action, that's how many powers will suppress its contributions to amplitudes. This means that the EFT is like a renormalizable theory *at each order in the expansion* (here in $1/M_W$), in that there is only a finite number of allowed vertices that contribute at each order (counterterms for which need to be fixed by a renormalization condition). The insatiable appetite for counterterms is still insatiable, but it eats only a finite number at each order in the expansion. Eventually you'll get to an order in the expansion that's too small to care about, at which point the EFT will have eaten only a finite number of counterterms.

There is a price for these wonderful features of mass-independent schemes, which has two aspects:

- Heavy particles (of mass $m$) don't decouple when $\mu < m$. For example, in a mass-independent scheme for a gauge theory, heavy charged particles contribute to the beta function for the gauge coupling even at $\mu \ll m$.
- Perturbation theory will break down at *low* energies, when $\mu < m$; in the example just mentioned this happens because the coupling keeps running.

We will show both these properties very explicitly in the next subsection. The solution of both these problems is to integrate out the heavy particles by hand at $\mu = m$, and make a new EFT for $\mu < m$ which simply omits that field. Processes for which we should set $\mu < m$ don't have enough energy to make the heavy particles in external states anyway. (For some situations where you should still worry about them, see Aneesh Manohar's notes linked above.)

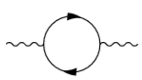
### 11.4.1 Comparison of schemes, case study

The case study we will make is the contribution of a charged fermion of mass $m$ to the running of the QED gauge coupling.

Recall that the QED Lagrangian is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \bar{\psi}\left(\mathbf{i}\not{D} - m\right)\psi$$

with $D_\mu = \partial_\mu - \mathbf{i}eA_\mu$. By redefining the field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ by a constant factor we can move around where the $e$ appears, *i.e.* by writing $\tilde{A} = eA$, we can make the gauge kinetic term look like $\frac{1}{4e^2}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$. This means that the charge renormalization can be seen *either* in the vacuum polarization, the correction to the photon propagator:



. I will call this diagram $\mathbf{i}\Pi_{\mu\nu}$.

So the information about the running of the coupling is encoded in the gauge field two-point function:

$$\Pi_{\mu\nu} \equiv \langle A_\mu(p)A_\nu(q)\rangle = \left(p_\mu p_\nu - p^2 g_{\mu\nu}\right)\not{\delta}(p+q)\Pi(p^2)\ .$$

The factor $P_{\mu\nu} \equiv p_\mu p_\nu - p^2 g_{\mu\nu}$ is guaranteed to be the polarization structure by the gauge invariance Ward identity: $p^\mu \langle A_\mu(p)A_\nu(q)\rangle = 0$. That is: $p^\mu P_{\mu\nu} = 0$, and there is no other symmetric tensor made from $p^\mu$ which satisfies this. This determines the correlator up to a function of $p^2$, which we have called $\Pi(p^2)$.

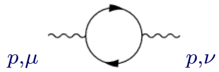
The choice of scheme shows up in our choice of renormalization condition to impose on $\Pi(p^2)$:

**Mass-dependent scheme:** subtract the value of the graph at $p^2 = -M^2$ (a very off-shell, euclidean, momentum). That is, we impose a renormalization condition which says

$$\Pi(p^2 = -M^2) \overset{!}{=} 1 \tag{11.9}$$

(which is the tree-level answer with the normalization above).

The contribution of a fermion of mass $m$ and charge $e$ is (factoring out the momentum-conserving delta function):




$$= -\int đ^D k \mathrm{tr}\left( (-\mathbf{i}e\gamma^\mu) \frac{-\mathbf{i}\,(\not{k}+m)}{k^2-m^2} (-\mathbf{i}e\gamma^\nu) \frac{-\mathbf{i}\,(\not{p}+\not{k}+m)}{(p+k)^2-m^2} \right)$$

The minus sign out front is from the fermion loop. Some boiling, which you can find in Peskin (page 247) or Zee (§III.7), reduces this to something manageable. The steps involved are: (1) a trick to combine the denominators, like the Feynman trick $\frac{1}{AB} = \int_0^1 dx \left(\frac{1}{(1-x)A+xB}\right)^2$. (2) some Dirac algebra, to turn the numerator into a polynomial in $k, p$. As Zee says, our job in this course is not to train to be professional integrators. The result of this boiling can be written

$$\mathbf{i}\Pi^{\mu\nu} = -e^2 \int đ^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{\left(\ell^2-\Delta\right)^2}$$

with $\ell = k + xp$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)p^2$, and the numerator is

$$N^{\mu\nu} = 2\ell^\mu\ell^\nu - g^{\mu\nu}\ell^2 - 2x(1-x)p^\mu p^\nu + g^{\mu\nu}\left(m^2 + x(1-x)p^2\right) + \text{terms linear in } \ell^\mu \quad .$$

---

In dim reg, the one-loop vacuum polarization correction satisfies the gauge invaraince Ward identity $\Pi^{\mu\nu} = P^{\mu\nu}\delta\Pi_2$ (unlike the euclidean momentum cutoff which is not gauge invariant). A peek at the tables of dim reg integrals shows that $\delta\Pi_2$ is:

$$\begin{aligned} \delta\Pi_2(p^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left( \frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right) \right) \end{aligned} \tag{11.10}$$

where we have introduced the heralded $\mu$:

$$\mu^2 \equiv 4\pi\bar{\mu}^2 e^{-\gamma_E}$$

where $\gamma_E$ is the Euler-Mascheroni constant. In the second line of (11.10), we expanded the $\Gamma$-function about $D=4$; there are other singularities at other integer dimensions.

**Mass-dependent scheme:** Now back to our discussion of schemes. I remind you that in a mass-independent scheme, we demand that the counterterm cancels $\delta\Pi_2$ when we set the external momentum to $p^2 = -M^2$, so that the whole contribution at order $e^2$ is :

$$0 \overset{(11.9)!}{=} \Pi_2^{(M)}(p^2 = -M^2) = \underbrace{\delta_{F^2}^{(M)}}_{\text{counterterm coefficient for } \frac{1}{4}F_{\mu\nu}F^{\mu\nu}} + \delta\Pi_2$$

$$\implies \Pi_2^{(M)}(p^2) = \frac{e^2}{2\pi^2}\int dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{m^2 + x(1-x)M^2}\right).$$

Notice that the $\mu$s go away in this scheme.

**Mass-Independent scheme:** This is to be contrasted with what we get in a mass-independent scheme, such as $\overline{\text{MS}}$, in which $\Pi$ is defined by the rule that we *subtract the* $1/\epsilon$ *pole*. This means that the counterterm is

$$\delta_{F^2}^{(\overline{\text{MS}})} = -\frac{e^2}{2\pi^2}\frac{2}{\epsilon}\underbrace{\int_0^1 dx x(1-x)}_{=1/6}.$$

(Confession: I don't know how to state this in terms of a simple renormalization condition on $\Pi_2$. Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and $\mu$.) The resulting vacuum polarization function is

$$\Pi_2^{(\overline{\text{MS}})}(p^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{\mu^2}\right).$$

Next we will talk about beta functions, and verify the claim above about the failure of decoupling. First let me say some words about what is failing. What is failing – the price we are paying for our power counting – is the basic principle of the RG, namely that physics at low energies shouldn't care about physics at high energies, except for small corrections to couplings. An informal version of this statement is: you don't need to know about nuclear physics to make toast. A more formal version is the *Appelquist-Carazzone Decoupling Theorem*, which I will not state (Phys. Rev. D11, 28565 (1975)). So it's something we must and will fix.

**Beta functions.** $\boxed{M}$: First in the mass-dependent scheme. Demanding that physics is independent of our made-up RG scale, we find

$$0 = M\frac{d}{dM}\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)} e\frac{\partial}{\partial e}\right)\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)}\underbrace{\cdot 2}_{\text{to this order}}\right)\Pi_2^{(M)}(p^2)$$

where I made the high-energy physics definition of the beta function[35]:

$$\beta_e^{(M)} \equiv \frac{1}{e}\left(M\partial_M e\right) = -\frac{\partial_\ell e}{e}, \quad M \equiv e^{-\ell}M_0 .$$

Here $\ell$ is the RG time again, it grows toward the IR. So we find

$$\beta_e^{(M)} = -\frac{1}{2}\left(\frac{e^2}{2\pi}\right)\int_0^1 dx x(1-x)\left(\frac{-2M^2x(1-x)}{m^2 + M^2x(1-x)}\right) + \mathcal{O}(e^3)$$

[35]I've defined these beta functions to be dimensionless, *i.e.* they are $\partial_{\log M}\log(g)$; this convention is not universally used.

$$\begin{cases} \overset{m \ll M}{\simeq} \frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) = \frac{e^2}{12\pi^2} \\ \\ \overset{m \gg M}{\simeq} \frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \frac{M^2 x(1-x)}{m^2} = \frac{e^2}{60\pi^2} \frac{M^2}{m^2} \end{cases} . \tag{11.11}$$

$$\boxed{\overline{\text{MS}}}: \quad 0 = \mu \frac{d}{d\mu} \Pi_2^{(\overline{\text{MS}})}(p^2) = \left( \mu \frac{\partial}{\partial \mu} + \beta_e^{(\overline{\text{MS}})} e \frac{\partial}{\partial e} \right) \Pi_2^{(\overline{\text{MS}})}(p^2) = \left( \mu \frac{\partial}{\partial \mu} + \beta_e^{(\overline{\text{MS}})} \underbrace{\cdot 2}_{\text{to this order}} \right) \Pi_2^{(\overline{\text{MS}})}(p^2)$$

$$\begin{aligned} \implies \beta_e^{(\overline{\text{MS}})} &= -\frac{1}{2} \frac{e^2}{2\pi^2} \underbrace{\int_0^1 dx x(1-x)}_{=1/6} \underbrace{\mu \partial_\mu \log \frac{m^2 - p^2 x(1-x)}{\mu^2}}_{=-2} \\ &= \frac{e^2}{12\pi^2}. \end{aligned} \tag{11.12}$$

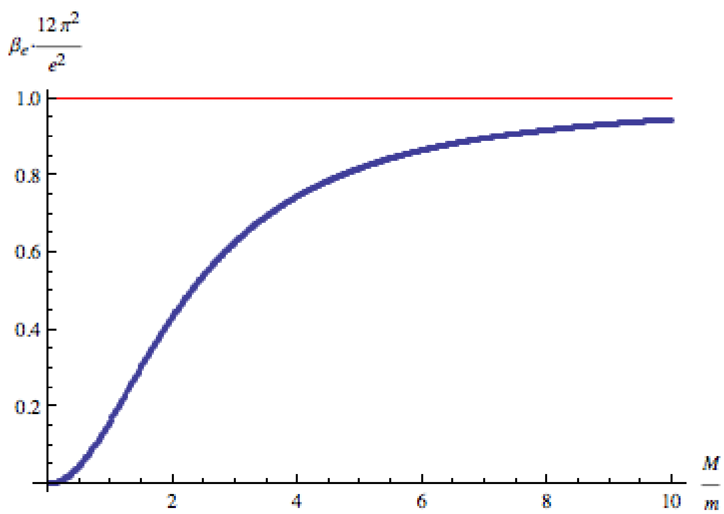
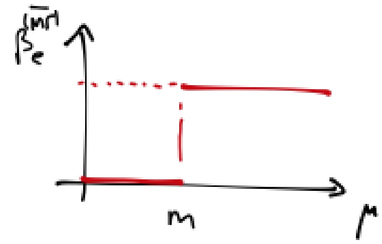



**Figure 2:** The blue curve is the mass-dependent-scheme beta function; at scales $M \ll m$, the mass of the heavy fermion, the fermion sensibly stops screening the charge. The red line is the $\overline{\text{MS}}$ beta function, which is just a constant, pinned at the UV value.

Also, the $\overline{\text{MS}}$ vacuum polarization behaves for small external momenta like

$$\Pi_2(p^2 \ll m^2) \simeq -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \underbrace{\log \frac{m^2}{\mu^2}}_{\gg 1, \text{for } \mu \ll m! \text{ bad!}}$$

As I mentioned, the resolution of both these problems is simply to define a new EFT for $\mu < m$ which omits the heavy field. Then the strong coupling problem goes away and the heavy fields do decouple. The price is that we have to do this by hand, and the beta function jumps at $\mu = m$; the coupling is continuous, though.




## 11.5 The Standard Model as an EFT.

**The Standard Model.** [Schwartz, §29]

| | $L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$ | $e_R$ | $\nu_R$ | $Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$ | $u_R$ | $d_R$ | $H$ |
|---|---|---|---|---|---|---|---|
| $\mathsf{SU}(3)$ | - | - | - | $\square$ | $\square$ | $\square$ | - |
| $\mathsf{SU}(2)$ | $\square$ | - | - | $\square$ | - | - | $\square$ |
| $\mathsf{U}(1)_Y$ | $-\frac{1}{2}$ | $-1$ | $0$ | $\frac{1}{6}$ | $\frac{2}{3}$ | $-\frac{1}{3}$ | $\frac{1}{2}$ |

**Table 1:** The Standard Model fields and their quantum numbers under the gauge group. $\square$ indicates fundamental representation, - indicates singlet. Except for the Higgs, each column is copied three times; each copy is called a *generation*. Except for the Higgs all the matter fields are Weyl fermions of the indicated handedness. Gauge fields as implied by the gauge groups. (Some people might leave out the right-handed neutrino, $\nu_R$.)

Whence the values of the charges under the $\mathsf{U}(1)$ ("hypercharge")? The condition $Y_L + 3Y_Q = 0$ (where $Y$ is the hypercharge) is required by anomaly cancellation. This implies that electrons and protons $p = \epsilon_{ijk} u_i u_j d_k$ have exactly opposite charges of the same magnitude.

The Lagrangian is just all the terms which are invariant under the gauge group $\mathsf{SU}(3) \times \mathsf{SU}(2) \times \mathsf{U}(1)$ with dimension less than or equal to four – all renormalizable terms. This includes a potential for the Higgs, $V(|H|) = m_H^2|H|^2 + \lambda|H|^4$, where it turns out that $m_H^2 \leq 0$. The resulting Higgs vacuum expectation value breaks the Electroweak part of the gauge group

$$\mathsf{SU}(2) \times \mathsf{U}(1)_Y \overset{\langle H \rangle}{\rightsquigarrow} \mathsf{U}(1)_{EM}.$$

The broken gauge bosons get masses from the Higgs kinetic term

$$|D_\mu H|^2|_{H = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}} \quad \text{with} \quad D_\mu H = \left( \partial_\mu - \mathbf{i} g W_\mu^a \tau^a - \frac{1}{2} \mathbf{i} g' Y_\mu \right) H$$

where $Y_\mu$ is the hypercharge gauge boson, and $W^a, a = 1,2,3$ are the $\mathsf{SU}(2)$ gauge bosons. There are two massive $W$-bosons with electric charge $\pm 1$ (as described in §11.3), with $M_W = \frac{vg}{2}$. The photon and $Z$ boson are the linear combinations of $Y$ and $W^3$ which diagonalize the remaining mass terms:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W^3_\mu \\ Y_\mu \end{pmatrix}.$$

Here $\tan\theta_w \equiv \frac{g'}{g}$ defines the Weinberg angle. The masses are $M_\gamma = 0$ and $M_Z = \frac{M_W}{\cos\theta_w} < M_W$.

Fermion masses come from (dimension-four) Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}} = -Y^\ell_{ij} \bar{L}_i H e^j_R - Y^u_{ij} \bar{Q}^i H d^j_R - Y^d_{ij} \bar{Q}^i \left(\mathbf{i}\tau^2 H^\star\right) u^j_R + h.c.$$

The contortion with the $\tau^2$ is required to make a hypercharge invariant. Plugging in the Higgs vev to *e.g.* the lepton terms gives $-m_e \bar{e}_L e_R + h.c.$ with $m_e = y_e v/\sqrt{2}$. There's lots of drama about the matrices $Y$ which can mix the generations. the mass for the $\nu_R$ (which maybe could not exist – it doesn't have any charges at all) you'll figure out on the homework.

Here is a useful mnemonic for remembering the table of quantum numbers (possibly it is more than that): There are larger simple Lie groups that contain the SM gauge group as subgroups:

$$\begin{array}{ccccc} \mathsf{SU}(3) \times \mathsf{SU}(2) \times \mathsf{U}(1)_Y & \subset & \mathsf{SU}(5) & \subset & \mathsf{SO}(10) \\ \text{one generation} & = & 10 \oplus \bar{5} \oplus 1 & = & 16 \end{array}$$

The singlet of $\mathsf{SU}(5)$ is the right-handed neutrino, but if we include it, one generation is an irreducible (spinor) representation of $\mathsf{SO}(10)$. This idea is called grand unification. It is easy to imagine that the gauge group is actually the larger groups on the right, and another instance of the Higgs mechanism accomplishes the breaking down to the Standard Model. (The running of the respective gauge couplings go in the right direction with approximately the right rate to unify to a single value at $M_{GUT} \sim 10^{16} GeV$.) Notice that this idea means leptons and quarks are in the same representations – they can turn into each other. This predicts that the proton should not be perfectly stable. Next we'll say more about this.

**Beyond the Standard Model with EFT.** At what energy does the Standard Model stop working? Because of the annoying feature of renormalizibility, it doesn't tell us. However, we have experimental evidence against a cutoff on the Standard Model (SM) at energies less than something like 10 TeV. The evidence I have in mind is the absence of interactions of the form

$$\delta L = \frac{1}{M^2} \left(\bar{\psi} A \psi\right) \cdot \left(\bar{\psi} B \psi\right)$$

(where $\psi$ represent various SM fermion fields and $A, B$ can be various gamma and flavor matrices) with $M \lesssim 10$ TeV. Notice that I am talking now about interactions *other* than the electroweak interactions, which as we've just discussed, for energies above $M_W \sim 80$GeV cannot be treated as contact interactions – you can see the $W$s propagate!

If such operators were present, we would have found different answers for experiments at LEP. But such operators would be present if we consider new physics in addition to the Standard Model (in most ways of doing it) at energies less than 10 TeV. For example, many interesting ways of coupling in new particles with masses that make them accessible at the LHC would have generated such operators.

A little more explicitly: the Standard Model Lagrangian $L_0$ contains all the renormalizable (*i.e.* engineering dimension $\leq 4$) operators that you can make from its fields (though the coefficients of the dimension 4 operators do vary through quite a large range, and the coefficients of the two relevant operators – namely the identity operator which has dimension zero, and the Higgs mass, which has engineering dimension two, are strangely small, and so is the QCD $\theta$ angle).

To understand what lies beyond the Standard Model, we can use our knowledge that whatever it is, it is probably heavy (it could also just be very weakly coupled, which is a different story), with some intrinsic scale $\Lambda_{\text{new}}$, so we can integrate it out and include its effects by corrections to the Standard Model:

$$L = L_0 + \frac{1}{\Lambda_{\text{new}}}\mathcal{O}^{(5)} + \frac{1}{\Lambda_{\text{new}}^2}\sum_i c_i \mathcal{O}_i^{(6)} + \cdots$$

where the $\mathcal{O}$s are made of SM fields, and have the indicated engineering dimensions, and preserve the necessary symmetries of the SM (Lorentz symmetry and gauge invariance).

In fact there is only one kind of operator of dimension 5 meeting these demands:

$$\mathcal{O}^{(5)} = c_5 \epsilon_{ij} \left(\bar{L}^c\right)^i H^j \epsilon_{kl} L^k H^l$$

where $H^i = (h^+, h^0)^i$ is the $\text{SU}(2)_{EW}$ Higgs doublet and $L^i = (\nu_L, e_L)^i$ is an $\text{SU}(2)_{EW}$ doublet of left-handed leptons, and $\bar{L}^c \equiv L^T C$ where $C$ is the charge conjugation matrix. (I say 'kind of operator' because we can have various flavor matrices in here.) On the problem set you get to see from whence such an operator might arise, and what it does if you plug in the higgs vev $\langle H \rangle = (0, v)$. This term violates lepton number symmetry ($L \to e^{\mathbf{i}\alpha_L} L, Q \to Q, H \to H$).

At dimension 6, there are operators that directly violate baryon number, such as

$$\epsilon_{\alpha\beta\gamma}(\bar{u}_R)^c_\alpha (u_R)_\beta \, (\bar{u}_R)^c_\gamma \, e_R.$$

You should read the above tangle of symbols as '$qqq\ell$' – it turns three quarks into a lepton. The epsilon tensor makes a color SU(3) singlet; this thing $\epsilon qqq$ has the quantum numbers of a baryon, such as the proton and neutron. The long lifetime of the proton (you can feel it in your bones – see Zee p. 413) then directly constrains the scale of new physics appearing in front of this operator.

Two more comments about this:

- If we didn't know about the Standard Model, (but after we knew about QM and GR and EFT (the last of which people didn't know before the SM for some reason)) we should have made the estimate that dimension-5 Planck-scale-suppressed operators like $\frac{1}{M_{\text{Planck}}} p\mathcal{O}$ would cause proton decay (into whatever $\mathcal{O}$ makes). This predicts $\Gamma_p \sim \frac{m_p^3}{M_{\text{Planck}}^2} \sim 10^{-13} s^{-1}$ which is *not* consistent with our bodies not glowing. Actually it is a remarkable fact that there are no gauge-invariant operators made of SM fields of dimension less than 6 that violate baryon number symmetry ($L \to L, Q \to e^{\mathbf{i}\alpha_B} Q, H \to H$). This is an *emergent* symmetry, expected to be violated by the UV completion.
- Surely nothing can prevent $\Delta L \sim \left(\frac{1}{M_{\text{Planck}}}\right)^2 qqq\ell$. Happily, this is consistent with the observed proton lifetime.

There are $\sim 10^2$ dimension 6 operators that preserve baryon number, and therefore are not as tightly constrained[36]. (Those that induce flavor-changing processes in the SM are more highly constrained and must have $\Lambda_{\text{new}} > 10^4$ TeV.) Two such operators are considered equivalent if they differ by something which vanishes by the tree-level SM equations of motion. This is the right thing to do, even for off-shell calculations (like green's functions and for fields running in loops). You know this from a previous problem set: the EOM are true as operator equations – Ward identities resulting from being free to change integration variables in the path integral[37]. [End of Lecture 35]

[36] Recently, humans have gotten better at counting these operators. See this paper.

[37] There are a few meaningful subtleties here, as you might expect if you recall that the Ward identity is only true up to contact terms. The measure in the path integral can produce a Jacobian which renormalizes some of the couplings; the changes in source terms will drop out of S-matrix elements (recall our discussion of changing field variables in the Consequences of Unitarity section.) but can change the form of Green's functions. For more information on the use of eom to eliminate redundant operators in EFT, see Arzt, hep-ph/9304230 and Georgi, "On-Shell EFT".

## 11.6 The color of the sky

[from hep-ph/9606222 and nucl-th/0510023] Why is the sky blue? Basically, it's because the blue light from the sun scatters in the atmosphere more than the red light, and you (I hope) only look at the scattered light.

Here is an understanding of this fact using the EFT logic. Consider the scattering of photons off atoms (in a gas) at low energies. Low energy means that the photon does not have enough energy to probe the substructure of the atom – it can't excite the electrons or the nuclei. This means that the atom is just a particle, with some mass $M$.

The dofs are just the photon field and the field that creates an atom.

The symmetries are Lorentz invariance and charge conjugation invariance and parity. We'll use the usual redundant description of the photon which has also gauge invariance.

The cutoff is the energy $\Delta E$ that it takes to excite atomic energy levels we've left out of the discussion. We allow no inelastic scattering. This means we require

$$E_\gamma \ll \Delta E \sim \frac{\alpha}{a_0} \ll a_0^{-1} \ll M_{\text{atom}} \tag{11.13}$$

Because of this separation of scales, we can also ignore the recoil of the atom, and treat it as infinitely heavy.

Since there are no charged objects in sight – atoms are neutral – gauge invariance means the Lagrangian can depend on the field strength $F_{\mu\nu}$. Let's call the field which destroys an atom with velocity $v$ $\phi_v$. $v^\mu v_\mu = 1$ and $v_\mu = (1,0,0,0)_\mu$ in the atom's rest frame. The (Lorentz-singlet) Lagrangian can depend on $v^\mu$. We can write a Lagrangian for the free atoms as

$$L_{\text{atom}} = \phi_v^\dagger \mathbf{i} v^\mu \partial_\mu \phi_v \quad .$$

This action is related by a boost to the statement that the atom at rest has zero energy – in the rest frame of the atom, the eom is just $\partial_t \phi_{v=(1,\vec{0})} = 0$. (If we didn't define the zero of energy to be at the rest mass, there would be an additional term $M_{\text{atom}}\phi_v^\dagger\phi_v$.) Notice that the kinetic term $\phi_v^\dagger \frac{\vec{\nabla}^2}{2M_{\text{atom}}}\phi_v$ is a very small correction given our hierarchy of scales (11.13).

So the Lagrangian density is

$$L_{\text{Maxwell}}[A] + L_{\text{atom}}[\phi_v] + L_{\text{int}}[A, \phi_v]$$

and we must determine $L_{\text{int}}$. It is made from local, Hermitian, gauge-invariant, Lorentz invariant operators we can construct out of $\phi_v, F_{\mu\nu}, v_\mu, \partial_\mu$ (it can only depend on $F_{\mu\nu} =$

$\partial_\mu A_\nu - \partial_\nu A_\mu$, and not $A_\mu$ directly, by gauge invariance, because the atom, and hence $\phi_v$, is neutral.). It should actually only depend on the combination $\phi_v^\dagger \phi_v$ since we will not create and destroy atoms. (Notice that we didn't have to specify the statistics of the atoms or $\phi_v$.) Therefore

$$L_{\text{int}} = c_1 \phi_v^\dagger \phi_v F_{\mu\nu} F^{\mu\nu} + c_2 \phi_v^\dagger \phi_v v^\sigma F_{\sigma\mu} v_\lambda F^{\lambda\mu} + c_3 \phi_v^\dagger \phi_v \left(v^\lambda \partial_\lambda\right) F_{\mu\nu} F^{\mu\nu} + \ldots$$

$\ldots$ indicates terms with more derivatives and more powers of velocity (*i.e.* an expansion in $\partial \cdot v$). Which are the most important terms at low energies? Demanding that the Maxwell term dominate, we get the power counting rules (so time and space should scale the same way):

$$[\partial_\mu] = 1, \quad [F_{\mu\nu}] = 2$$

This then implies $[\phi_v] = 3/2, [v] = 0$ and therefore

$$[c_1] = [c_2] = -3, [c_3] = -4 \ .$$

Terms with more partials are more irrelevant.

What makes up these dimensions? They must come from the length scales that we have integrated out to get this description – the size of the atom $a_0 \sim \alpha m_e$ and the energy gap between the ground state and the electronic excited states $\Delta E \sim \alpha^2 m_e$. For $E_\gamma \ll \Delta E, a_0^{-1}$, we can just keep the two leading terms.

In the rest frame of the atom, these two leading terms $c_{1,2}$ represent just the scattering of $E$ and $B$ respectively. To determine their coefficients one would have to do a matching calculation to a more complete theory (compute transition rates in a theory that does include extra energy levels of the atom). But a reasonable guess is just that the scale of new physics (in this case atomic physics) makes up the dimensions: $c_1 \simeq c_2 \simeq a_0^3$. (In fact the magnetic term $c_2$ comes with extra factor of $v/c$ which suppresses it.) The scattering cross section then goes like $\sigma \sim c_i^2 \sim a_0^6$; dimensional analysis ($[\sigma] = -2$ is an area, $[a_0^6] = -6$) then tells us that we have to make up four powers with the only other scale around:

$$\sigma \propto E_\gamma^4 a_0^6.$$

(The factor of $E_\gamma^2$ in the amplitude arises from $\vec{E} \propto \partial_t \vec{A}$.) Blue light, which has about twice the energy of red light, is therefore scattered 16 times as much.

The leading term that we left out is the one with coefficient $c_3$. The size of this coefficient determines when our approximations break down. We might expect this to come from the next smallest of our neglected scales, namely $\Delta E$. That is, we expect

$$\sigma \propto E_\gamma^4 a_0^6 \left(1 + \mathcal{O}\left(\frac{E_\gamma}{\Delta E}\right)\right).$$

The ratio in the correction terms is appreciable for UV light.

## 11.7 Pions

[Schwartz §28.1] Below the scale of electroweak symmetry breaking, we can forget the $W$ and $Z$ bosons. Besides the 4-Fermi interactions, the remaining drama is QCD and electromagnetism:

$$\mathcal{L}_{QCD_2} = -\frac{1}{4}F_{\mu\nu}^2 + \mathbf{i}\sum_{\alpha=L,R}\sum_f \bar{q}_{\alpha f} \not{D} q_{\alpha f} - \bar{q}Mq.$$

Here $f$ is a sum over quark flavors, which includes the electroweak doublets, $u$ and $d$. Let's focus on just these two lightest flavors, $u$ and $d$. We can diagonalize the mass matrix by a field redefinition (this is what makes the CKM matrix meaningful): $M = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix}$. If it were the case that $m_u = m_d$, we would have *isospin* symmetry

$$\begin{pmatrix} u \\ d \end{pmatrix} \to U \begin{pmatrix} u \\ d \end{pmatrix}, \quad U \in \mathsf{SU}(N_f = 2).$$

If, further, there were no masses $m = 0$, then $L$ and $R$ decouple and we also have chiral symmetry, $q \to e^{\mathbf{i}\gamma_5\alpha}q$, *i.e.*

$$q_L \to Vq_L, q_R \to V^{-1}q_R, \; V \in \mathsf{SU}(N_f = 2).$$

Why do I restrict to $\mathsf{SU}(2)$ and not $\mathsf{U}(2)$? The central bit of the axial symmetry $\mathsf{U}(1)_A$ is anomalous – its divergence is proportional to the *gluon* theta term operator $F \wedge F$, which has all kinds of nonzero matrix elements. It's not a symmetry (see Peskin page 673 for more detail). The missing non-Goldstone boson is called the $\eta'$. The central bit of the vectorlike transformation $q \to e^{\mathbf{i}\alpha}q$ is *baryon number*, $B$. (Actually this is anomalous under the full electroweak symmetry, but $B - L$ is not).

The groundstate of QCD is mysterious, because of infrared slavery. Here's one piece of input from experiment and numerical simulation. Apparently it is the case that in the groundstate

$$\langle \bar{q}_f q_f \rangle = V^3 \tag{11.14}$$

independent of flavor $f$. This condensate spontaneously breaks

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R \to \mathsf{SU}(2)_{\text{isospin}}, \tag{11.15}$$

the diagonal combination. $\begin{pmatrix} u \\ d \end{pmatrix}$ is a doublet. Since $p = u_\alpha u_\beta d_\gamma \epsilon_{\alpha\beta\gamma}, n = u_\alpha d_\beta d_\gamma \epsilon_{\alpha\beta\gamma}$, this means that $\begin{pmatrix} p \\ n \end{pmatrix}$ is also a doublet. This symmetry is (explicitly) weakly broken by the difference of the masses $m_d = 4.7\text{MeV} \neq m_u = 2.15\text{MeV}$ and by the electromagnetic interactions, since $q_d = -1/3 \neq q_u = 2/3$.

This symmetry-breaking structure enormously constrains the dynamics of the color singlets which are the low-energy excitations above the QCD vacuum (hadrons). Let us use the EFT strategy. We know that the degrees of freedom must include (pseudo-)Goldstone bosons for the symmetry breaking (11.15) ('pseudo' because of the weak explicit breaking).

**Effective field theory.** Since QCD is strongly coupled in this regime, let's use the knowing-the-answer trick: the low energy theory must include some fields which represent the breaking of the symmetry (11.15). One way to do this is to introduce a field $\Sigma$ which transforms like

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R : \Sigma \to g_L \Sigma g_R^\dagger, \quad \Sigma^\dagger \to g_R \Sigma^\dagger g_L^\dagger$$

(this will be called a *linear* sigma model, because $\Sigma$ transforms linearly) – we have in mind $\bar{q}_\alpha q_\beta \sim \Sigma_{\alpha\beta}$. We can make singlets (hence an action) out of $\Sigma_{ij}\Sigma^\dagger_{ji} = \text{tr}\Sigma\Sigma^\dagger \equiv |\Sigma|^2$:

$$\mathcal{L} = |\partial_\mu \Sigma|^2 + m^2|\Sigma|^2 - \frac{\lambda}{4}|\Sigma|^4 + \cdots \tag{11.16}$$

which is designed to have a minimum at $\langle\Sigma\rangle = \frac{V}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, with $V = 2m/\sqrt{\lambda}$ (here $V$ is from (11.14)), which preserves $\mathsf{SU}(2)_{\text{isospin}}$ (under which $\Sigma \to g\Sigma g^\dagger$). We can parametrize the fluctuations about this configuration as

$$\Sigma(x) = \frac{V + \sigma(x)}{\sqrt{2}} e^{\frac{2\mathbf{i}\pi^a(x)\tau^a}{F_\pi}}$$

where $F_\pi = V = \frac{2m}{\sqrt{\lambda}}$ is be chosen to give $\pi^a(x)$ canonical kinetic terms. The $\pi^a$ parametrize the directions of field space in which the potential is flat (like the field $\theta$ in the discussion of the Mexican hat in §9.1). Under $g_{L/R} = e^{\mathbf{i}\theta_{L/R}\tau^a}$, the pion field transforms as

$$\pi^a \to \pi^a + \underbrace{\frac{F_\pi}{2}\left(\theta^a_L - \theta^a_R\right)}_{\text{nonlinear realization of } \mathsf{SU}(2)_{\text{axial}}} - \underbrace{\frac{1}{2}f^{abc}\left(\theta^a_L + \theta^a_R\right)\pi^c}_{\text{linear realiz'n (adj rep) of } \mathsf{SU}(2)_{\text{isospin}}}.$$

The fields $\pi^\pm, \pi^0$ create pions, they transform in the adjoint representation of the diagonal $\mathsf{SU}(2)_{\text{isospin}}$, and they *shift* under the broken symmetry. This shift symmetry forbids mass terms $\pi^2$. The radial excitation $\sigma$, on the other hand, is a fiction which we've introduced in (11.16), and which has no excuse to stick around at low energies (and does not). We can put it out of its misery by taking $m \to \infty, \lambda \to \infty$ fixing $F_\pi$.

In the limit, the useful field to use is

$$U(x) \equiv \frac{\sqrt{2}}{V}\Sigma(x)|_{\sigma=0} = e^{\frac{2\mathbf{i}\pi^a\tau^a}{F_\pi}}$$

which is unitary $UU^\dagger = U^\dagger U = \mathbb{1}$. This last identity means that all terms in an action for $U$ require derivatives, so (again) no mass for $\pi$. The most general Lagrangian for $U$ can be written as an expansion in derivatives, and is called the *chiral Lagrangian*:

$$\mathcal{L}_\chi = \frac{F_\pi^2}{4}\mathrm{tr}D_\mu U D^\mu U^\dagger + L_1 \mathrm{tr}\left(D_\mu U D^\mu U^\dagger\right)^2 + L_2 \mathrm{tr}D_\mu U D_\nu U^\dagger \mathrm{tr}D^\nu U^\dagger D_\mu U + L_3 \mathrm{tr}D_\mu U D^\mu U^\dagger D_\nu U D^\nu U^\dagger + \cdots \tag{11.17}$$

In terms of $\pi$, the leading term expands into

$$L_\chi = \frac{1}{2}\partial_\mu\pi^a\partial^\mu\pi^a + \frac{1}{F_\pi^2}\left(-\frac{1}{3}\pi^0\pi^0 D_\mu\pi^+ D^\mu\pi^- + \cdots\right) + \frac{1}{F_\pi^4}\left(\frac{1}{18}\left(\pi^-\pi^+\right)^2 D_\mu\pi^0 D^\mu\pi^0 + \cdots\right)$$

This fixes the relative coefficients of many irrelevant interactions, all with two derivatives, suppressed by powers of $F_\pi$. The expansion of the $L_i$ terms have four derivatives, and are therefore suppressed by further powers of $E/F_\pi$, the promised small parameter of this EFT.

**Pion masses.** The pions aren't actually massless: $m_{\pi^\pm} \sim 140\mathrm{MeV}$. In terms of quarks, one source for such a thing is the quark mass term $\mathcal{L}_{QCD} \ni \bar{q}Mq$. This explicitly breaks the isospin symmetry if the eigenvalues of $M$ aren't equal. But an *invariance* of $\mathcal{L}_{QCD}$ is

$$q_{L/R} \to g_{L/R}q_{L/R}, \;\; M \to g_L M g_R^\dagger. \tag{11.18}$$

Think of $M$ as a background field (such a thing is sometimes called a *spurion*). If $M$ were an actual dynamical field, then (11.18) would be a symmetry. In the effective action which summarizes all the drama of strong-coupling QCD in terms of pions, the field $M$ must still be there, and if we transform it as in (11.18), it should still be an invariance. Maybe we're going to do the path integral over $M$ later. (This 'spurion' trick has applications all over physics.)

So the chiral lagrangian $\mathcal{L}_\chi$ should depend on $M$ and (11.18) should be an invariance. This determines

$$\Delta\mathcal{L}_\chi = \frac{V^3}{2}\mathrm{tr}\left(MU + M^\dagger U^\dagger\right) + \cdots = V^3(m_u + m_d) - \frac{V^3}{2F_\pi^2}(m_u + m_d)\sum_a \pi_a^2 + \mathcal{O}(\pi^2).$$

The coefficient $V^3$ is chosen so that the first term matches $\langle\bar{q}Mq\rangle = V^3(m_u + m_d)$. The second term then gives

$$m_\pi^2 \simeq \frac{V^3}{F_\pi^2}(m_u + m_d)$$

which is called the Gell-Mann Oakes Renner relation.

**Electroweak interactions.** You may have noticed that I used covariant-looking $D$s in (11.17). That's because the $\mathsf{SU}(2)_L$ symmetry we've been speaking about is actually gauged by $W_\mu^a$. (The electroweak gauge boson kinetic terms are in the $\cdots$ of (11.17).) Recall that

$$\mathcal{L}_{\text{Weak}} \ni gW_\mu^a \left( \underbrace{J_\mu^a - J_\mu^{5a}}_{\text{`V' - `A'}} \right) = gW_\mu^a \left( V_{ij}\bar{Q}_i \gamma^\mu \frac{1-\gamma^5}{2}\tau^a Q_j + \bar{L}_i \gamma^\mu \tau^a \frac{1-\gamma^5}{2} L_i \right)$$

where $Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}, L_1 = \begin{pmatrix} e \\ \nu_e \end{pmatrix}$ are doublets of $\mathsf{SU}(2)_L$.

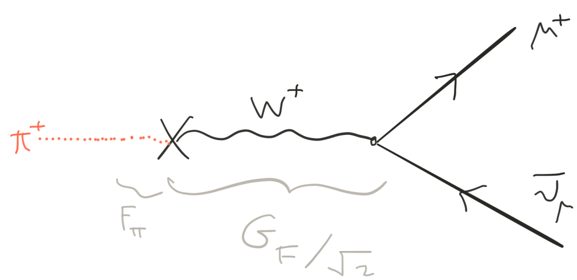
Now, in equations, the statement "a pion is a Goldstone boson for the axial $\mathsf{SU}(2)$" is:

$$\langle 0| J_\mu^{5a}(x) \left|\pi^b(p)\right\rangle = \mathbf{i} p_\mu F_\pi e^{-\mathbf{i}p\cdot x}\delta^{ab}$$

where the state $\left|\pi^b(p)\right\rangle$ is a one-pion state of momentum $p$. If the vacuum were invariant under the symmetry transformation generated by $J_\mu$, the BHS would vanish. The momentum dependence implements the fact that a global rotation ($p_\mu = 0$) does not change the energy. Contracting the BHS with $p^\mu$ and using current conservation (ignoring the explicit breaking just mentioned) would give $0 = p^2 F_\pi^2 = m_\pi^2 F_\pi^2$, a massless dispersion for the pions.

Combining the previous two paragraphs, we see that the following process can happen

$$\pi \overset{\text{Goldstone}}{\rightarrow} J_\mu^5 \overset{\text{electroweak interaction}}{\rightarrow} \text{leptons}$$




(11.19)

and in fact is responsible for the dominant decay channel of charged pions. (Time goes from left to right in these diagrams, sorry.)

$$\mathcal{M}(\pi^+ \to \mu^+\nu_\mu) = \frac{G_F}{\sqrt{2}} F_\pi p^\mu \bar{v}_{\nu_\mu} \gamma^\mu (1-\gamma^5) u_\mu$$

where the Fermi constant $G_F \sim 10^{-5} GeV^{-2}$ (known from *e.g.* $\mu^- \to e^- \bar{\nu}_e \nu_\mu$ ) is a good way to parametrize the Weak interaction amplitude. Squaring this and integrating over two-body phase space gives the decay rate

$$\Gamma(\pi^+ \to \mu^+ \nu_\mu) = \frac{G_F^2 F_\pi^2}{4\pi} m_\pi m_\mu^2 \left(1 - \frac{m_\mu^2}{m_\pi^2}\right)^2 .$$

(You can see from the answer why the decay to muons is more important than the decay to electrons, since $m_\mu / m_e \sim 200$. This is called *helicity suppression* – the decay of the helicity-zero $\pi^-$ into back-to-back spin-half particles by the weak interaction (which only produces $L$ particles and $R$ antiparticles) can't happen if helicity is conserved – the mass term is required to flip the $e_L$ into an $e_R$.) This contributes most of $\tau_{\pi^+} = \Gamma^{-1} = 2.6 \cdot 10^{-8} s$.

Knowing further the mass of the muon $m_\mu = 106\text{MeV}$ then determines $F_\pi = 92\text{MeV}$ which fixes the leading terms in the chiral Lagrangian. This is why $F_\pi$ is called the *pion decay constant*. This gives a huge set of predictions for *e.g.* pion scattering $\pi^0\pi^0 \to \pi^+\pi^-$ cross sections.

Note that the neutral pion can decay by an anomaly into two photons:

$$q_\mu \langle p_1, p_2 | J_\mu^{5,a=3}(q) |0\rangle = -c \frac{e^2}{4\pi^2} \epsilon^{\nu\lambda\alpha\beta} p_1^\nu \epsilon_1^\lambda p_2^\alpha \epsilon_2^\beta$$

where $\langle p_1, p_2|$ is a state with two photons of polarizations $\epsilon_{1,2}$. I know this because it is a matrix element of the $J_e J_e J_{\mathsf{SU}(2)-\text{axial}}$ anomaly,

$$\partial_\mu J^{\mu 5a} = -\frac{e^2}{16\pi^2} \epsilon^{\nu\lambda\alpha\beta} F_{\nu\lambda} F_{\alpha\beta} \text{tr}\left(\tau^a Q^2\right)$$

where $Q = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix}$ is the quark charge matrix. Comments: (1) this symmetry acts by $u \to e^{\mathbf{i}\theta\gamma^5} u, d \to e^{-\mathbf{i}\theta\gamma^5}$, and is *not* the same as the anomalous $\mathsf{U}(1)_A$ (which does $q_i \to e^{\mathbf{i}\theta\gamma^5} q_i$ for every flavor), and it's also not the same as isospin $u \to e^{\mathbf{i}\theta} u, d \to e^{-\mathbf{i}\theta}$, which is not chiral, and not spontaneously broken. Confusing! (2) The rate of $\pi^0$ decay (known since the 1940s) gives a measurement of the number of colors of QCD! (3) This effect can be encoded in the Lagrangian for the pions by a term

$$L \ni N_c \frac{e^2}{16\pi^2} \pi^0 \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma},$$

where $N_c = 3$ is the number of colors. The effective field theory consistently realizes the anomalies of the microscopic theory. This is an example of *'t Hooft anomaly matching*, a principle which can be used, for example, to prove that QCD *must* spontaneously break the $\mathsf{SU}(3)_L \times \mathsf{SU}(3)_R$ chiral symmetry (see Schwartz §30.6). [End of Lecture 36]

Wait – what $\mathsf{SU}(3)$?

**$\mathsf{SU}(3)$ and baryons.** The strange quark mass is also pretty small $m_s \sim 95\text{MeV}$, and $\langle \bar{s}s\rangle \sim V^3$. This means the approximate invariance and symmetry breaking pattern is actually $\mathsf{SU}(3)_L \times \mathsf{SU}(3)_R \to \mathsf{SU}(3)_{\text{diag}}$, meaning that there are $16-8=8$ pseudo NGBs. Besides $\pi^{\pm,0}$, the others are the kaons $K^{\pm,0}$ and $\eta$. It's still only the $\mathsf{SU}(2)_L$ that's gauged.

We can also include baryons $B = \epsilon_{\alpha\beta\gamma}q_\alpha q_\beta q_\gamma$. Since $q = (u,d,s) \in 3$ of the flavor $\mathsf{SU}(3)$, the baryons are in the representation

$$\begin{array}{ccccc} 3\otimes 3\otimes 3 & = (6\oplus \bar{3})\otimes 3 & = 10\oplus 8\oplus 8\oplus 1 \\ \square\otimes\square\otimes\square & = \left(\square\!\square \oplus \begin{smallmatrix}\square\\ \square\end{smallmatrix}\right)\otimes\square & = \square\!\square\!\square \oplus \begin{smallmatrix}\square\square\\ \square\;\,\end{smallmatrix} \oplus \begin{smallmatrix}\square\square\\ \square\;\,\end{smallmatrix} \oplus \begin{smallmatrix}\square\\ \square\\ \square\end{smallmatrix} \end{array} \tag{11.20}$$

The proton and neutron are in one of the octets. This point of view brought some order (and some predictions) to the otherwise-bewildering zoo of hadrons.

Returning to the two-flavor $\mathsf{SU}(2)$ approximation, we can include the nucleons $N_{L/R} = \begin{pmatrix} p \\ n \end{pmatrix}_{L/R}$ and couple them to pions by the symmetric coupling

$$\mathcal{L} \ni \lambda_{NN\pi}\bar{N}_L \Sigma N_R.$$

The expectation value for $\Sigma$ gives a nucleon mass: $m_N = \lambda_{NN\pi}F_\pi$, where $\lambda_{NN\pi}$ can be measured by scattering. This is a cheap version of the *Goldberger-Treiman relation*; for a better one see Peskin pp. 670-672.

**WZW terms in the chiral Lagrangian.** Finally, I would be remiss not to mention that the chiral Lagrangian must be supplemented by WZW terms to have the correct realization of symmetries (in order to encode all the effects of anomalies, and in order to violate $\pi \to -\pi$ which is not a symmetry of QCD).

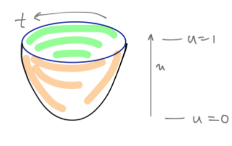
The chiral Lagrangian governs a *non-linear sigma model* (NL$\sigma$M)– a QFT whose fields are maps from spacetime into some *target space*. In this case the target space is the coset space $\mathsf{G}/\mathsf{H}$, where $\mathsf{G}$ is the full symmetry group ($\mathsf{SU}(N_f)_L \times \mathsf{SU}(N_f)_R$) and $\mathsf{H}$ is the *unbroken* subgroup $\mathsf{SU}(N_f)_{\text{diagonal}}$. We can parametrize this space by $U = e^{\mathbf{i}\pi^a T^a \frac{2}{F_\pi}}$ where the $T^a$ includes only generators of the broken part of the group, so the $\pi^a$ are coordinates on $\mathsf{G}/\mathsf{H}$.

A WZW term is a term which we can sometimes add to a NL$\sigma$M action; it is defined by the fact that it is symmetric under some group $\mathsf{G}$, but isn't the integral of a symmetric local Lagrangian density in $D$ dimensions. Making it manifestly symmetric requires the introduction of a fictitious extra dimension. This has the dramatic and surprising consequence that its coefficient is quantized.

To get the idea, consider a model in $D = 0 + 1$ where the field variable $\check{n}$ takes values on the unit sphere $S^2$, $1 = \sum_{a=1,2,3} \check{n}_a^2$. This is a special case of a coset space $\mathsf{G}/\mathsf{H} = \mathsf{SU}(2)/\mathsf{U}(1)$.

In order to write the WZW term in a manifestly symmetric way (under the $\mathsf{SO}(3)$ of rotations of the sphere, we have to extend the field into a (possibly fictitious) extra dimension whose coordinate is $u$.
We do this in such a way that the real system lives at $u = 1$:

$$\check{n}(t, u = 1) \equiv \check{n}(t), \quad \check{n}(t, u = 0) \equiv (0, 0, 1)$$




it goes to the north pole at the other end of the extra dimension for all $t$. Consider periodic boundary conditions in time $\check{n}(2\pi) = \check{n}(0)$. Then this means that the full space is really a disk with the origin at $u = 0$, and the boundary at $u = 1$. Call this disk $B$, its boundary $\partial B = \mathcal{M}$ is the real spacetime (here a circle).

We can write the WZW term in terms of the $S^2$-valued field $\check{n}^{1,2,3}$ as

$$\mathcal{W}_1[\check{n}] = \frac{2\pi}{\Omega_2} \int_{B_2} \check{n}^a \mathrm{d}\check{n}^b \wedge \mathrm{d}\check{n}^c \epsilon_{abc} = \frac{1}{4\pi} \int_{\mathcal{M}} \mathrm{d}t \, (1 - \cos\theta) \, \partial_t \phi.$$

The integrand here is the volume element of the image of a chunk of spacetime in the target $S^2$. If we integrate over the union of two balls with cancelling boundaries $B_2 \cup \bar{B}_2$, we get an integer multiple of $2\pi$ (the integer is the winding number of the map).

The coefficient $k$ of $\mathcal{W}_1$ in the action $\Delta S[\check{n}] = k\mathcal{W}_1[\check{n}]$ must be an integer since $B_1$ and $\bar{B}_1$ give equally good definitions of $\mathcal{W}_2$, which differ by $2\pi k$. So this ambiguity will not affect the path integral if $k \in \mathbb{Z}$.

---

The generalization to a group-valued variable $U$ in any dimension is of the form

$$\mathcal{W}_D = c \int_{B_{D+1}} \mathrm{tr} \underbrace{U^{-1}dU \wedge U^{-1}dU \wedge \cdots \wedge U^{-1}dU}_{D+1 \text{ of these}}.$$

Such terms are interesting when $\pi_{D+1}(\mathcal{M})$ is nontrivial, where $\mathcal{M}$ is the space where the fields live (the target space), that is, there are maps from $S^{D+1}$ to $\mathcal{M}$ which cannot be smoothly deformed to the trivial map where every point in the base space goes to the same point in the target. The variation of $\mathcal{W}_D$ with respect to $U$ is (for even $D$)[38]:

$$\delta\mathcal{W}_D = (D+1)c\int_{B_{D+1}} \mathrm{tr}\left\{ \left(U^{-1}dU\right)^D \underbrace{\delta\left(U^{-1}dU\right)}_{=U^{-1}d(\delta UU^{-1})U} \right\} \tag{11.22}$$

$$= (D+1)c\int_{B_{D+1}} \mathrm{tr}\left\{ \left(dUU^{-1}\right)^D d(\delta UU^{-1}) \right\} \tag{11.23}$$

$$= (D+1)c\int_{B_{D+1}} d\mathrm{tr}\left\{ \left(U^{-1}dU\right)^D U^{-1}\delta U \right\} \tag{11.24}$$

$$\overset{\text{Stokes}}{=} (D+1)c\int_{\mathcal{M}} \mathrm{tr}\left\{ \left(U^{-1}dU\right)^D U^{-1}\delta U \right\}$$

which only depends on the field configuration on $\mathcal{M}$, not on the extension to $B_{D+1}$. Again there can be topologically distinct ways to make the extension; demanding that they always give the same answer determines $c$ in terms of volumes of spheres (so that $c\int_{S^{D+1}} \mathrm{tr}(U^{-1}dU)^{D+1} \in \mathbb{Z}$ is the winding number), and the coefficient must be an integer. (In $D=4$, we have $c = \frac{\mathbf{i}}{240\pi^2}$.)

This WZW term is *less* topological than the theta term we discussed above, in the sense that it affects the equations of motion for $\check{n}(t)$. The variation of $\mathcal{W}$ *is* local in $D$ dimensions. The following table gives a comparison between theta terms and WZW terms for a field theory in $D$ spacetime dimensions, on a spacetime $M_D$:

[38] Why do I restrict to even $D$?

$$\mathrm{tr}\left(U^{-1}dU\right)^{D+1} = \epsilon^{\mu_1\cdots\mu_{D+1}}\mathrm{tr}\left(U^{-1}\partial_{\mu_1}U\cdots U^{-1}\partial_{\mu_{D+1}}\right)$$

but $\epsilon^{\mu_1\cdots\mu_{D+1}} = -(-1)^{D+1}\epsilon^{\mu_{D+1}\mu_1\cdots\mu_D}$ so $\mathcal{W}_D = (-1)^D\mathcal{W}_D$ vanishes in odd dimensions. The step from (11.23) to (11.24) also relies on this fact. Using $1 = U^{-1}U$ and hence $0 = \delta(U^{-1}U) = d(U^{-1}U)$, so that

$$dU^{-1} = -U^{-1}dUU^{-1}, \tag{11.21}$$

the term by which (11.23) and (11.24) differ is

$$\begin{aligned} &\mathrm{tr}\left\{\left(d\left(U^{-1}dU\right)^D\right)\delta UU^{-1}\right\} \\ \overset{\text{product rule}}{=} &\mathrm{tr}\left\{\left(dU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-1} - (U^{-1}dU\wedge dU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-2}+\cdots\right)\delta UU^{-1}\right\} \\ \overset{(11.21)}{=} &-\mathrm{tr}\left\{\left(U^{-1}dU\wedge U^{-1}dU\wedge\left(U^{-1}dU\right)^{D-1} - U^{-1}dUU^{-1}\wedge dUU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-2}+\cdots\right)\delta UU^{-1}\right\} \\ = &\mathrm{tr}\left\{\underbrace{(1-1+1-1...)}_{D-1\text{ of these}}\left(U^{-1}dU\right)^{D-1}\delta UU^{-1}\right\} \overset{D-1\text{ even}}{=} 0. \end{aligned}$$

See Weinberg, vol 2, §23.4 for more.

| theta term | WZW term |
|---|---|
| $\mathcal{H} = \int_{M_D} h$ <br> $h = \mathrm{d}q$ <br> Doesn't affect EOM | $\mathcal{W}_D = \int_{B_{D+1}} w, \;\; \partial B_{D+1} = M_D$ <br> $w = \mathrm{d}v$ <br> Affects EOM |
| Invisible in perturbation theory | Appears in perturbation theory, <br> *e.g.* in beta functions |
| $\mathcal{H} \in \mathbb{Z}$ for $M_D$ closed | Coefficient of $\mathcal{W} \in \mathbb{Z}$ <br> in order for path integral to be well-defined. |

Pion physics is the context where these terms were first discovered, and where it was realized that their coefficients are quantized. In particular the coefficient of the WZW term $W_4[U]$ here is $N_c$, the number of colors, as Witten shows by explicitly coupling to electromagnetism, and finding the term that encodes $\pi^0 \to \gamma\gamma$. One dramatic consequence here is that the chiral Lagrangian (with some higher-derivative terms) has a topological soliton solution (the skyrmion) which is a *fermion* if the number of colors of QCD is odd. The field configuration $U(x,t)$ is constant in time and approaches the vacuum at infinity, so we can regard it as a map $U : (\text{space} \cup \infty \sim S^d) \to G/H$, where $G$ is the full symmetry group and $H$ is the unbroken subgroup, so $G/H$ is the space of Goldstones (in the chiral Lagrangian, $G/H = \mathsf{SU}(3) \times \mathsf{SU}(3)/\mathsf{SU}(3)_{\text{preserved}} \simeq \mathsf{SU}(3)_{\text{broken}}$). The configuration is topological in the sense that as a map from $S^3 \to G/H$, it cannot be smoothly deformed to the trivial map – it represents a nontrivial element of $\pi_3(G/H)$. Its nontriviality is witnessed by a winding number, which can be written as the integral of a local density. In fact, the baryon number of this configuration comes from the anomalous (WZW) contribution to the baryon number current $B_\mu = \frac{\epsilon_{\mu\nu\alpha\beta}}{24\pi^2} \mathrm{tr} U^{-1}\partial_\nu U U^{-1}\partial_\alpha U U^{-1}\partial_\beta U$ whose conserved charge $\int_{\text{space}} B_0$ is exactly the winding number of the map from space (plus the point at infinity) to the space of goldstones. And finally this object a fermion because the WZW term evaluates to $\pi$ on a spacetime trajectory where the soliton makes a $2\pi$ rotation. So this object is a fermionic particle which carries baryon number. It also carries isospin. It's a nucleon! Above we added nucleon fields to the chiral Lagrangian, but we actually didn't need to – they were already there as solitonic excitations. Note that the *size* of the soliton (the region of space over which the fields vary) is determined by the higher-derivative terms in the chiral lagrangian, so we shouldn't take too seriously the substructure of the proton predicted by this picture. But it doesn't do too badly.

I should also mention that WZW terms are important in the study of interacting spin systems, for example in our understanding the dependence on the $s$ of Heisenberg spin-$s$ chains, and in phase transitions beyond the Landau-Ginzburg (symmetry-breaking) paradigm (*i.e.* deconfined quantum criticality). This is a subject for Physics 215C.

## 11.8 Superconductors

Recall from §9.1 our effective (Landau-Ginzburg) description of superconductors which reproduces the Meissner effect, the Abelian Higgs model:

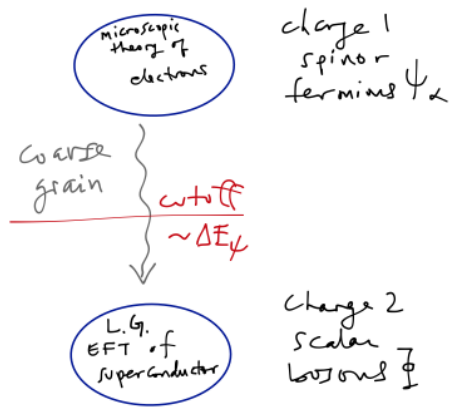
$$\mathcal{F} = \frac{1}{4} F_{ij} F_{ij} + |D_i \Phi|^2 + a|\Phi|^2 + \frac{1}{2} b |\Phi|^4 + \ldots \tag{11.25}$$

with $D_i \Phi \equiv (\partial_i - 2e\mathbf{i}A_i)\,\Phi$.

I want to make two more comments about this:

**Symmetry breaking by fluctuations (Coleman-Weinberg).** [Zee problem IV.6.9.] What happens near the transition, when $a = 0$ in (11.25)? Quantum fluctuations can lead to symmetry breaking.

**New IR dofs.** A feature of this example that I want you to notice: the microscopic description of real superconductor involves electrons – charge $1e$ spinor fermions, created by some fermionic operator $\psi_\alpha$, $\alpha = \uparrow, \downarrow$.




We are describing the low-energy physics of a system of electrons in terms of a bosonic field, which (in simple '$s$-wave' superconductors) is roughly related to the electron field by

$$\Phi \sim \psi_\alpha \psi_\beta \epsilon^{\alpha\beta} \;; \tag{11.26}$$

$\Phi$ is called a Cooper pair field. At least, the charges and the spins and the statistics work out. The details of this relationship are not the important point I wanted to emphasize. Rather I wanted to emphasize the dramatic difference in the correct choice of variables between the UV description (spinor fermions) and the IR description (scalar bosons). One reason that this is possible is that it costs a large energy to make a fermionic excitation of the superconductor. This can be understood roughly as follows: The microscopic theory of the electrons looks something like (ignoring the coupling to electromagnetism for now)

$$S[\psi] = S_2[\psi] + \int \mathrm{d}t\mathrm{d}^d x \; u\psi^\dagger\psi\psi^\dagger\psi + h.c. \tag{11.27}$$

where

$$S_2 = \int \mathrm{d}t \int đ^d k \psi_k^\dagger \left( \mathbf{i}\partial_t - \epsilon(k) \right) \psi_k .$$

Spin is important here so that $\psi_\uparrow^\dagger \psi_\uparrow \psi_\downarrow^\dagger \psi_\downarrow$ is nonzero. A mean field theory description of the condensation of Cooper pairs (11.26) is obtained by replacing the quartic term in (11.27) by expectation values:

$$\begin{aligned} S_{MFT}[\psi] &= S_2[\psi] - \int \mathrm{d}t\mathrm{d}^d x \; u \left\langle \psi\psi \right\rangle \psi^\dagger \psi^\dagger + h.c. \\ &= S_2[\psi] - \int \mathrm{d}t\mathrm{d}^d x \; u \Phi \psi^\dagger \psi^\dagger + h.c. \end{aligned} \tag{11.28}$$

So an expectation value for $\Phi$ is a mass for the fermions. It is a funny kind of symmetry-breaking mass, but if you diagonalize the quadratic operator in (11.28) (actually it is done below) you will find that it costs an energy of order $\Delta E_\psi = u \left\langle \Phi \right\rangle$ to excite a fermion. That's the cutoff on the LG EFT.

A general lesson from this example is: the useful degrees of freedom at low energies can be very different from the microscopic dofs.

---

### 11.8.1 Lightning discussion of BCS.

I am sure that some of you are nervous about the step from $S[\psi]$ to $S_{MFT}[\psi]$ above. To make ourselves feel better about it, I will say a few more words about the steps from the microscopic model of electrons (11.27) to the LG theory of Cooper pairs (these steps were taken by Bardeen, Cooper and Schreiffer (BCS)).

First recall the *Hubbard-Stratonovich transformation* aka *completing the square*. In 0+0 dimensional field theory:

$$e^{-\mathbf{i}ux^4} = \frac{1}{\sqrt{\mathbf{i}\pi u}} \int_{-\infty}^{\infty} \mathrm{d}\sigma \; e^{-\frac{1}{\mathbf{i}u}\sigma^2 - 2\mathbf{i}x^2\sigma} \; . \tag{11.29}$$

At the cost of introducing an extra field $\sigma$, we turn a quartic term in $x$ into a quadratic term in $x$. The RHS of (11.29) is gaussian in $x$ and we know how to integrate it over $x$. (The version with $\mathbf{i}$ is relevant for the real-time integral.) Notice the weird extra factor of $\mathbf{i}$ lurking in (11.29). This can be understood as arising because we are trying to use a scalar field $\sigma$, to mediate a repulsive interaction (which it is, for positive $u$) (see Zee p. 193, 2nd Ed).

Actually, we'll need a complex H-S field:

$$e^{-\mathbf{i}ux^2\bar{x}^2} = \frac{1}{\mathbf{i}\pi u} \int_{\mathbb{C}} \mathrm{d}^2\sigma \; e^{-\frac{1}{\mathbf{i}u}|\sigma|^2 - \mathbf{i}x^2\bar{\sigma} - \mathbf{i}\bar{x}^2\sigma} \; , \tag{11.30}$$

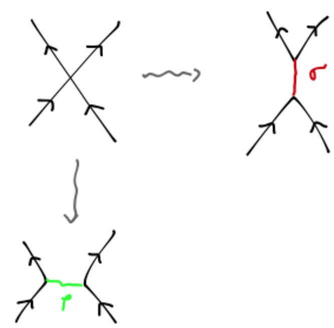
where $\int_{\mathbb{C}} \mathrm{d}^2\sigma ... \equiv \int_{-\infty}^{\infty} d\mathrm{Re}\,\sigma \int_{-\infty}^{\infty} d\mathrm{Im}\,\sigma ...$ (The field-independent prefactor is, as usual, not important for path integrals.)

We can use a field theory generalization of (11.30) to 'decouple' the 4-fermion interaction in (11.27):

$$Z = \int [D\psi D\psi^\dagger] e^{\mathbf{i}S[\psi]} = \int [D\psi D\psi^\dagger D\sigma D\sigma^\dagger] e^{\mathbf{i}S_2[\psi] + \mathbf{i}\int d^Dx (\bar{\sigma}\psi\psi + h.c.) - \int \mathrm{d}^Dx \frac{|\sigma|^2(x)}{\mathbf{i}u}} . \quad (11.31)$$

The point of this is that now the fermion integral is gaussian. At the saddle point of the $\sigma$ integral (which is exact because it is gaussian), $\sigma$ is the Cooper pair field, $\sigma_{\text{saddle}} = u\psi\psi$.

Notice that we made a choice here about in which 'channel' to make the decoupling – we could have instead introduces a different auxiliary field $\rho$ and written $S[\rho,\psi] = \int \rho\psi^\dagger\psi + \int \frac{\rho^2}{2u}$, which would break up the 4-fermion interaction in the $t$-channel (as an interaction of the fermion density $\psi^\dagger\psi$) instead of the $s$ (BCS) channel (as an interaction of Cooper pairs $\psi^2$). At this stage both are correct, but they lead to different mean-field approximations below. That the BCS mean field theory wins is a consequence of the RG.




How can you resist doing the fermion integral in (11.31)? Let's study the case where the single-fermion dispersion is $\epsilon(k) = \frac{\vec{k}^2}{2m} - \mu$.

$$I_\psi[\sigma] \equiv \int [D\psi D\psi^\dagger] e^{\mathbf{i}\int \mathrm{d}t\mathrm{d}^dx \left(\psi^\dagger\left(\mathbf{i}\partial_t - \frac{\nabla^2}{2m} - \mu\right)\psi + \psi\bar{\sigma}\psi + \bar{\psi}\bar{\psi}\sigma\right)}$$

The action here can be written as the integral of

$$L = \left(\bar{\psi}\ \psi\right) \begin{pmatrix} \mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla) & \sigma \\ \bar{\sigma} & -\left(\mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla)\right) \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} \equiv \left(\bar{\psi}\ \psi\right) M \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix}$$

so the integral is

$$I_\psi[\sigma] = \det M = e^{\operatorname{tr}\log M(\sigma)}.$$

If $\sigma$ is constant (which will lower the energy), the matrix $M$ is diagonal in momentum space, and the integral remaining to be done is

$$\int [D\sigma D\sigma^\dagger] e^{-\int \mathrm{d}^Dx \frac{|\sigma(x)|^2}{2\mathbf{i}u} + \int \mathrm{d}^Dk \log\left(\omega^2 - \epsilon_k^2 - |\sigma|^2\right)}.$$

It is often possible to do this integral by saddle point. This can justified, for example, by the largeness of the volume of the Fermi surface, $\{k|\epsilon(k) = \mu\}$, or by large $N$ number of species of fermions. The result is an equation which determines $\sigma$, which as we saw earlier determines the fermion gap.

$$0 = \frac{\delta \text{exponent}}{\delta \bar{\sigma}} = \mathbf{i}\frac{\sigma}{2u} + \int đ\omega đ^d k \frac{2\sigma}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} \ .$$

We can do the frequency integral by residues:

$$\int đ\omega \frac{1}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} = \frac{1}{2\pi} 2\pi \mathbf{i} \frac{1}{2\sqrt{\epsilon_k^2 + |\sigma|^2}}.$$

The resulting equation is naturally called the *gap equation*:

$$1 = -2u \int đ^d p' \frac{1}{\sqrt{\epsilon(p')^2 + |\sigma|^2}} \tag{11.32}$$

which you can imagine solving self-consistently for $\sigma$. Plugging back into the action (11.31) says that $\sigma$ determines the energy cost to have electrons around; more precisely, $\sigma$ is the energy required to break a Cooper pair.

Comments:

- Notice that a solution of (11.32) requires $u < 0$, an *attractive* interaction. Superconductivity happens because the $u$ that appears here is not the bare interaction between electrons, which is certainly repulsive (and long-ranged). This is where the phonons come in in the BCS discussion.

- If we hadn't restricted to a delta-function 4-fermion interaction $u(p, p') = u_0$ at the outset, we would have found a more general equation like

$$\sigma(\vec{p}) = -\frac{1}{2} \int đ^d p' \frac{u(p, p')\sigma(\vec{p}')}{\sqrt{\epsilon(p')^2 + |\sigma(p')|^2}} \ .$$

- A conservative perspective on the preceding calculation is that we have made a variational ansatz for the groundstate wavefunction, and the equation we solve for $\sigma$ is minimizing the variational energy – finding the best wavefunction within the ansatz.

- I haven't included here effects of the fluctuations of the fermions. In fact, they make the four-fermion interaction which leads to Cooper pairing marginally *relevant*. This breaks the degeneracy in deciding how to split up the $\psi\psi\psi^\dagger\psi^\dagger$ into *e.g.* $\psi\psi\sigma$ or $\psi^\dagger\psi\rho$. BCS wins. This is explained beautifully in Polchinski, lecture 2, and R. Shankar. If there were time, I would summarize the EFT framework for understanding this in §11.9.

- I've tried to give the most efficient introduction I could here. I left out any possibility of $k$-dependence or spin dependence of the interactions or the pair field, and I've conflated the pair field with the gap. In particular, I've been sloppy about the dependence on $k$ of $\sigma$ above.

- You can study a very closely related manipulation on the problem set, in an example (the Gross-Neveu model) where the saddle point is justified by large $N$.

[End of Lecture 37]

---

## 11.9 Effective field theory of Fermi surfaces

I skipped this subsection in lecture. Perhaps we can do this as part of the RG class. Skip to §12. [Polchinski, lecture 2, and R. Shankar] Electrically conducting solids are a remarkable phenomenon. An arbitrarily small electric field $\vec{E}$ leads to a nonzero current $\vec{j} = \sigma\vec{E}$. This means that there must be gapless modes with energies much less than the natural cutoff scale in the problem.

---

**Scales involved:** The Planck scale of solid state physics (made by the logic by which Planck made his quantum gravity energy scale, namely by making a quantity with dimensions of energy out of the available constants) is

$$E_0 = \frac{1}{2}\frac{e^4 m}{\hbar^2} = \frac{1}{2}\frac{e^2}{a_0} \sim 13\text{eV}$$

(where $m \equiv m_e$ is the electron mass and the factor of 2 is an abuse of outside information) which is the energy scale of *chemistry*. Chemistry is to solids as the melting of spacetime is to particle physics. There are other scales involved however. In particular a solid involves a lattice of nuclei, each with $M \gg m$ (approximately the proton mass). So $m/M$ is a useful small parameter which controls the coupling between the electrons and the lattice vibrations. Also, the actual speed of light $c \gg v_F$ can generally also be treated as $\infty$ to first approximation. $v_F/c$ suppresses spin orbit couplings (though large atomic numbers enhance them: $\lambda_{\text{SO}} \propto Z v_F/c$).

---

Let us attempt to construct a Wilsonian-natural effective field theory of this phenomenon. The answer is called Landau Fermi Liquid Theory. What are the right low-energy degrees of freedom? Let's make a guess that they are like electrons – fermions with spin and electric charge. They will not have exactly the properties of free electrons, since they must incorporate the effects of interactions with all their friends. The 'dressed' electrons are called quasielectrons, or more generally quasiparticles.

Given the strong interactions between so many particles, why should the dofs have anything at all to do with electrons? Landau's motivation for this description (which is not always correct) is that we can imagine starting from the free theory and adiabatically turning up the interactions. If we don't encounter any phase transition along the way, we can follow each state of the free theory, and use the same labels in the interacting theory.

We will show that there is a nearly-RG-stable fixed point describing gapless quasi-electrons. Notice that we are not trying to match this description directly to some microscopic lattice model of a solid; rather we will do bottom-up effective field theory.

Having guessed the necessary dofs, let's try to write an action for them consistent with the symmetries. A good starting point is the free theory:

$$S_{\text{free}}[\psi] = \int dt\, đ^d p \left( \mathbf{i}\psi^\dagger_\sigma(p)\partial_t\psi_\sigma(p) - (\epsilon(p) - \epsilon_F)\, \psi^\dagger_\sigma(p)\psi_\sigma(p) \right)$$

where $\sigma$ is a spin index, $\epsilon_F$ is the Fermi energy (zero-temperature chemical potential), and $\epsilon(p)$ is the single-particle dispersion relation. For non-interacting non-relativistic electrons in free space, we have $\epsilon(p) = \frac{p^2}{2m}$. It will be useful to leave this as a general function of $p$. [39] [40]

The groundstate is the filled Fermi sea:

$$|\text{gs}\rangle = \prod_{p|\epsilon(p)<\epsilon_F} \psi^\dagger_p |0\rangle, \quad \psi_p |0\rangle = 0, \quad \forall p.$$

(If you don't like continuous products, put the system in a box so that $p$ is a discrete label.) The Fermi surface is the set of points in momentum space at the boundary of the filled states:

$$\text{FS} \equiv \{p|\epsilon(p) = \epsilon_F\}.$$

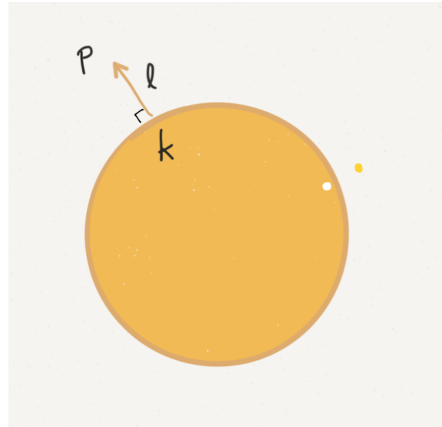
The low-lying excitations are made by adding an electron just above the FS or removing an electron (creating a hole) just below.

We would like to define a scaling transformation which focuses on the low-energy excitations. We scale energies by a factor $E \to bE, b < 1$. In relativistic QFT, $\vec{p}$ scales like $E$, toward zero, $\vec{p} \to b\vec{p}$, since all the low-energy stuff is near the single special point $\vec{p} = 0$. Here the situation is much more interesting because there is a whole surface of low-energy stuff on the FS. This will lead to what's called *hyperscaling violation* – we can't just count powers of momentum.

[39] Notice that we are assuming translation invariance. I am not saying anything at the moment about whether translation invariance is discrete (the ions make a periodic potential) or continuous.

[40] We have chosen the normalization of $\psi$ to fix the coefficient of the $\partial_t$ term (this rescaling may depend on $p$).

One way to implement this is to introduce a hierarchical labeling of points in momentum space, by breaking the momentum space into *patches* around the FS. (An analogous strategy of labeling is also used in heavy quark EFT and in SCET.)




We'll use a slightly different strategy, following Polchinski. To specify a point $\vec{p}$, we pick the nearest point $\vec{k}$ on the FS, $\epsilon(\vec{k}) = \epsilon_F$ (draw a line perpendicular to the FS from $\vec{p}$), and let

$$\vec{p} = \vec{k} + \vec{\ell}.$$

So $d-1$ of the components are determined by $\vec{k}$ and one is determined by $\ell$. (Clearly there are some exceptional cases if the FS gets too wiggly. Ignore these for now.)

$$\epsilon(p) - \epsilon_F = \ell v_F(\vec{k}) + \mathcal{O}(\ell^2), \qquad v_F \equiv \partial_p \epsilon|_{p=k}.$$

So a scaling rule which accomplishes our goal of focusing on the FS is

$$E \to bE, \quad \vec{k} \to \vec{k}, \quad \vec{l} \to b\vec{\ell}.$$

This implies

$$dt \to b^{-1}dt, \quad d^{d-1}\vec{k} \to d^{d-1}\vec{k}, \quad d\vec{\ell} \to bd\vec{\ell}, \quad \partial_t \to b\partial_t$$

$$S_{\text{free}} = \int \underbrace{dt\ d^{d-1}\vec{k}\ d\vec{\ell}}_{\sim b^0} \left( \mathbf{i}\psi^\dagger(p) \underbrace{\partial_t}_{\sim b^1} \psi(p) - \underbrace{\ell v_F(k)}_{\sim b^1} \psi^\dagger(p)\psi(p) \right)$$

In order to make this go like $b^0$ we require $\psi \to b^{-\frac{1}{2}}\psi$ near the free fixed point.

Next we will play the EFT game. To do so we must enumerate the symmetries we demand of our EFT:

1. Particle number, $\psi \to e^{\mathbf{i}\theta}\psi$

2. Spatial symmetries: either (a) continuous translation invariance and rotation invariance (as for *e.g.* liquid $^3$He) or (b) lattice symmetries. This means that momentum space is periodically identified, roughly $p \simeq p + 2\pi/a$ where $a$ is the lattice spacing (the set of independent momenta is called the Brillouin zone (BZ)) and $p$ is only conserved modulo an inverse lattice vector $2\pi/a$; the momentum There can also be some remnant of rotation invariance preserved by the lattice. Case (b) reduces to case (a) if the Fermi surface does not go near the edges of the BZ.

3. Spin rotation symmetry, $\mathsf{SU}(n)$ if $\sigma = 1..n$. In the limit with $c \to \infty$, this is an internal symmetry, independent of rotations.

4. Let's assume that $\epsilon(p) = \epsilon(-p)$, which is a consequence of *e.g.* parity invariance.

Now we enumerate all terms analytic in $\psi$ (since we are assuming that there are no other low-energy dofs integrating out which is the only way to get non-analytic terms in $\psi$) and consistent with the symmetries; we can order them by the number of fermion operators involved. Particle number symmetry means every $\psi$ comes with a $\psi^\dagger$. The possible **quadratic terms** are:

$$\int \underbrace{dt\ d^{d-1}\vec{k}\ d\vec{\ell}}_{\sim b^0}\ \mu(k)\, \underbrace{\psi_\sigma^\dagger(p)\psi_\sigma(p)}_{\sim b^{-1}} \sim b^{-1}$$

is relevant. This is like a mass term. But don't panic: it just shifts the FS around. The *existence* of a Fermi surface is Wilson-natural (*i.e.* a stable assumption given generic coefficients of all possible terms in the action); any precise location or shape (modulo something enforced by symmetries, like roundness) is not.

Adding one extra $\partial_t$ or factor of $\ell$ costs a $b^1$ and makes the operator marginal; those terms are already present in $S_{\text{free}}$. Adding more than one makes it irrelevant.

**Quartic terms:**

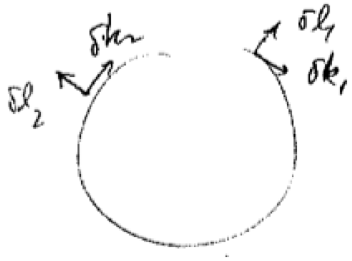
$$S_4 = \int \underbrace{dt\ \prod_{i=1}^{4} d^{d-1}\vec{k}_i d\vec{\ell}_i}_{\sim b^{-1+4-4/2}}\, u(4\cdots 1)\psi_\sigma^\dagger(p_1)\psi_\sigma(p_3)\psi_{\sigma'}^\dagger(p_2)\psi_{\sigma'}(p_4)\delta^d(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4)$$

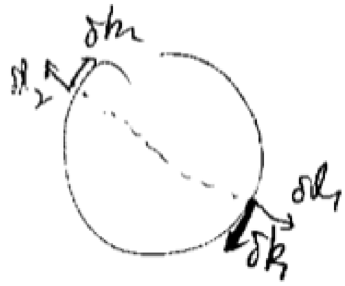
Note the similarity with the discussion of the XY model in §**??**. The minus signs on $p_{3,4}$ is because $\psi(p)$ *removes* a particle with momentum $p$. We assume $u$ depends only on $k, \sigma$, so does not scale – this will give the most relevant piece. How does the delta function scale?

$$\delta^d\left(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4\right) = \delta^d\left(k_1+k_2-k_3-k_4+\ell_1+\ell_2-\ell_3-\ell_4\right) \overset{?}{\simeq} \delta^d\left(k_1+k_2-k_3-k_4\right)$$

In the last (questioned) step, we used the fact that $\ell \ll k$ to ignore the contributions of the $\ell$s. If this is correct then the delta function does not scale (since $k$s do not), and $S_4 \sim b^1$ is irrelevant (and quartic interactions with derivatives are moreso). If this were correct, the free-fixed point would be exactly stable.

There are two important subtleties: (1) there exist phonons. (2) the questioned equality above is questionable because of kinematics of the Fermi surface. We will address these two issues in reverse order.



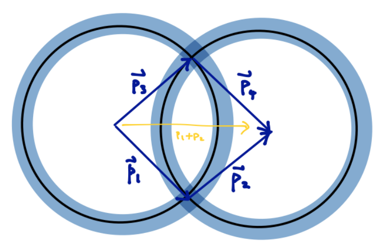

The **kinematic subtlety** in the treatment of the scaling of $\delta(p_1 + p_2 - p_3 - p_4)$ arises because of the geometry of the Fermi surface. Consider scattering between two points on the FS, where (in the labeling convention above)

$$p_3 = p_1 + \delta k_1 + \delta \ell_1, \quad p_4 = p_2 + \delta k_2 + \delta \ell_2,$$

in which case the momentum delta function is

$$\delta^d(p_1 + p_2 - p_3 - p_4) = \delta^d(\delta k_1 + \delta \ell_1 + \delta k_2 + \delta \ell_2).$$

For generic choices of the two points $p_{1,2}$ (top figure at left), $\delta k_1$ and $\delta k_2$ are linearly independent and the $\delta \ell$s can indeed be ignored as we did above. However, for two points with $p_1 = -p_2$ (they are called *nested*, as depicted in the bottom figure at left), then one component of $\delta k_1 + \delta k_2$ is automatically zero, revealing the tiny $\delta \ell$s to the force of (one component of) the delta function. In this case, $\delta(\ell)$ scales like $b^{-1}$, and for this particular kinematic configuration the four-fermion interaction is (classically) marginal. Classically marginal means quantum mechanics has a chance to make a big difference.




A useful visualization is at right ($d = 2$ with a round FS is shown; this is what's depicted on the cover of the famous book by Abrikosov-Gorkov-Dzyaloshinski): the blue circles have radius $k_F$; the yellow vector is the sum of the two initial momenta $p_1 + p_2$, both of which are on the FS; the condition that $p_3 + p_4$, each also on the FS, add up to the same vector means that $p_3$ must lie on the intersection of the two circles (spheres in $d > 2$). But when $p_1 + p_2 = 0$, the two circles are on top of each other so they intersect everywhere! Comments:

1. We assumed that both $p_1$ and $-p_2$ were actually *on* the FS. This is automatic if $\epsilon(p) = \epsilon(-p)$, *i.e.* if $\epsilon$ is only a function of $p^2$.

2. This discussion works for any $d > 1$.

3. **Forward scattering.** There is a similar phenomenon for the case where $p_1 = p_3$ (and hence $p_2 = p_4$). This is called *forward scattering* because the final momenta are the same as the initial momenta. (We could just as well take $p_1 = p_4$ (and hence $p_2 = p_3$).) In this case too the delta function will constrain the $\ell$s and will therefore scale.

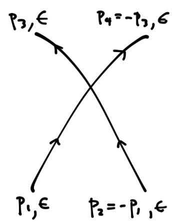
The tree-level-marginal 4-Fermi interactions at special kinematics leads to a *family* of fixed points labelled by 'Landau parameters'. In fact there is whole *function*'s worth of fixed points. In 2d, the fixed point manifold is parametrized by the forward-scattering function

$$F(\theta_1, \theta_2) \equiv u(\theta_4 = \theta_2, \theta_3 = \theta_1, \theta_2, \theta_1)$$

(Fermi statistics implies that $u(\theta_4 = \theta_1, \theta_3 = \theta_2, \theta_2, \theta_1) = -F(\theta_1, \theta_2)$ .) and the BCS-channel interaction:

$$V(\theta_1, \theta_3) = u(\theta_4 = -\theta_3, \theta_3, \theta_2 = -\theta_1, \theta_1).$$

Now let's think about what decision the fluctuations make about the fate of the nested interactions. The first claim, which I will not justify here, is that $F$ is not renormalized at one loop. The interesting bit is the renormalization of the BCS interaction:




The electron propagator, obtained by inverting the kinetic operator $S_{\text{free}}$, is

$$G(\epsilon, p = k + l) = \frac{1}{\epsilon(1 + \mathbf{i}\eta) - v_F(k)\ell + \mathcal{O}(\ell)^2}$$

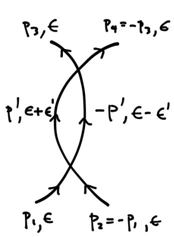
where I used $\eta \equiv 0^+$ for the infinitesimal specifying the contour prescription. (To understand the contour prescription for the *hole* propagator, it is useful to begin with

$$G(t, p) = \langle \epsilon_F | \, c_p^\dagger(t) c_p(0) \, | \epsilon_F \rangle \,, \quad c_p^\dagger(t) \equiv e^{-\mathbf{iH}t} c_p^\dagger e^{\mathbf{iH}t}$$

and use the free-fermion fact $[\mathbf{H}, c_p^\dagger] = \epsilon_p c_p^\dagger$.)

Let's assume rotation invariance. Then $V(\theta_3, \theta_1) = V(\theta_3 - \theta_1)$, $V_l = \int đ\theta e^{\mathbf{i}l\theta} V(\theta)$. Different angular momentum sectors decouple from each other at one loop.

We will focus on the $s$-wave bit of the interaction, so $V$ is independent of momentum. We will integrate out just a shell in energy (depicted by the blue shaded shell in the Fermi surface figures). The interesting contribution comes from the following diagram:

$\delta^{(1)} V =$ 


$$= \mathbf{i} V^2 \int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon' d^{d-1}k' d\ell'}{(2\pi)^{d+1}} \frac{1}{(\epsilon + \epsilon' - v_F(k')\ell') \, (\epsilon - \epsilon' - v_F(k')\ell')}$$

$$\text{do } \int d\ell' \text{ by residues} \qquad = \mathbf{i} V^2 \int \frac{d\epsilon' d^{d-1}k'}{(2\pi)^{d+1}} \frac{1}{v_F(k')} \left( \underbrace{\epsilon - \epsilon' - (\epsilon + \epsilon')}_{=-2\epsilon'} \right)^{-1}$$

$$= -V^2 \underbrace{\int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon'}{\epsilon'}}_{=\log(1/b)} \underbrace{\int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}}_{\text{dos at FS}} \tag{11.33}$$

Between the first and second lines, we did the $\ell'$ integral by residues. The crucial point is that we are interested in external energies $\epsilon \sim 0$, but we are integrating out a shell near the cutoff, so $|\epsilon'| > |\epsilon|$ and the sign of $\epsilon + \epsilon'$ is opposite that of $\epsilon - \epsilon'$; therefore there is a pole on either side of the real $\ell$ axis and we get the same answer by closing the contour either way. On one side the pole is at $\ell' = \frac{1}{v_F(k')}(\epsilon + \epsilon')$. (In the t-channel diagram (what Shankar calls ZS), the poles are on the *same* side and it therefore does not renormalize the four-fermion interaction.)

The result to one-loop is then

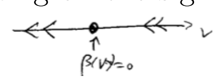
$$V(b) = V - V^2 N \log(1/b) + \mathcal{O}(V^3)$$

with $N \equiv \int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}$ is the density of states at the Fermi surface. From this we derive the beta function

$$b\frac{d}{db}V(b) = \beta_V = NV^2(b) + \mathcal{O}(V^3)$$

and the solution of the flow equation at $E = bE_1$ is

$$V(E) = \frac{V_1}{1 + NV_1 \log(E_1/E)} \begin{cases} \to 0 & \text{in IR for } V_1 > 0 \text{ (repulsive)} \\ \to -\infty & \text{in IR for } V_1 < 0 \text{ (attractive)} \end{cases} \tag{11.34}$$

There is therefore a very significant dichotomy depending on the sign of the coupling at the microscopic scale $E_1$, as in this phase diagram: 


The conclusion is that if the interaction starts attractive at some scale it flows to large attractive values. The thing that is decided by our perturbative analysis is that (if $V(E_1) > 0$) the decoupling we did with $\sigma$ ('the BCS channel') wins over the decoupling with $\rho$ ('the particle-hole channel'). What happens at $V \to -\infty$? Here we need non-perturbative physics.

The non-perturbative physics is in general hard, but we've already done what we can in §11.8.1.

The remaining question is: Who is $V_1$ and why would it be attractive (given that Coulomb interactions between electrons, while screened and therefore short-ranged, are repulsive)? The answer is:

**Phonons.** The lattice of positions taken by the ions making up a crystalline solid spontaneously break many spacetime symmetries of their governing Hamiltonian. This implies a collection of gapless Goldstone modes in any low-energy effective theory of such a solid[41]. The Goldstone theorem is satisfied by including a field

$$\vec{D} \propto \text{(local) displacement } \delta\vec{r} \text{ of ions from their equilibrium positions}$$

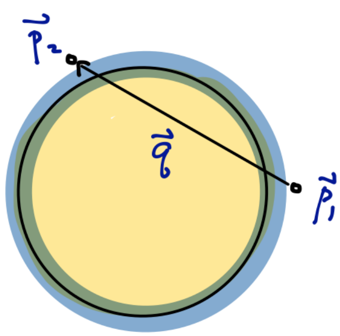
Most microscopically we have a bunch of coupled springs:

$$L_{\text{ions}} \sim \frac{1}{2}M\left(\dot{\delta\vec{r}}\right)^2 - k_{ij}\delta r^i \delta r^j + \ldots$$

with spring constants $k$ independent of the nuclear mass $M$. It is useful to introduce a canonically normalized field in terms of which the action is

$$S[\vec{D} = (M)^{1/2}\,\delta\vec{r}] = \frac{1}{2}\int dt d^d q \left(\partial_t D_i(q)\partial_t D_i(-q) - \omega_{ij}^2(q) D_i(q) D_j(-q)\right).$$

Here $\omega^2 \propto M^{-1}$. Their status as Goldstones means that the eigenvalues of $\omega_{ij}^2(q) \sim |q|^2$ at small $q$: moving everyone by the same amount does not change the energy. This also constrains the coupling of these modes to the electrons: they can only couple through derivative interactions.




For purposes of their interactions with the electrons, a nonzero $q$ which keeps the $e^-$ on the FS must scale like $q \sim b^0$. Therefore

$$dt d^d q \left(\partial_t D\right)^2 \sim b^{+1+2[D]} \qquad \implies D \sim b^{-\frac{1}{2}}$$

and the restoring force $dtdqD^2\omega^2(q) \sim b^{-2}$ is relevant, and dominates over the $\partial_t^2$ term for

$$E < E_D = \sqrt{\frac{m}{M}} E_0 \qquad \text{the Debye energy.}$$

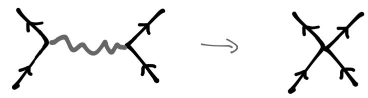
This means that phonons mediate *static* interactions below $E_D$ – we can ignore retardation effects, and their effects on the electrons can be fully incorporated by the four-fermion interaction we used above (with some $\vec{k}$ dependence). How do they couple to the electrons?

$$S_{\text{int}}[D,\psi] = \int dtq^3 q d^2 k_1 d\ell_1 d^2 k_2 d\ell_2\ M^{-\frac{1}{2}} g_i(q,k_1,k_2) D_i(q) \psi_\sigma^\dagger(p_1)\psi_\sigma(p_2)\delta^3(p_1 - p_2 - q)$$

[41]Note that there is a subtlety in counting Goldstone modes from spontaneously broken spacetime symmetries: there are more symmetry generators than Goldstones. Basically it's because the associated currents differ only by functions of spacetime; but a localized Goldstone particle is anyway made by a current times a function of spacetime, so you can't sharply distinguish the resulting particles. Some useful references on this subject are Low-Manohar and most recently Watanabe-Murayama.

$$\sim b^{-1+1+1-3/2} = b^{-1/2} \tag{11.35}$$

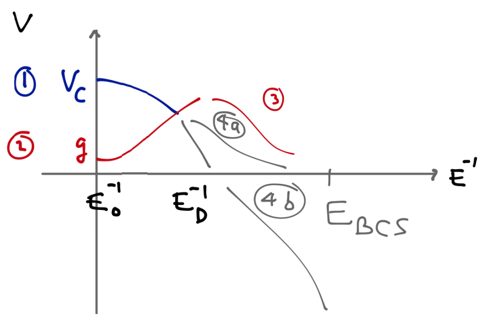
– here we took the delta function to scale like $b^0$ as above. This is relevant when we use the $\dot{D}^2$ scaling for the phonons; when the restoring force dominates we should scale $D$ differently and this is irrelevant for generic kinematics. This is consistent with our previous analysis of the four-fermion interaction.

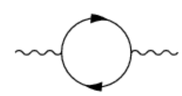
The summary of this discussion is: phonons do not destroy the Fermi surface, but they *do* produce an attractive contribution to the 4-fermion interaction, which is relevant in some range of scales (above the Debye energy). Below the Debye energy, it amounts to an addition to $V$ that goes like $-g^2$: 

Notice that the scale at which the coupling $V$ becomes strong ($V(E_{\text{BCS}}) \equiv 1$ in (11.34)) is

$$E_{\text{BCS}} \sim E_D e^{-\frac{1}{NV_D}}.$$

Two comments about this: First, it is non-perturbative in the interaction $V_D$. Second, it provides some verification of the role of phonons, since $E_D \sim M^{-1/2}$ can be varied by studying the same material with different isotopes and studying how the critical superconducting temperature ($\sim E_{\text{BCS}}$) scales with the nuclear mass.




Here's the narrative, proceeding as a function of decreasing energy scale, beginning at $E_0$, the Planck scale of solids: (1) Electrons repel each other by the Coulomb interaction. However, in a metal, this interaction is *screened* by processes like this: 

(the intermediate state is an electron-hole pair) and is short-ranged. It is still repulsive, however. As we coarse-grain more and more, we see more and more electron-hole pairs and the force weakens. (2) While this is happening, the electron-phonon interaction is relevant and growing. This adds an attractive bit to $V$. This lasts until $E_D$. (3) At $E_D$ the restoring force term in the phonon lagrangian dominates (for the purposes of their interactions with the electrons) and we can integrate them out. (4) What happens next depends on the sign of $V(E_D)$. If it's positive, $V$ flows harmlessly to zero. If it's negative, it becomes moreso until we exit the perturbative analysis at $E_{\text{BCS}}$, and vindicate our choice of Hubbard-Stratonovich channel above.

Further brief comments, for which I refer you to Shankar:

1. Putting back the possible angular dependence of the BCS interaction, the result

at one loop is

$$\frac{dV(\theta_1-\theta_3)}{d\ell} = -\frac{1}{8\pi^2}\int_0^{2\pi} đ\theta V(\theta_1-\theta)V(\theta-\theta_3)$$

or in terms of angular momentum components,

$$\frac{dV_l}{d\ell} = -\frac{V_l^2}{4\pi}.$$

2. This example is interesting and novel in that it is a (family of) fixed point(s) characterized by a dimensionful quantity, namely $k_F$. This leads to a phenomenon called *hyperscaling violation* where thermodynamic quantities need not have their naive scaling with temperature.

3. The one loop analysis gives the right answer to all loops in the limit that $N \equiv k_F/\Lambda \gg 1$, where $\Lambda$ is the UV cutoff on the momentum.

4. The forward scattering interaction (for any choice of function $F(\theta_{13})$) is not renormalized at one loop. This means it is exactly marginal at leading order in $N$.

5. Like in $\phi^4$ theory, the sunrise diagram at two loops is the first appearance of wavefunction renormalization. In the context of the Fermi liquid theory, this leads to the renormalization of the effective mass which is called $m^\star$.

---

Another consequence of the FS kinematics which I should emphasize more: it allows the quasiparticle to be stable. The leading contribution to the decay rate of a one-quasiparticle state with momentum $k$ can be obtained applying the optical theorem to the following process.

The intermediate state is two electrons with momenta $k'+q$ and $k-q$, and one *hole* with momentum $k'$. The hole propagator has the opposite $\mathbf{i}\eta$ prescription. After doing the frequency integrals by residues, we get

$$\Sigma(k,\epsilon) = \int đq\, đk' \frac{|u_q|^2}{D-\mathbf{i}\eta}$$

$$D \equiv \epsilon_k(1+\mathbf{i}\eta) + \epsilon_{k'}(1-\mathbf{i}\eta) - \epsilon_{k'+q}(1+\mathbf{i}\eta) - \epsilon_{k-q}(1+\mathbf{i}\eta)$$

(Notice that this is the eyeball diagram which gives the lowest-order contribution to the wavefunction renormalization of a field with quartic interactions.) By the optical theorem, its imaginary part is the (leading contribution to the) inverse-lifetime of the quasiparticle state with fixed $k$:

$$\tau^{-1}(k) = \mathrm{Im}\,\Sigma(k,\epsilon) = \pi\int đq\, đk' \delta(D)|u_q|^2 f(-\epsilon_{k'})f(\epsilon_{k'+q})f(\epsilon_{k-q})$$

where

$$f(\epsilon) = \lim_{T \to 0} \frac{1}{e^{\frac{\epsilon - \epsilon_F}{T}} + 1} = \theta(\epsilon < \epsilon_F)$$

is the Fermi function. This is just the demand that a particle can only scatter into an empty state and a hole can only scatter into a filled state. These constraints imply that all the energies are near the Fermi energy: both $\epsilon_{k'+q}$ and $\epsilon_{k'}$ lie in a shell of radius $\epsilon$ about the FS; the answer is proportional to the density of possible final states, which is thus

$$\tau^{-1} \propto \left(\frac{\epsilon}{\epsilon_F}\right)^2 .$$

So the width of the quasiparticle resonance is

$$\tau^{-1} \propto \epsilon^2 \ll \epsilon$$

much smaller than its frequency – it is a sharp resonance, a well-defined particle.

---

## 12 Non-abelian gauge fields in perturbation theory

**Gauge fixing.** [Peskin §16.2, Scwhartz §25] Consider the partition function for Yang-Mills theory (it will be easy to add the quarks later):

$$Z \equiv \frac{1}{\text{vol}(\mathcal{G})}\int [DA] e^{\mathbf{i}S[A]}.$$

We assume that $S[A]$ is some gauge-invariant functional of $A$, such as the Yang-Mills action. The integral over $[DA]$ goes over all configurations of the gauge field $A$. Here $\text{vol}(\mathcal{G})$ is the volume of the gauge group – a copy of $\mathsf{G}$ for each point in space. We divide by it to cancel out the contributions from gauge-equivalent configurations of $A$. We would like to make this cancellation (which is $\infty/\infty$) more explicit by fixing a gauge, $G(A) = 0$. Note that the gauge-fixing function $G(A)$ must be a function of $A$ which is *not* gauge invariant, such as $\partial^\mu A_\mu$. To do so, we will insert the following form of the number 1:

$$1 = \Delta[A]\int [D\alpha]\delta[G(A^\alpha)]. \tag{12.1}$$

Here

$$A^\alpha_\mu = (A^\alpha_\mu)^a T^a = e^{\mathbf{i}\alpha\cdot T}\left(A_\mu + \frac{\mathbf{i}}{g}\partial_\mu\right)e^{-\mathbf{i}\alpha\cdot T} \tag{12.2}$$

$$= \left(A^a_\mu + \frac{1}{g}\partial_\mu\alpha^a + f^{abc}A^b_\mu\alpha^c\right)T^a + \mathcal{O}(\alpha^2) \tag{12.3}$$

$$= \left(A^a_\mu + \frac{1}{g}D_\mu\alpha^a\right)T^a + \mathcal{O}(\alpha^2) \tag{12.4}$$

is the gauge-image of $A$ under a gauge transformation with parameter $\alpha^A(x)$ – at the last step $D$ is the covariant derivative on a field in the adjoint representation. The (Fadeev-Popov) determinant $\Delta$ defined by (12.1) is a Jacobian $\Delta[A] = \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right)$. A crucial property of $\Delta$ is that it is gauge invariant:

$$\Delta[A^{\alpha_1}]^{-1} \overset{(12.1)}{=} \int [D\alpha]\delta[G(A^{\alpha+\alpha_1})] = \int [D(\alpha+\alpha_1)]\delta[G(A^{\alpha+\alpha_1})] = \Delta[A]^{-1}. \tag{12.5}$$

(Here we used the fact that the measure on the group is invariant, $[D\alpha] = [D(\alpha+\alpha_1)]$.)

Inserting this form of 1 (12.1) into $Z$ (and changing the order of integration) gives:

$$Z = \frac{1}{\text{vol}(\mathcal{G})}\int [D\alpha]\int [DA]\delta[G(A^\alpha)]\Delta[A]e^{\mathbf{i}S[A]} \tag{12.6}$$

$$= \frac{1}{\text{vol}(\mathcal{G})}\int [D\alpha]\int [DA^\alpha]\delta[G(A^\alpha)]\Delta[A^\alpha]e^{\mathbf{i}S[A^\alpha]} \tag{12.7}$$

$$\overset{\tilde{A}\equiv A^\alpha}{=} \underbrace{\left(\frac{1}{\text{vol}(\mathcal{G})}\int [D\alpha]\right)}_{=1}\int [D\tilde{A}]\delta[G(\tilde{A})]\Delta[\tilde{A}]e^{\mathbf{i}S[\tilde{A}]}\ . \tag{12.8}$$

In the first step we use the fact that $\int [DA] = \int [DA^\alpha], S[A] = S[A^\alpha]$ and (12.5). In the second step we change integration variables to $A^\alpha \equiv \tilde{A}$, and promptly drop the tilde. So we've cancelled the offending volume of the gauge group, and inserted a gauge-fixing delta function in the path integral.

The only price is the FP determinant $\Delta$ that we've acquired. What is it? It depends on the choice of gauge fixing function. Let's choose

$$G[A] = \partial^\mu A_\mu^A(x) - \omega^A(x).$$

Rather than picking a particular $\omega$, let's average over all possibilities with gaussian measure:

$$1 = N(\xi) \int [D\omega] e^{-\mathbf{i} \int d^D x \frac{\omega^2(x)}{2\xi}}.$$

The normalization factor is just a constant which we can forget. Therefore

$$Z = N(\xi) \int [DA] \underbrace{\int [D\omega] \delta[\partial \cdot A - \omega]}_{=1} e^{-\mathbf{i} \int \frac{\omega^2}{2\xi}} \Delta[A] e^{\mathbf{i} S[A]} \tag{12.9}$$

$$= N(\xi) \int [DA] \Delta[A] e^{\mathbf{i}\left(S[A] - \int \frac{(\partial \cdot A)^2}{2\xi}\right)}. \tag{12.10}$$

Finally we must figure out what is $\Delta[A]$. Comparing to (12.4) (and remembering that there is a factor of $\delta[G]$ multiplying everything), $\Delta$ is the determinant of the operator

$$\frac{\delta G[A^\alpha]}{\delta \alpha} = \frac{1}{g} \partial^\mu D_\mu.$$

Notice that in the abelian case, this is independent of $A$ (the covariant derivative $D$ acting on the adjoint representation of $\mathsf{U}(1)$ is just $\partial$) and we can forget about it; that's why we didn't bother doing this for QED.

$$\Delta = \det\left(\frac{1}{g} \partial^\mu D_\mu\right) = \int [Dc D\bar{c}] e^{\mathbf{i} \int d^D x \bar{c}(-\partial^\mu D_\mu) c}.$$

At the last step we used the integration formula for gaussian grassmann integrals (and absorbed a factor of $g$ into the definition of $c$, and ignored a constant factor). $c$ here is a new complex scalar field in the theory ($\bar{c} \equiv c^\dagger$). Since $D$ is the covariant derivative in the adjoint, it's a field in the adjoint of the gauge group. There's just one weird thing about it – it's a *fermionic* field with second-order kinetic terms, a ghost!

We saw all kinds of bad things about fermions with second-order kinetic terms in our discussion of spin-statistics. But those bad things only happen if the particles occur in external states. One purpose of a lot of the fancy stuff on this subject (such as

BRST symmetry) is guaranteeing that we'll never make ghost particles while scattering the real particles. The loops of the ghosts, though, are crucial for getting correct and unitary answers. In particular, the optical theorem relates scattering states to particles appearing in loops. The contributions to the imaginary part of loops from the ghosts are required to cancel the unitarity-violating contributions from the unphysical polarization states of the gluons. (For the details of what is being cancelled see Peskin pp. 508-511, and for the cancellation itself, see 515-516.)

Altogether,

$$Z = \int [DADcD\bar{c}] e^{\mathbf{i}\left(S[A] - \int \frac{(\partial \cdot A)^2}{2\xi} + \int \bar{c}(-\partial^\mu D_\mu)c\right)}. \tag{12.11}$$

More generally, the ghosts are *negative degrees of freedom* whose role in life is to *cancel* the unphysical contributions of the timelike and longitudinal components of the gluon field. For example, in the free ($g = 0$) path integral in Feynman gauge, we have

$$Z = \left(\det\left(-\partial^2\right)^{-D/2} \det\left(-\partial^2\right)^{+1}\right)^{\dim(\mathsf{G})}.$$

The contribution of the ghosts cancels two components' worth of the contribution from the gluons.

**Feynman rules.** More explicitly, the ghost action is

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a \left(-\partial^2 \delta^{ab} - g\vec{\partial}^\mu f^{abc} A_\mu^c\right) c^b.$$

The ghost propagator is then:

$$\left\langle c^a(x)\bar{c}^b(y)\right\rangle = \int đ^4k \; e^{-\mathbf{i}k(x-y)} \frac{-\mathbf{i}}{k^2 + \mathbf{i}\epsilon}\delta^{ab}.$$

Let us use the lovely Yang-Mills action, $S[A] = \int \mathcal{L}_{YM}$

$$\mathcal{L}_{YM} = -\frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu}, \qquad F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c.$$

The resulting gluon propagator is

$$\left\langle A_\mu^a(x) A_\nu^b(y)\right\rangle = \int đ^4k \; e^{-\mathbf{i}k(x-y)} \frac{-\mathbf{i}}{k^2 + \mathbf{i}\epsilon}\delta^{ab}\left(\eta_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2}\right)$$

The gluon propagator is just like the photon one, times a $\delta^{ab}$ which conserves the color. $\xi = 1$ is Feynman gauge, which I'll use everywhere below.

The new Feynman rules are

 $= \dfrac{-\mathbf{i}}{k^2}$,  $= -gf^{abc}k^\mu$,

 $= gf^{abc}\left(\eta_{\mu\nu}(k_1-k_2)_\lambda + \eta_{\nu\lambda}(k_2-k_3)_\mu + \eta_{\lambda\mu}(k_3-k_1)_\nu\right),$

 $= -\mathbf{i}g^2\left(f^{abe}f^{cde}\left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\rho}\eta_{\nu\lambda}\right) + f^{ade}f^{cbe}\left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\nu}\eta_{\rho\lambda}\right) + f^{ace}f^{bde}\left(\eta_{\mu\nu}\eta_{\lambda\rho} - \eta_{\mu\rho}\eta_{\nu\lambda}\right)\right).$

(Patterns: in the cubic coupling, the three terms cyclically permute the color and kinematic indices. In the quartic coupling, the second term is obtained from the first by the interchange $(b,\nu) \leftrightarrow (d,\rho)$, and the third is obtained from the first by $(b,\nu) \leftrightarrow (c,\lambda)$.)

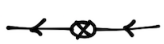
Including quarks doesn't mess with the gauge-fixing stuff. We'll take

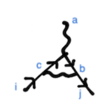
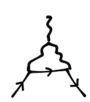
$$\mathcal{L}_{\text{quarks}} = \bar{q}\left(\mathbf{i}\not{D} - m\right)q = \bar{q}_i\left(\gamma^\mu\left(\mathbf{i}\partial_\mu\delta_{ij} + gA^a_\mu t^a_{ij}\right) - m\delta_{ij}\right)q_j$$

Here $i,j$ are color indices. For QCD, $i,j = 1..3$ and and $t^a_{ij}$ are the generators of $\mathsf{SU}(3)$ in the fundamental representation (a good basis of which are called Gell-Mann matrices). Then there's also a quark propagator, and the $qqg$ vertex,  $= \mathbf{i}g\gamma^\mu t^a$.

We'll also need to add some counterterms

 $= -\mathbf{i}\left(k^2\eta^{\mu\nu} - k^\mu k^\nu\right)\delta^{ab}\delta_3$ $\leftarrow$ 

 $= \mathbf{i}\not{k}\delta_2$ $\leftarrow$ 

 $= \mathbf{i}gt^a\gamma^\mu\delta_1$ $\leftarrow$  $+$  .

On the right, I've indicated which one-loop diagrams require us to add these respective counterterms.

**Beta function.** We're going to calculate the beta function for the QCD coupling $g$. We'll use dim reg, so the beta function is defined as $\beta(g_R) = \mu\partial_\mu g_R$, where $\mu$ is the scale that appears when we replace 4-dimensional integrals with $D$-dimensional integrals, and where $g_R$ is the renormalized coupling. Here is a good device for working out the beta function in dim reg. Very explicitly, the whole Lagrangian is

$$\mathcal{L} = -\frac{1}{4} Z_3 (\partial A)^2 + Z_2 \bar{q} \left( \mathbf{i}\not{\partial} - Z_m m_R \right) q - Z_{3c} \bar{c}^a \Box c^a \tag{12.12}$$

$$- \mu^{\epsilon/2} g_R Z_{A^3} f^{abc} (\partial_\mu A^a_\nu) A^{\mu b} A^{\nu c} - \frac{1}{4} \mu^\epsilon g_R^2 Z_{A^4} \left( f^{abc} A^b_\mu A^c_\nu \right) \left( f^{ade} A^{\mu d} A^{\nu e} \right) \tag{12.13}$$

$$+ \mu^{\epsilon/2} g_R Z_1 A^a_\mu \bar{q} \gamma^\mu t^a q + \mu^{\epsilon/2} g_R Z_{1c} f^{abc} \partial_\mu \bar{c}^a A^b_\mu c^c \tag{12.14}$$

Here I've written the counterterms in terms of $Z = 1 - \delta$. Notice that there are *four* counterterms $(Z_1, Z_{A^3}, Z_{A^4}, Z_{3c})$ all of which describe corrections to $g$ – they are related by gauge invariance, just like how in QED the vertex correction and the electron self-energy were related.[42]

The *bare fields* are the ones whose quadratic terms are just $(\partial A^0)^2$ and $\overline{q^0} \mathbf{i}\not{\partial} q^0$, *i.e.* $A^0_\mu = \sqrt{Z_3} A_\mu, q^0 = \sqrt{Z_2} q, c^0 = \sqrt{Z_{3c}} c$. The bare coupling is the coefficient of the interaction written in terms of the bare fields, *e.g.*

$$\mathcal{L}_{qqg} = \underbrace{\mu^{\frac{4-D}{2}} g_R Z_1 Z_3^{-1/2} Z_2^{-1}}_{=g_0} A^{0a}_\mu \bar{q}^0 \gamma^\mu t^a q^0.$$

Now here comes the trick: the bare coupling doesn't know anything about our choice of $\mu$.[43] Therefore

$$0 = \mu \partial_\mu g_0 = \mu \partial_\mu \left( \mu^{\frac{\epsilon}{2}} g_R Z_1 Z_3^{-1/2} Z_2^{-1} Z_1 \right) = g_0 \left( \frac{\epsilon}{2} + \frac{1}{g_R} \underbrace{\mu \partial_\mu g_R}_{=\beta(g_R)} + \mu \partial_\mu \left( \delta_1 - \frac{1}{2} \delta_3 - \delta_2 \right) \right)$$

Now the counterterms $\delta$ will depend on $\mu$ through $g_R(\mu)$, so we can use the chain rule:

$$\mu \partial_\mu \delta = \mu \frac{d g_R}{d\mu} \frac{\partial}{\partial g_R} \delta = \beta(g_R) \frac{\partial \delta}{\partial g_R}.$$

[42]Note that I also include the dimensional-analysis-restoring dim reg factor of $\mu^{\epsilon/2}$ explicitly in $\mathcal{L}$. We can see that this is the right thing to do by rescaling $\tilde{A} = gA$ so that the coupling appears only in $-\frac{1}{4g^2}\mathrm{tr}F^2$. In that case, the action and the coupling $g$ are both dimensionless in $D = 4 - \epsilon$ dimensions if we write $S = -\int d^D x \frac{\mu^{D-4}}{4g^2} \mathrm{tr} F^2$. This is the same as replacing $g \to \mu^{\epsilon/2} g$.

[43]This is a different perspective than we have when the scale introduced in the renormalization scheme is a UV cutoff. There, if we hold fixed the physical coupling we *must* vary the bare coupling with the UV cutoff, and in fact its variation defines the beta function, as in §7.1 The two perspectives are related to each other like active and passive transformations; the object under study is the transformation itself which here is the beta function.

So we have

$$\beta(g_R) = -\frac{\epsilon}{2}g_R - g_R\mu\partial_\mu\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) \tag{12.15}$$

$$= -\frac{\epsilon}{2}g_R - g_R \underbrace{\beta(g_R)}_{=-\frac{\epsilon}{2}g_R+\mathcal{O}(g_R^2)} \partial_{g_R}\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) \tag{12.16}$$

$$= -\frac{\epsilon}{2}g_R + \frac{\epsilon}{2}g_R^2\partial_{g_R}\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g_R^4). \tag{12.17}$$

At the last step we solved the equation perturbatively. So we need to know how the counterterms $\delta_{1,2,3}$ depend on the coupling. We could have chosen a different term to focus on, which would have required knowing a different set of the counterterms; we'd have to get the same answer.

**Gluon vacuum polarization.** The Ward identity in QCD still implies that the gluon self-energy is transverse:

$$\mathbf{i}\Pi^{\mu\nu}_{ab}(q) = -\mathbf{i}\Pi_{ab}\left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)$$

$=$  $+$  $+$  $+$  $+$ 

$$\equiv \mathcal{M}_q + \mathcal{M}_3 + \mathcal{M}_4 + \mathcal{M}_{\text{ghost}} - \mathbf{i}\left(k^2\eta^{\mu\nu} - k^\mu k^\nu\right)\delta^{ab}\delta_3$$

through one loop. In Feynman gauge, we have

$\mathbf{i}\mathcal{M}^{\mu\nu ab}_q(q) =$  $= -\text{tr}_F(t^at^b)\left(\mathbf{i}g\right)^2\bar{\mu}^{4-D}\int đ^Dk\frac{\mathbf{i}}{(q-k)^2-m^2}\frac{\mathbf{i}}{k^2-m^2}\text{tr}\left[\gamma^\mu\left(\not{k}-\not{q}+m\right)\gamma^\nu\left(\not{k}+m\right)\right].$

There are no surprises here – it looks just like the electron loop contribution to the photon vacuum polarization. The color trace is $\text{tr}_F(t^at^b) = T_F\delta^{ab} = \frac{1}{2}\delta^{ab}$ for the fundamental representation. Since we're interested in the UV singular structure, we can simplify our lives by setting the quark masses to zero. Using exactly the same tricks as for QED, the answer is then, near $D = 4-\epsilon$,

$$\mathbf{i}\mathcal{M}^{\mu\nu ab}_q(q) = N_fT_F\left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)\delta^{ab}\frac{g^2}{16\pi^2}\left(-\frac{8}{3}\frac{1}{\epsilon} - \frac{20}{9} - \frac{4}{3}\ln\frac{\mu^2}{-q^2} + \mathcal{O}(\epsilon)\right) \tag{12.18}$$

where $N_f$ is the number of flavors of quarks (*e.g.* up, down...), counting Dirac multiplets,

More novel are the gluon and ghost loops:

$\mathbf{i}\mathcal{M}^{\mu\nu ab}_{\text{ghost}} =$  $= (-1)(-g)^2\int đ^Dk\frac{\mathbf{i}}{(k-q)^2}\frac{\mathbf{i}}{k^2}f^{cad}k^\mu f^{dbc}(k-q)^\nu$

$$\overset{(12.19)}{=} g^2\frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}}\delta^{ab}C_2(\mathsf{G})\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}\left(\frac{1}{2}\Gamma\left(2-\frac{D}{2}\right)\Delta\right) + q^\mu q^\nu\left(x(1-x)\Gamma\left(2-\frac{D}{2}\right)\right.\right.$$

The big $(-1)$ is because the ghosts are fermionic. To get to the second line, we used Feynman parameters and completed the square and did the integral over $\ell \equiv k + xq$. $\Delta \equiv x(x-1)q^2$. The new ingredient is the color stuff.

**Quadratic Casimir.** Recall that the total angular momentum $\mathbf{J}^2 = j(j+1)\mathbb{1}$ has $[\mathbf{J}^2, \vec{J}] = 0$ – it's a Casimir for $\mathsf{SU}(2)$, proportional to the identity on each irrep. This works for any Lie algebra:

$$T^2 \equiv T^a T^a \quad \text{satisfies} \quad [T^b, T^2] = 0, \forall b.$$

In any representation $r$ then we have $T_r^a T_r^a = C_2(r)\mathbb{1}_{d(r)\times d(r)}$. In particular, for the adjoint rep,

$$T^a_{\text{adj}} T^a_{\text{adj}} = (f^a)_{bc}(f^a)_{cd} = f^{abc} f^{acd} \equiv C_2(\mathsf{G})\delta^{bd}. \tag{12.19}$$

$C_2(r)$ is related to the normalization of the generators: $\text{tr}T_r^a T_r^b = c(r)\delta^{ab}$ (remember, we chose $c(\text{fundamental}) = \frac{1}{2}$). Contracting with $\delta^{ab}$ gives $d(r)C_2(r) = d(\mathsf{G})c(r)$. For the $\mathsf{SU}(N)$, $c_2(\mathsf{G}) = N$. See Peskin page 502 for a derivation.

$\mathbf{i}\mathcal{M}_3^{\mu\nu ab} =$ 

$$= \frac{g^2}{2}\bar{\mu}^{4-D}\int đ^D k \frac{-\mathbf{i}}{k^2}\frac{-\mathbf{i}}{(k-q)^2} f^{acd} f^{bcd} N^{\mu\nu} \tag{12.20}$$

$$= -\frac{g^2}{2}\frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}}\delta^{ab} C_2(\mathsf{G})\int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}A + q^\mu q^\nu B + \eta^{\mu\nu} q^2 C\right). \tag{12.21}$$

The $\frac{1}{2}$ is a symmetry factor, since gluons are real, the two internal gluon lines can be exchanged. $\Delta = x(x-1)q^2$ is the same as before.

$$A = 3(D-1)\Gamma\left(1-\frac{D}{2}\right)\Delta, B = (6(x^2-x+1)-D(1-2x)^2)\Gamma\left(2-\frac{D}{2}\right), C = (-2x^2+2x-5)\Gamma\left(2-\frac{D}{2}\right).$$

The term with $A$ represents a would-be-quadratic divergence. In dim reg this shows up as a pole at $D = 2$.

In the diagram which uses the quartic coupling, too, we find a quadratic divergence $\mathcal{M}_4 \sim \int \frac{đ^4 k}{k^2} \sim \Lambda^2$:

$\mathbf{i}\mathcal{M}_4^{\mu\nu ab} =$ 

$$= \frac{\mathbf{i}g^2}{2}\bar{\mu}^{4-D}\int đ^D k \frac{-\mathbf{i}}{k^2}\eta^{\rho\lambda}\delta^{cd}\left(f^{abe}f^{cde}\left(\delta^\mu_\lambda\delta^\nu_\rho - \delta^\mu_\rho\delta^\nu_\lambda\right)\right.$$
$$\left. + f^{ade}f^{cbe}\left(\delta^\mu_\lambda\delta^\nu_\rho - \eta^{\mu\nu}\eta_{\rho\lambda}\right) + f^{ace}f^{bde}\left(\eta^{\mu\nu}\eta_{\lambda\rho} - \delta^\mu_\rho\delta^\nu_\lambda\right)\right)$$
$$= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int \frac{đ^D k}{k^2}\frac{(q-k)^2}{(q-k)^2}$$
$$= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\left(-\frac{D}{2}\Gamma\left(1-\frac{D}{2}\right)\Delta + (1-x)^2 q^2\Gamma\left(2-\frac{D}{2}\right)\right). \tag{12.22}$$

The monstrosity in the first line is just the quartic vertex. The first term vanishes by antisymmetry $\delta^{cd} f^{cde} = 0$. At the second line we multipled by $1 = \frac{(q-k)^2}{(q-k)^2}$ in order to put the integral into the same form as the other terms. [44]

The glue contributions to the gluon vacuum polarization (not including quarks yet, since those are optional) are then

$$\mathcal{M}^{\mu\nu ab}_{\text{glue}}(q) = (\mathcal{M}_3 + \mathcal{M}_4 + \mathcal{M}_{\text{ghost}})^{\mu\nu ab} = \delta^{ab} C_2(\mathsf{G}) g^2 \frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}} \int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2} \cdot$$
$$\left[\eta^{\mu\nu}\Gamma\left(1-\frac{D}{2}\right)\Delta\left(\underbrace{-\frac{1}{2}+\frac{3(D-1)}{2}-\frac{D(D-1)}{2}}_{=-\frac{1}{2}(D-2)^2}\right) + q^\mu q^\nu \Gamma\left(2-\frac{D}{2}\right)\mathfrak{a} + \eta^{\mu\nu} q^2 \Gamma\left(2-\frac{D}{2}\right)\mathfrak{b}\right]$$

Here $\mathfrak{a} = -3(x^2-x+1) + \frac{D}{2}(1-2x)^2 + x(1-x)$ and $\mathfrak{b} = x^2 - x + \frac{5}{2} - (1-x)^2(D-1)$. The coefficient of $\Gamma\left(1-\frac{D}{2}\right)$ has a factor of $D-2$, which cancels the pole at $D=2$. Then using $\Gamma\left(1-\frac{D}{2}\right)(D-2) = -2\Gamma\left(2-\frac{D}{2}\right)$, this term combines with the other two. After some boiling using the $x \leftrightarrow 1-x$ symmetry, this is

$$\mathcal{M}^{\mu\nu ab}_{\text{glue}}(q) = \delta^{ab} C_2(\mathsf{G}) g^2 \frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}} \int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2} (\eta^{\mu\nu} - q^\mu q^\nu)\,\Gamma\left(2-\frac{D}{2}\right)\left(\left(1-\frac{D}{2}\right)(1-2x)^2 + 2\right)$$
$$\stackrel{D=4-\epsilon}{=} C_2(\mathsf{G})\,(\eta^{\mu\nu} - q^\mu q^\nu)\,\delta^{ab} \frac{g^2}{(4\pi)^2}\left(\frac{10}{3}\frac{1}{\epsilon} + \frac{31}{9} + \frac{5}{3}\ln\frac{\mu^2}{-q^2} + \mathcal{O}(\epsilon)\right).$$

Notice that compared to (12.18), the coefficient of the log (and of the pole in $\epsilon$) has the opposite sign. From this we conclude that to cancel the $\epsilon^{-1}$ pole in the vacuum polarization we must take

$$\delta_3 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(\frac{10}{3}C_2(\mathsf{G}) - \frac{8}{3}N_f T_F\right).$$

We're almost there. To get the beta function we also need $\delta_1$ and $\delta_2$.

**Quark self-energy.** This determines $\delta_2$ and $\delta_m$ (the latter we won't need). The UV bit doesn't care about the mass, so let's set $m=0$. Again it is just like QED except for the color factors. The one-loop correction to the quark self-energy is

$$\mathbf{i}\Sigma_2^{ij}(\not{p}) = \int đ^D k\, t^a_{ik} \gamma^\mu \frac{\mathbf{i}\not{k}\delta^{kl}}{k^2 + \mathbf{i}\epsilon} t^b_{lj} \gamma_\mu \frac{-\mathbf{i}\delta^{ab}}{(k-p)^2 + \mathbf{i}\epsilon}$$

[44] Actually, there is a sense in which this contribution is zero in dim reg. After the rewriting, it's still zero, but only after doing the $x$ integral. How can it affect anything then? The difference after the rewriting is merely that the cancellation of the pole at $D=2$ happens in the integrand of the $x$ integral, rather than only after integration.

The color factors are

$$t^a_{ik}t^b_{lj}\delta^{ab}\delta^{kl} = \sum_a (t^a t^a)_{ij} = C_2(F)\delta_{ij}$$

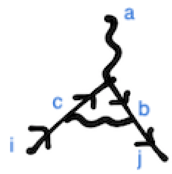
where $C_2(F)$ is the quadratic Casimir in the fundamental representation, which for $\mathsf{SU}(N)$ is $C_2(F) = \frac{N^2-1}{2N}$. The momentum integral is the same as in QED and we find

$$\Sigma_2^{ij}(\not p) = \frac{g^2}{8\pi^2}\delta^{ij}C_2(F)\frac{1}{\epsilon}\not p + \text{finite} \,. \tag{12.23}$$

The quark wavefunction renormalization counterterm $\delta_2$ contributes as $\Sigma^{ij} = ... + \delta_2 \not p$, so we must set

$$\delta_2 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(-2C_2(F)\right).$$

**Vertex correction.** The vertex correction gets two contributions at one loop.




$$= \mathbf{i}g(t^c t^a t^b)_{ij}\delta^{bc}\Gamma^\mu_{\text{QED}}$$

where

$$\Gamma^\mu_{\text{QED}} = F_1(p^2)\gamma^\mu + \frac{\mathbf{i}\sigma^{\mu\nu}}{2m}p_\nu F_2(p^2)$$

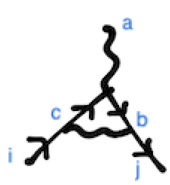
is identical to the QED answer (notice that it's useful to keep the quark mass around for a bit here) and the color factors are

$$t^c t^a t^b \delta^{bc} = t^b t^a t^b = t^b t^b t^a + t^b[t^a, t^b] = C_2(F)T^a + \mathbf{i}f^{abc}t^b t^c.$$

By antisymmetry of $f^{abc}$, the second term is

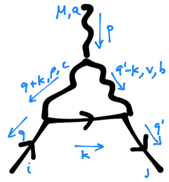
$$\mathbf{i}f^{abc}t^b t^c = \mathbf{i}f^{abc}\frac{1}{2}[t^b, t^c] = -\frac{1}{2}f^{abc}f^{bcd}t^d = -\frac{1}{2}C_2(\mathsf{G})t^a.$$

Altogether, the divergent bit of this diagram is




$$= \mathbf{i}g\left(C_2(F) - \frac{1}{2}C_2(\mathsf{G})\right)t^a_{ij}\gamma^\mu\frac{g^2}{16\pi^2}\left(\frac{2}{\epsilon} + \ln\frac{\mu^2}{-p^2} + \text{finite}\right).$$

The other diagram is new:




$$= \mathbf{i}gf^{abc}(t^c t^b)_{ij}\Gamma^\mu_{\text{new}}$$

with

$$-\mathbf{i}g\Gamma^{\mu}_{\text{new}}(p^2) = (\mathbf{i}g)^2 g\bar{\mu}^{4-D}\int đ^D k \gamma^\rho \frac{\mathbf{i}}{\not{k}}\gamma^\nu \frac{-\mathbf{i}}{(q+k)^2}\frac{-\mathbf{i}}{(q'-k)^2}\cdot$$
$$\left(\eta^{\mu\nu}(2q+q'+k)^\rho + \eta^{\nu\rho}(-q+q'-2k)^\mu + \eta^{\rho\mu}(k-2q'-q)^\nu\right)$$

(I find the opposite sign from Schwartz here. This sign cancels against the one in (12.24).) The horrible numerator comes from the 3-gluon vertex, but in computing the UV divergence we can set the external momenta to zero. This gives

$$\begin{aligned}
-\Gamma^{\mu}_{\text{new}}(0) &= g^2\bar{\mu}^{4-D}\int đ^D k \frac{\gamma_\rho \not{k}\gamma_\nu}{k^6}\left(\eta^{\mu\nu}k^\rho - 2\eta^{\nu\rho}k^\mu + \eta^{\rho\mu}k^\nu\right) \\
&= g^2\bar{\mu}^{4-D}\int \frac{đ^D k}{k^6}\left(2k^2\gamma^\mu - 2\gamma_\rho \not{k}\gamma^\rho k^\mu\right) \\
&= g^2\left(4-\frac{4}{D}\right)\gamma^\mu \bar{\mu}^{4-D}\int\frac{đ^D k}{k^4} \qquad \gamma_\rho\gamma^\nu\gamma^\rho = (2-D)\gamma^\nu, \int k^\mu k^\nu ... = \int \frac{k^2}{D}\eta^{\mu\nu}... \\
&= g^2 \mathbf{i}\gamma^\mu \frac{g^2}{16\pi^2}\left(\frac{6}{\epsilon} + 3\log\frac{\mu^2}{-p^2} + \text{finite}\right)
\end{aligned}$$

where at the last step we put back the gluon momentum to make up the dimensions. Finally the color factor is

$$f^{abc}(t^c t^b) = \frac{1}{2}\mathbf{i}f^{abc}f^{cbd}t^d = -\mathbf{i}\frac{1}{2}C_2(\mathsf{G})t^a. \tag{12.24}$$

Altogether, the divergent part of the qqg vertex at one loop is then

$$\frac{1}{\epsilon}\mathbf{i}gt^a_{ij}\gamma^\mu\left(\left(2\left(C_2(F)-\frac{1}{2}C_2(\mathsf{G})\right)+3C_2(\mathsf{G})\right)\frac{g^2}{16\pi^2}+\delta_1\epsilon\right) \implies \delta_1 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(-2C_2(F)-2C_2(\mathsf{G})\right).$$

Combining all of this information using (12.17), the QCD beta function is (dropping the $R$ subscripts on $g_R$)

$$\begin{aligned}
\beta(g) &= -\frac{\epsilon}{2}g + \frac{\epsilon}{2}g^2\partial_g\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g^4) \\
&\overset{D\to 4}{=} \frac{g^3}{16\pi^2}\left(-2C_2(F) - 2C_2(\mathsf{G}) - \frac{1}{2}\left(\frac{10}{3}C_2(\mathsf{G}) - \frac{8}{3}N_f T_F\right) - (-2C_2(F))\right) + \mathcal{O}(g^4) \\
&= -\frac{g^3}{16\pi^2}\left(\frac{11}{3}C_2(\mathsf{G}) - \frac{4}{3}N_f T_F\right) \\
&\overset{\mathsf{SU}(N)\text{ with fundamental quarks}}{=} -\frac{g^3}{16\pi^2}\left(\frac{11}{3}N - \frac{2}{3}N_f\right).
\end{aligned}$$

If there are not too many species of quarks ($N_f < 6N = 18$, which is true in the SM), $\beta$ is negative, in which case such a non-Abelian gauge theory is asymptotically free, as promised many times. Defining a running coupling as in (7.45), we find a crucial plus sign relative to (7.45)

$$g_{\text{eff}}^2(q^2) = \frac{g^2}{1 + \frac{g^2}{16\pi^2} C \log\left(\frac{|q^2|}{M^2}\right)}, \quad C \equiv \frac{11}{3}N - \frac{2}{3}N_f$$

and the coupling grows as $q$ decreases, and shrinks at large $q$. Actually it is a bit tricky to define the effective coupling in QCD, but (a more precise version of) this curve has been measured (see Peskin fig. 17.23 and Schwartz §26.3).

**Qualitative picture of asymptotic freedom.** [Peskin §16.7] The sign of the beta function in QED can be understood as charge screening by the vacuum – electron-positron pairs fluctuate into existence, and respond to the presence of a source in such a way as to decrease its field at long distance.

How does non-Abelian gauge theory manage to produce antiscreening? There is certainly still screening from the quarks, and since the gluons are charged, they will also produce screening. So it makes sense that too many quarks will spoil the soup. But whence the terms of the opposite sign in the beta function?

Following Peskin §16.7, consider pure (no quarks) $\mathsf{SU}(2)$ gauge theory, in Coulomb gauge $\partial_i A^{ia} = 0$. In this gauge, we sacrifice Lorentz covariance for more manifest unitarity – no ghosts, and no longitudinal and timelike polarization states. The equation of motion for $A^{0a}$ is the Gauss law (in terms of $E^{ia} \equiv F^{0ia}$):

$$g\rho^a = D_i E^{ia} = \partial_i E^{ia} + g f^{abc} A_i^b E^{ic},$$

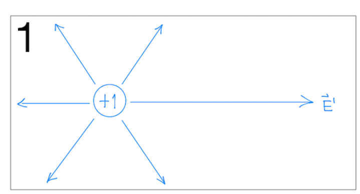
where $\rho$ is the charge density (*e.g.* the number density of quarks if we included them in the theory), and for $\mathsf{SU}(2)$, the structure constants are $f^{abc} = \epsilon^{abc}$. Instead of dynamical quarks, let's consider a static color source particle: $\rho^a(x) = \delta^{(3)}(\vec{x})\delta^{a1}$ pointing in a particular color direction, so the equation we wish to solve is

$$\partial_i E^{ia} = g\delta^{(3)}(\vec{x})\delta^{a1} + g f^{abc} A^{bi} E^{ic}. \tag{12.25}$$

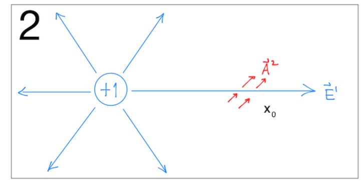
Let's solve this perturbatively in $g$ in several steps.

1. At leading order, the source produces a Coulomb field:

   $\vec{E}^a(x) = \frac{g\delta^{1a}\check{x}}{x^2}$. So far, this is just classical physics.

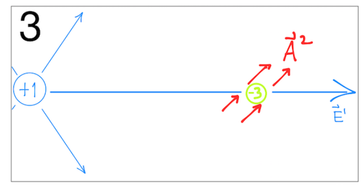


2. The quantum mechanics comes in here:
   consider a fluctuation of the vector potential in the 2d color direction $A^{b=2,i}(\vec{x})$, with support localized somewhere, call it $x_0$, away from the origin. Suppose it points in some direction, somewhat aligned with $\vec{x}$, its displacement from the source.

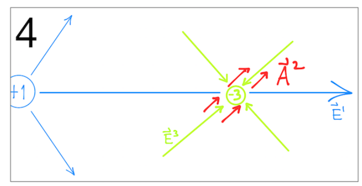



3. Here comes the iteration.

   The second term on the RHS of (12.25) is then $g\epsilon^{abc}\vec{A}^b \cdot \vec{E}^c \propto -\delta^{c3}\vec{A}^2 \cdot \vec{E}^1$, a sink for the color-electric field in the 3rd color direction.



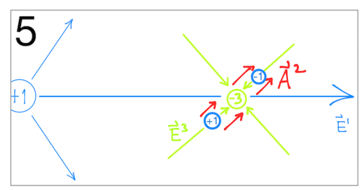

4. This produces a new Coulomb field $\vec{E}^3(\vec{x}) \sim -\frac{\vec{x}-\vec{x}_0}{|x-x_0|^3}$ pointing towards $x_0$.



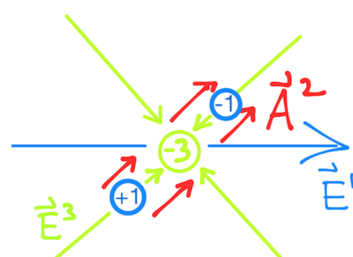

5. Now look at the second term on the RHS of (12.25) again:

$$\vec{\nabla} \cdot \vec{E}^1 = ... + g\epsilon^{123}\vec{A}^2 \cdot \vec{E}^3$$

   it is a *source* (*sink*) for the color field in direction 1 where $\vec{A}^2$ and $\vec{E}^3$ are parallel (antiparallel).




But if the fluctuation $\vec{A}^2$ points away from the source, then in the region closer to the source, $\vec{A}^2 \cdot \vec{E}^3 > 0$, and farther from the source they are anti-aligned. This produces a dipole source for $\vec{E}^1$ which points *toward* the original charge, and therefore anti-screens its field.




Warning: on the other hand, if the fluctuation $\vec{A}^2$ points toward the original source then this process produces a dipole pointing away from the original source, which contributes to screening. I'm not sure if this picture can be made quantitative.

**Disclaimer.** This discussion just scratches the surface of the physics of QCD! Many measurable phenomena can be calculated using the machinery we've set up. Please see Peskin §17 and Schwartz §32. [End of Lecture 38]