# Physics 215B: Particles and Fields
# Winter 2017

Lecturer: McGreevy

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

Last updated: 2017/03/15, 23:19:56

# 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

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

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

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

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.

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.

# 6 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.

## 6.1 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 exponents 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

$$\begin{aligned} \left(-\frac{1}{2}\nabla^2 - E\right)\psi(q) &= -V(q)\psi(q) \\ \int đ^2 p e^{ip\cdot q}\left(\frac{p^2}{2} - E\right)\varphi(p) &= +g_0\delta^2(q)\psi(0) \\ &= +g_0\left(\int đ^2 p e^{\mathbf{i}p\cdot q}\right)\psi(0) \end{aligned} \tag{6.1}$$

which (integrating the both-hand side of (6.1) over $q$: $\int d^2q e^{\mathbf{i}p\cdot q}\left((6.1)\right)$ ) 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 on 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 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:

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

So in our cutoff theory, the boundstate condition is:

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

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 $\epsilon_B$.

$$\epsilon_B = \frac{\Lambda^2}{2}\frac{1}{e^{\frac{2\pi\hbar^2}{g_0}}-1} . \tag{6.2}$$

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 (6.2):

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

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 coupilngs.)

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 $\Lambda$. But the general 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.

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.

## 6.2 A simple example of perturbative renormalization in QFT

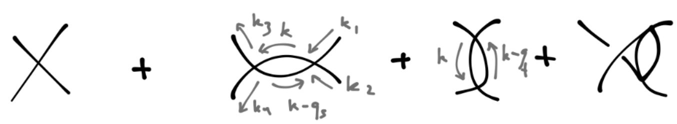
[Zee §III.1, Schwartz §15.4] Now let's consider an actual field theory but a simple one, namely the theory of a real scalar field in four dimensions, with

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - m^2\phi^2 - \frac{g}{4!}\phi^4. \tag{6.3}$$

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

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

**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'.

[End of Lecture 21]

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:

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

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.)

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

$$\begin{aligned} -\mathbf{i}g &= -\mathbf{i}g_P - \mathbf{i}Cg^2 L_0 + \mathcal{O}(g^3) \\ &= -\mathbf{i}g_P - \mathbf{i}Cg_P^2 L_0 + \mathcal{O}(g_P^3). \end{aligned} \tag{6.4}$$

and eliminate $g$ from the discussion:

$$\begin{aligned} \mathbf{i}\mathcal{M}(s,t,u) &= -\mathbf{i}g + \mathbf{i}Cg^2 L + \mathcal{O}(g^3) \\ &\overset{(6.4)}{=} -\mathbf{i}g_P - \mathbf{i}Cg_P^2 L_0 + \mathbf{i}Cg_P^2 L + \mathcal{O}(g_P^3) \\ &= -\mathbf{i}g_P + \mathbf{i}Cg_P^2 \left( \log \frac{s_0}{s} + \log \frac{t_0}{t} + \log \frac{u_0}{u} \right) + \mathcal{O}(g_P^3). \end{aligned} \tag{6.5}$$

[4] You might hesitate here about my referring to the amplitude $\mathcal{M}$ as an 'observable'. The difficult 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.

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 §6.1, it was eliminated by letting the coupling vary with it, $g = g(\Lambda)$. 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 (6.3), with $m = 0$ for simplicity):

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 \tag{6.6}$$

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 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, and we satisfy the observational condition $\mathcal{M}(s_0, t_0, u_0) = -g_P$.

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.

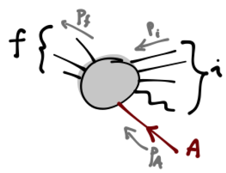
The step whereby we were able to absorb all the dependence on the cutoff into the bare coupling constant involved some apparent magic. It is not so clear that the same magic will happen if we study the next order $\mathcal{O}(g_P^3)$ terms. A QFT where all the cutoff dependence to all orders can be removed 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.

## 6.3 Classical interlude: Mott formula

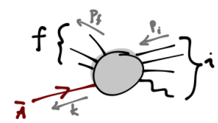
As a prelude to studying loops in QED, and to make clear what is at stake, I want to fill a hole in our discussion of last quarter. By studying scattering of an electron from a heavy charged fermion (a muon is convenient) we will reconstruct the cross section for scattering off a Coulomb potential (named after Mott). Then we'll figure out how it is corrected by other QED processes.

**Crossing symmetry.** If you look at a Feynman diagram on its side (for example because someone else fails to use the convention that time goes to the left) it is still a valid amplitude for some process. Similarly, dragging particles between the initial and final state also produces a valid amplitude. Making this relation precise can save us some work. The precise relation for dragging an incoming particle into the final state, so that it is an outgoing antiparticle, is:

$$\mathbf{i}\mathcal{M}_{f\leftarrow iA}(p_f; p_i, p_A) =$$




$$= \mathbf{i}\mathcal{M}_{\bar{A}f\leftarrow i}(p_f, k = -p_A; p_i) =$$




.

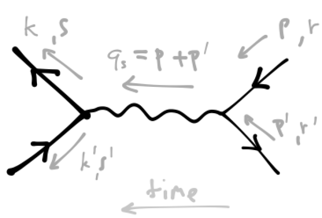
(If you must, note that this is another sense in which an antiparticle is a particle going backwards in time.) If $A$ is a spinor particle, the sum relations for particles and antiparticles are different:

$$\sum_r u^r(p)\bar{u}^r(p) = \not{p} + m, \qquad \sum_r v^r(k)\bar{v}^r(k) = \not{k} - m = -(\not{p} + m)$$

– after accounting for $k = -p_A$, they differ by an overall sign. Hence we must also append a fermion sign factor $(-1)^{\text{number of fermions shuffled between in and out}}$ in the unpolarized scattering probability. Here is an example.

$\boxed{\mu^+\mu^- \leftarrow e^+e^-}$. For example, we studied the process $\mu^+\mu^- \leftarrow e^+e^-$ in some detail at the very end of last quarter. To try to keep things straight, I'll call the electron momenta $p, p'$ and the muon momenta $k, k'$, since that won't change under crossing. We found the amplitude

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




$$= \left(-\mathbf{i}e\bar{u}^s(k)\gamma^\mu v^{s'}(k')\right)_{\text{muons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu q_\nu}{q^2}\right)}{q^2} \left(-\mathbf{i}e\bar{v}^{r'}(p')\gamma^\nu u^r(p)\right)_{\text{electrons}} \tag{6.7}$$

(with $q \equiv p + p' = k + k'$)[5] and the (unpolarized) scattering probability density

$$\frac{1}{4}\sum_{\text{spins}} |\mathcal{M}|^2 \overset{\text{spinor traces}}{=} \frac{1}{4}\frac{e^4}{s^2} E^{\mu\nu} M_{\mu\nu},$$

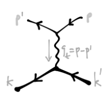
[5]Relative to the notation I used last quarter, $p_1 = p, p_2 = p', p_3 = k, p_4 = k'$.

where the tensor objects $E^{\mu\nu}, M^{\mu\nu}$ come respectively from the electron and muon lines,

$$\frac{1}{4}E_{\mu\nu} = p_\mu p'_\nu + p'_\mu p_\nu - \eta_{\mu\nu}(p \cdot p' + m_e^2)$$
$$\frac{1}{4}M_{\mu\nu} = k_\mu k'_\nu + k'_\mu k_\nu - \eta_{\mu\nu}(k \cdot k' + m_\mu^2).$$

and they are contracted by the photon line, with $s = q^2 = (p + p')^2$.

$\boxed{e^-\mu^- \leftarrow e^-\mu^-}$. To get from this the amplitude (tree level, so far) for the process $e^-\mu^- \leftarrow e^-\mu^-$, we must move the incoming positron line to an outgoing electron line, and move the outgoing antimuon line to an incoming muon line (hence the sign in $\sigma$ will be $(-1)^{\text{number of fermions shuffled between in and out}} = (-1)^2 = 1$). Relative to the amplitude for $\mu^+\mu^- \leftarrow e^+e^-$ (6.7), we must replace the relevant $v$s with $u$s for the initial/final antiparticles that were moved into final/initial particles, and we must replace $p' \to -p', k' \to -k'$:

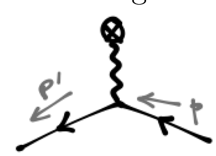
$$\mathbf{i}\mathcal{M} = $$

$$= (-\mathbf{i}e\bar{u}(p')\gamma^\mu u(p)))_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu^t q_\nu^t}{q_t^2}\right)}{q_t^2} (-\mathbf{i}e\bar{u}(k)\gamma^\nu u(k'))_{\text{muons}} \quad (6.8)$$

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

$$\frac{1}{4}\sum_{s,s',r,r'} |\mathcal{M}|^2 = 4\frac{e^4}{t^2}\left(-p_\mu p'_\nu - p'_\mu p_\nu - \eta_{\mu\nu}(-p \cdot p' + m_e^2)\right)$$
$$\cdot \left(-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k \cdot k' + m_\mu^2)\right) \quad (6.9)$$

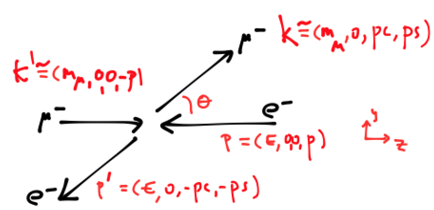
On the Mandelstam variables, this is just the permutation $(s, t, u) \to (t, u, s)$.

**Payoff: the Mott formula.** Recall other ways of figuring out the scattering cross section from a Coulomb potential from a point charge of charge $ze$.




We think about scattering from a fixed electrostatic potential $A_0 = \frac{ze}{r}$ and do classical mechanics. I can never remember how this goes. Instead, let's just scatter an electron off a heavy charge, such as a muon. If the charge of the heavy object were $z$ times that of the electron, we would multiply the amplitude by $z$ and the cross section by $z^2$.

'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, $c \equiv \cos\theta, s \equiv \sin\theta$.



$$-\frac{1}{4}M_{\mu\nu} \simeq k_\mu k'_\nu + k'_\mu k_\nu - \eta_{\mu\nu}\left(\underbrace{k\cdot k' - m_\mu^2}_{=m_\mu^2-m_\mu^2=0}\right) \simeq \delta_{\mu 0}\delta_{\nu 0}2m_\mu^2.$$

This means that the muon-line tensor factor $M_{\mu\nu}$ in (6.9) simplifies dramatically: In the electron line, we get

$$-p\cdot p' + m_e^2 = -E^2 + \vec{p}^2\cos\theta + m_e^2 = -\vec{p}^2(1-\cos\theta). \tag{6.10}$$

So

$$\begin{aligned} E^{\mu\nu}M_{\mu\nu} = 32m_\mu^2 E^{00} &= 32m_\mu^2(2E^2 + \eta^{00}(p\cdot p' - m_e^2)) \\ &\overset{(6.10)}{=} 32m_\mu^2(2E^2 - \vec{p}^2(1-\cos\theta)) \\ &\overset{\text{trig}}{=} 32m_\mu^2 2(E^2 - \vec{p}^2\sin^2\theta/2) \overset{\beta^2\equiv\vec{p}^2/E^2}{=} 64m_\mu^2E^2(1-\beta^2\sin^2\theta/2). \end{aligned} \tag{6.11}$$

Noting that $t = (p-p')^2 = -2\vec{p}^2(1-\cos\theta)$, the cross section is

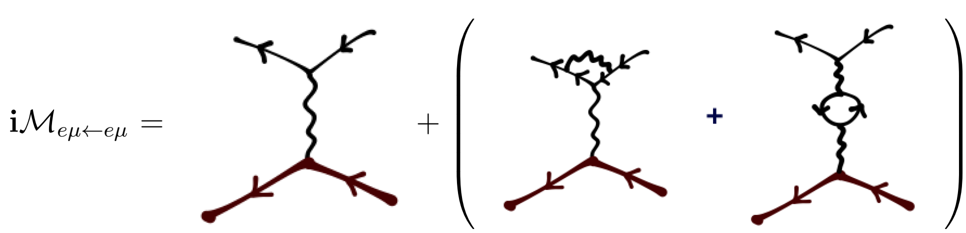
$$\begin{aligned} d\sigma &= \underbrace{\frac{1}{v_{\text{rel}}}}_{=\beta}\frac{1}{2E}\frac{1}{2m_\mu}\frac{z^2e^4}{t^2}64m_\mu^2E^2(1-\beta^2\sin^2\theta/2)\frac{d\Omega}{16\pi^2}\frac{p}{E_{\text{total}}} \\ &\overset{E_{\text{total}}\sim m_\mu}{=} \frac{4E}{\beta}\frac{z^2e^4(1-\beta^2\sin^2\theta/2)}{t^2}d\Omega \end{aligned}$$

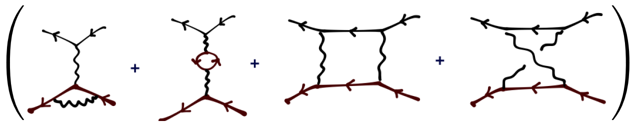
from which we get

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

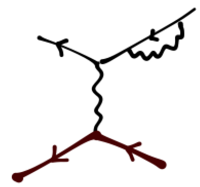
If we take $\beta \ll 1$ in this formula we get the Rutherford formula. Notice that it blows up at $\theta \to 0$. This is a symptom of the long-range nature of the Coulomb potential, *i.e.* the masslessness of the photon.

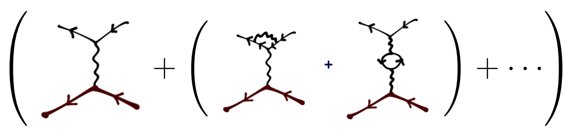
**Radiative corrections.** Now it's time to think about perturbative corrections to this cross section. Given that the leading-order calculation reproduced the classical physics of the Coulomb potential, you can think of what we are doing as effectively discovering (high-energy or short-distance) quantum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



$+$  $+\,\mathcal{O}(e^3)$

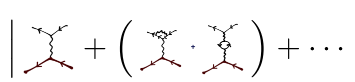
• What do the one-loop diagrams in the second line have in common? They have an internal muon line. Why does this matter? When the energy going through the line is much smaller than the muon mass, then the propagator is $\frac{\mathbf{i}(\not{k}+m_\mu)}{k^2-m_\mu^2} \sim \frac{1}{m_\mu}$ and its relative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

$S_{e\mu\leftarrow e\mu} = \sqrt{Z_e}^2\sqrt{Z_\mu}^2$  amputated, on-shell

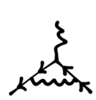
and in determining the locations of the poles whose residues are the $S$-matrix elements.

• 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

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

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

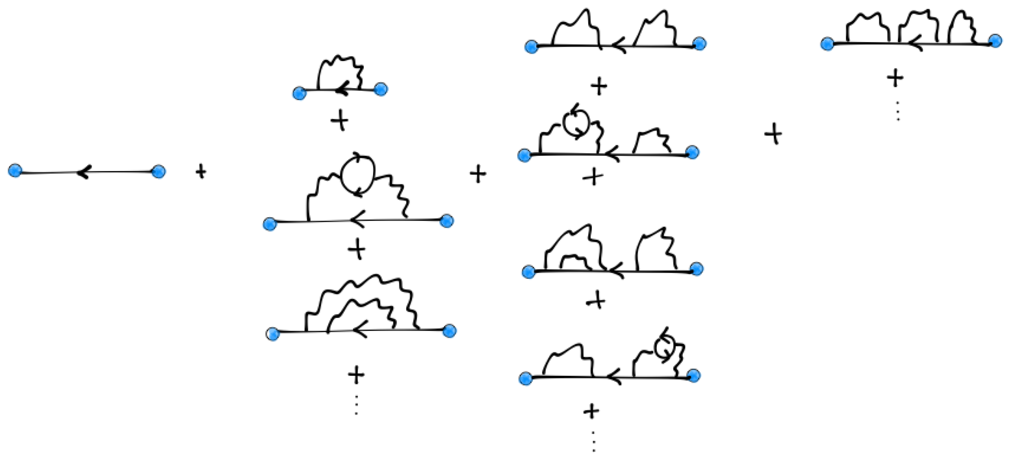
In the above discussion, we encounter all three 'primitive' one-loop divergent amplitudes of QED, which we'll study in turn:

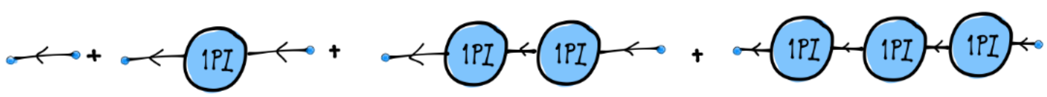
- electron self-energy: 
- vertex correction: 
- vacuum polarization: 

[End of Lecture 22]

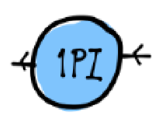
## 6.4 Electron self-energy in QED

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

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

$=$  (6.12)

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:

$$\begin{aligned} \tilde{G}^{(2)}(p) &= \mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \cdots \\ &= \mathbf{i}S\left(1 + \Sigma S + \Sigma S \Sigma S + \cdots\right) = \mathbf{i}S\frac{1}{1 - \Sigma S} \\ &= \frac{\mathbf{i}}{\not{p} - m_0}\frac{1}{1 - \Sigma\frac{1}{\not{p} - m_0}} = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)}. \end{aligned} \tag{6.13}$$

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;

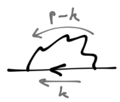
we could do these manipulations in the eigenbasis of $\not{p}$. This fully corrected propagator has a pole at

$$\not{p} = m \equiv m_0 + \Sigma(m) \ . \tag{6.14}$$

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 (6.12), 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:

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

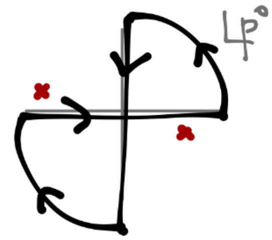
Step 2: Now we can complete the square

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

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

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

where the numerator is

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

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 $\ell$

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

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

$$\begin{aligned}\Sigma_2(p) &= e^2 \int_0^1 dx \int \frac{\ell^2 d\ell^2}{2}\frac{(2\pi^2)}{(2\pi)^4}\frac{2(2m_0 - x\not{p})}{\left(\ell^2 + \Delta\right)^2} \\ &= \frac{e^2}{8\pi^2}\int_0^1 dx (2m_0 - x\not{p})\mathcal{J} \end{aligned} \tag{6.15}$$

with

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

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

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

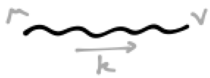
You knew this UV divergence was coming. To be more precise, let's add zero:

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

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.

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 by 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

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

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

$$\Sigma_2(p)_{PV} = \frac{\alpha}{2\pi}\int_0^1 dx(2m_0 - x\not{p}) \ln \frac{x\Lambda^2}{x\mu^2 + (1-x)m_0^2 - x(1-x)p^2}. \tag{6.16}$$

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 (6.14): the shift in the mass a a result of this one-loop self-energy is

$$\begin{aligned}
\delta m \equiv m - m_0 &= \Sigma_2(\not{p} = m) + \mathcal{O}(e^4) = \Sigma_2(\not{p} = m_0) + \mathcal{O}(e^4) \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \ln \underbrace{\frac{x\Lambda^2}{x\mu^2 + (1-x)m_0^2 + x(1-x)m_0^2}}_{\equiv f(x,m_0,\mu)} \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \left( \underbrace{\ln \frac{\Lambda^2}{m_0^2}}_{\text{divergent}} + \underbrace{\ln \frac{x m_0^2}{f(x,m_0,\mu)}}_{\text{relatively small}} \right) \\
&\approx \frac{\alpha}{2\pi}\left(2 - \frac{1}{2}\right) m_0 \ln \frac{\Lambda^2}{m_0^2} = \frac{3\alpha}{4\pi} m_0 \ln \frac{\Lambda^2}{m_0^2}.
\end{aligned} \tag{6.17}$$

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.

[End of Lecture 23]

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

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)} \overset{p\sim m}{\simeq} \frac{\mathbf{i}Z}{\not{p} - m} + \text{regular terms.} \tag{6.18}$$

The residue of the pole at the electron mass is no longer equal to one, but rather $Z$. To see what is, 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 (6.18) becomes

$$\tilde{G}^{(2)}(p) \overset{p\sim m}{\sim} \frac{\mathbf{i}}{\not{p} - m - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}(\not{p} - m)} = \frac{\mathbf{i}}{(\not{p} - m)\left(1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}\right)} \tag{6.19}$$

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 order

$$\begin{aligned}\delta Z = \frac{\partial \Sigma_2}{\partial \not{p}}|_{m_0} &\overset{(6.16)}{=} \frac{\alpha}{2\pi}\int_0^1 dx \left(-x \ln \frac{x\Lambda^2}{f(x, m_0, \mu)} + (2m_0 - xm_0)\frac{-2x(1-x)}{f(x, m_0, \mu)}\right) \\ &= -\frac{\alpha}{4\pi}\left(\ln\frac{\Lambda^2}{f} + \text{finite}\right).\end{aligned} \tag{6.20}$$

Here $f = f(x, m_0, \mu)$ is the same quantity defined in the second line of (6.17). 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.

## 6.5 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.

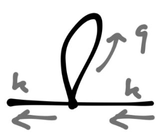
### 6.5.1 Self-energy in $\phi^4$ theory

[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:

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

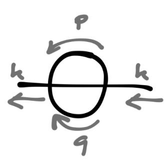
Here is a way in which $\phi^4$ theory is weird: At one loop there is no wavefunction renormalization. That is,

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


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

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 (6.19)), this diagram only contributes to the mass renormalization.

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

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


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

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 $\int^{\Lambda} \frac{d^8 P}{P^6} \sim \Lambda^2$, 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

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

The $\cdots$ here includes both higher orders in $g$ ($\mathcal{O}(g^3)$) and higher powers of $k^2$, *i.e.* higher derivative terms.

Therefore, the propagator is

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

with

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

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 (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 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

 $= -\mathbf{i} g_P,$  $= \dfrac{\mathbf{i}}{k^2 - m_P^2 + \mathbf{i}\epsilon}$ (6.21)

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 (6.21) 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)?

### 6.5.2 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 $\mathcal{A} \sim \Lambda^{D_{\mathcal{A}}}$. 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

$B_I \equiv$ # of internal scalar lines (*i.e.*, propagators)
$V \equiv$ # of $\phi^4$ vertices
$L \equiv$ # of loops

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

? 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,

$$L = B_I - (V - 1). \tag{6.22}$$

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, (6.22) 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

$$B_I = 2V - B_E/2. \tag{6.23}$$

Now we count powers of momenta:

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

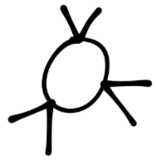
Since we are interested in the UV structure, I've set the mass to zero, as well as all the external momenta. The only scale left in the problem is the cutoff, so the dimensions of $\mathcal{A}$ must be made up by the cutoff:

$$\begin{aligned} D_{\mathcal{A}} = [\mathcal{A}] &= DL - 2B_I \\ &\overset{(6.22)}{=} B_I(D-2) - D(V-1) \\ &\overset{(6.23)}{=} D + \frac{2-D}{2}B_E + V(D-4). \end{aligned}$$

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

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

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.

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]

[End of Lecture 24]

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 Yukawa interaction:

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

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 $V_y + 4V_g = 2B_E + B_I$. The number of loops is $L = B_I + F_I - (V_y + V_g - 1)$. Putting these together (I used Mathematica) we get

$$D_{\mathcal{A}} = D + (D-4)\left(V_g + \frac{1}{2}V_y\right) + B_E\left(\frac{2-D}{2}\right) + F_E\left(\frac{1-D}{2}\right). \tag{6.24}$$

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 in

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

$$\mathcal{I} = \int^{\Lambda} \frac{d^4 p}{p^{10}} \int^{\Lambda} d^4 k.$$

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.

between). 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. 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.

### 6.5.3 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 $\phi$:

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

Notice that the $D=1$ case agrees with our quantum mechanics counting from §6.1. Quantum field theory in $D=1$ spacetime dimensions *is* quantum mechanics. (Quantum field theory in $D=0$ spacetime dimensions is integrals. This sounds trivial but it actually has some useful lessons for us in the form of *random matrix theory* and for understanding the large-order behavior of perturbation theory and its relation to non-perturbative effects. More later on this, I hope.)

Then the self-interaction term has dimensions

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

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 $g(\phi)(\partial\phi)^2$, where the coupling $g(\phi)$ is a function of $\phi$. This allows for scale-invariant non-linear sigma models; we will explore this further later on.

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

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

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

$$[g] = D \cdot \frac{p - p_D}{p_D} \begin{cases} > 0 \text{ when } p > p_D, & \text{non-renormalizable or irrelevant} \\ = 0 \text{ when } p = p_D, & \text{renormalizable or marginal} \\ < 0 \text{ when } p < p_D, & \text{super-renormalizable or relevant.} \end{cases} \tag{6.25}$$

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

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

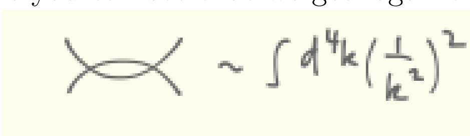
So if $[g] > 0$, $c_n$ must have more and more powers of some mass (inverse length) as $n$ increases. What dimensionful quantity makes up the difference? Sometimes it is masses or external momenta. But generically, it gets made up by UV divergences (if everything is infinite, dimensional analysis can 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.

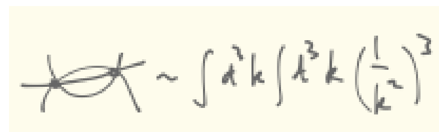
Consider the renormalizable (classically scale invariant) case: $[c_n] = [f]$, since $[g] = 0$. But in fact, what you'll get is something like

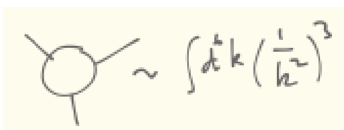
$$c_n = \tilde{c}_n \log^{\nu(n)} \left( \frac{\Lambda_{UV}}{\Lambda_{IR}} \right) ,$$

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

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 $D = 3$: 

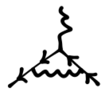

$\phi^3$ in $D = 6$: 

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 (6.25) are somewhat old-fashioned and come from a high-energy physics point of view where the short-distance physics is unkown, 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). The latter terms are natural in a 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.

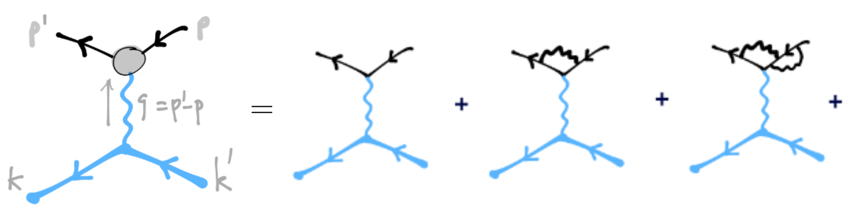
## 6.6 Vertex correction in QED

[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. 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! (Actually it is zero, but the one-loop correction is infinity.) 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.

  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.




$$\equiv \mathbf{i}\mathcal{M} \qquad = \mathbf{i}e^2 \left(\bar{u}(p')\Gamma^\mu(p,p')u(p)\right) \frac{1}{q^2}\bar{u}(K')\gamma_\mu u(K) \tag{6.26}$$

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 $q = p' - p$. Then $q^2 = 2m^2 - 2p' \cdot p$. We'll assume that the electron momenta $p, p'$ are on-shell, but $q^\mu$ is not, as in the $e\mu$ scattering process.

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

$$\Gamma^\mu(p,p') = A\gamma^\mu + B(p+p')^\mu + C(p-p')^\mu \tag{6.27}$$

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:

$$0 \overset{\text{Ward}}{=} q_\mu \bar{u}(p')\Gamma^\mu u(p) \overset{(6.27)}{=} \bar{u}(p')\left( A \underbrace{\not{q}}_{=\not{p}'-\not{p} \overset{\bar{u}(p')\ldots u(p)}{=} m-m=0} + B\underbrace{(p+p')\cdot(p-p')}_{=m^2-m^2=0} + Cq^2 \right) u(p)$$

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 (actually this is why we didn't include a $\gamma^{\mu\nu}$ term. 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* $F_{1,2}$:

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

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. The first term at zero momentum $eF_1(q^2 = 0)$ is the electric charge of the electron (if you don't believe it, there are some details on p. 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 hw 9 of 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 $2F_2(q^2 = 0)$.

[End of Lecture 25]

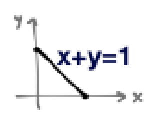
Now that we have some expectation about the form of the answer, and some ideas about what it's for, we sketch the evaluation of the one-loop QED vertex correction:


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

with $k' \equiv k + q$.

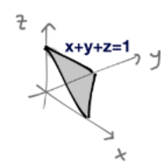
(1) Feynman parameters again. The one we showed before can be rewritten more symmetrically as:

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




Now how can you resist the generalization[8]:

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




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

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

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

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

[8] Peskin outlines a proof by induction of the whole family of such identities on page 190.

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

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

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:

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

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.

The numerator is

$$\begin{aligned} N^\mu &= \bar{u}(p')\gamma^\nu\left(\not{k}+\not{q}+m_e\right)\gamma^\mu\left(\not{k}+m_e\right)\gamma_\nu \\ &= -2\left(\mathcal{A}\bar{u}(p')\gamma^\mu u(p)+\mathcal{B}\bar{u}(p')\sigma^{\mu\nu}q_\nu u(p)+\mathcal{C}\bar{u}(p')q^\mu u(p)\right) \end{aligned} \tag{6.28}$$

where

$$\begin{aligned} \mathcal{A} &= \frac{1}{2}\ell^2+(1-x)(1-y)q^2+(1-4z+z^2)m^2 \\ \mathcal{B} &= \mathbf{i}mz(1-z) \\ \mathcal{C} &= m(z-2)(x-y)\ . \end{aligned} \tag{6.29}$$

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 is a correction to the charge of the electron and will be UV divergent. More explicitly, we get

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

The other bits are finite, and we ignore the terms that go like negative powers of $\Lambda$.

### 6.6.1 Anomalous magnetic moment

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

$$\begin{aligned} F_2(q^2) &= \frac{2m}{e} \cdot (\text{the term with } \mathcal{B}\ ) \\ &= \frac{2m}{e} 4e^3 m \int dxdydz\delta(x+y+z-1)z(1-z) \underbrace{\int \frac{đ^4\ell}{(\ell^2-\Delta)^3}}_{=\frac{-\mathbf{i}}{32\pi^2\Delta}} \\ &= \frac{\alpha}{\pi} m^2 \int dxdydz\delta(x+y+z-1) \frac{z(1-z)}{(1-z)^2m^2 - xyq^2}. \end{aligned} \tag{6.30}$$

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: $g = 2.00232$.

### 6.6.2 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 $m_\gamma = 0$)

$$\int dxdydz\delta(x+y+z-1)\frac{m^2(1-4z+z^2)}{\Delta} = m^2 \int^1 dz \int_0^{1-z} dy \frac{-1+(1-z)(3-z)}{(1-z)^2m^2} = \int^1 dz \frac{-2}{(1-z)} + \text{finite}.$$

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

$$\Gamma^\mu = \gamma^\mu \left(1 - \frac{\alpha}{2\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{{m_\gamma}^2}\right)\right) + \text{ stuff which is finite as } m_\gamma \to 0. \tag{6.31}$$

Notice that the IR divergent stuff is independent of $p$. 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 (6.31) 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. e

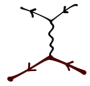
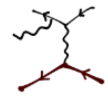
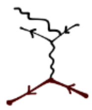
We can be more explicit if we consider $-q^2 \gg m_e^2$ so that we can ignore the electron mass everywhere. Then we could choose the counterterm $\delta_1$ so that

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

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

$$\begin{aligned} f(q^2)|_{m_e=0} &= \frac{e^2}{8\pi^2}\int dxdydz\delta(x+y+z-1)\left(\underbrace{\ln\frac{(1-x-y)\Lambda^2}{\Delta}}_{\text{IR finite}} + \frac{q^2(1-x)(1-y)}{-xyq^2+(1-x-y)m_\gamma^2}\right) \\ &= -\frac{e^2}{16\pi^2}\left(\ln^2\frac{-q^2}{m_\gamma^2} + 3\ln\frac{-q^2}{m_\gamma^2}\right) + \text{finite}. \end{aligned}$$

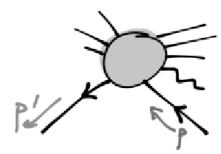
In doing the integrals, we had to remember the $\mathbf{i}\epsilon$ in the propagators, which can be reproduced by the replacement $q^2 \to q^2 + \mathbf{i}\epsilon$. This $\ln^2(q^2/m_\gamma)$ is called a *Sudakov double logarithm.* Notice that taking differences of these at different $q^2$ will not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

[End of Lecture 26]

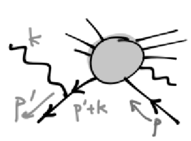
Being more inclusive, and allowing also photons in the final state, we must consider amplitudes of the form

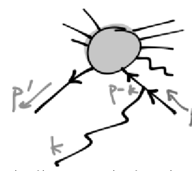
$$\bar{u}(p')\mathcal{M}_0(p',p)u(p) \equiv -\mathbf{i}\left( \quad \right)$$




in terms of which the more inclusive amplitudes look like

$$-\mathbf{i}\left( \quad + \quad \right)$$




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

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.* $(p-k)^2 - m^2 = p^2 - m_e^2 - 2p\cdot k + k^2 \sim -2p\cdot k$ since the electron is on shell and $k \ll p$. Therefore

$$\mathcal{M}\left(e\mu + \text{one soft } \gamma \leftarrow e\mu\right) = e\bar{u}(p')\mathcal{M}_0(p',p)u(p)\left(\frac{p'\cdot\epsilon^\star}{p'\cdot k} - \frac{p\cdot\epsilon^\star}{p\cdot k}\right) \tag{6.32}$$

**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 [9]

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

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 (6.31)) arises classically as the number of soft photons in 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.

---

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

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:

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

---

[9] Check it:

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

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

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[10]

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

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. Something you may have been tempted to regard as an ugly detail has saved the day.

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 called 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

[10] Notice that we add the cross-sections, not the amplitudes for these processes with different final states (despite some salacious rumors to the contrary in last week's lecture). 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.

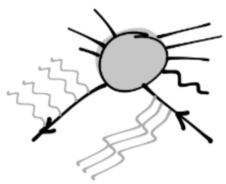
sense even if the electron and muon charges $Q_e, Q_m$ were different, so the calculation should make sense term-by-term in an expansion in $Q_m$.

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*.

---

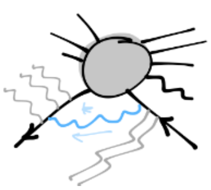
**Sketch of exponentiation of soft photons.** [Peskin §6.5] Consider a diagram with $n$ soft external photons, summed over ways of distributing them on an initial and final electron line:

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



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

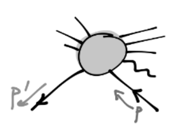
Here the difference in each factor is just as in (6.32), one term from initial and one from final-state emission; expanding the product gives the sum over $n_f = 1 - n_i$, the number coming from the final-state line. From this expression, we can make a diagram with a soft-photon *loop* by picking an 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 $k$:



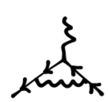
$$= \frac{e^2}{2} \int đ^4 k \frac{-\mathbf{i}\eta_{\rho\sigma}}{k^2} \left( \frac{p'}{p' \cdot k} - \frac{p}{p \cdot k} \right)^\rho \left( \frac{p'}{-p' \cdot k} - \frac{p}{-p \cdot k} \right)^\sigma$$

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

$$\bar{u}\mathbf{i}\mathcal{M}_0 u \cdot \mathbf{X} = \Bigg($$



$$\Bigg) \cdot \Bigg($$



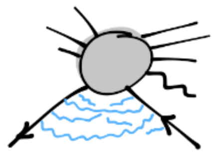
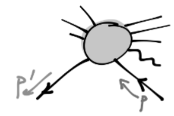
$$\Bigg)_{\text{soft part}}$$

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

$$\mathcal{M}_{\text{virtual soft}} = \sum_{m=0}^{\infty} \frac{1}{m!} \Bigg($$

$$\Bigg) = \Bigg($$

$$\underbrace{\qquad\qquad}_{i\mathcal{M}_0} \Bigg) \underbrace{\sum_m \frac{1}{m} \mathbf{X}^m}_{e^{\mathbf{X}}}.$$

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

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

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 $\alpha$:

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

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.

---

### 6.6.3 Some magic from gauge invariance of QED

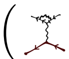
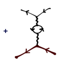
We found that the self-energy of the electron gave a wavefunction renormalization factor

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

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

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

Combining these together

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

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

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.

## 6.7 Vacuum polarization

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

$$\mathcal{L} = \bar{\psi}\left(\not{\partial} + \mathbf{i}e\tilde{\not{A}} - m\right)\psi - \frac{1}{4}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}.$$

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:

$$\mathcal{L} = \bar{\psi}\left(\not{\partial} + \mathbf{i}\not{A} - m\right)\psi - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}$$

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

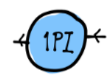
$$\langle A_\mu A_\nu\rangle \sim \frac{-\mathbf{i}\eta_{\mu\nu}e^2}{q^2}.$$

None of the physics is different, since each internal photon line still has two ends on a $\bar{\psi}\not{A}\psi$ 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 $\bar{\psi}\not{\partial}\psi$ and $\bar{\psi}\not{A}\psi$ terms[11]. Therefore, any counterterm we need for the $\bar{\psi}\not{\partial}\psi$ term (which comes from the electron self-energy correction and is called $\delta Z_2$) must be the *same* as the counterterm for the $\bar{\psi}\not{A}\psi$ 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 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 might seem plausible given what a mess the proton is, and would be a big deal for atoms. Gauge invariance forbids it.

[End of Lecture 27]

Just as we defined the electron self-energy (amputated 2-point function) as  $=$ $-\mathbf{i}\Sigma(\not{p})$ (with two spinor indices implied), we define the photon self-energy as

$+\mathbf{i}\Pi_{\mu\nu}(q^2) \equiv$  $=$  $+$  $+\,\mathcal{O}(e^4)$.

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

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

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

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

[11] Notice that the gauge transformation of the rescaled $A_\mu$ is $A_\mu \to A_\mu + q\partial_\mu\lambda(x), \psi(x) \to e^{\mathbf{i}q\lambda(x)}\psi(x)$ 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 $\tilde{A}_\mu \to \tilde{A}_\mu - q\partial_\mu\lambda(x)/e$.

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

$$\Delta^{\mu}_{T\rho}\Delta_{T\nu}^{\rho} = \Delta^{\mu}_{T\rho} \tag{6.33}$$

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

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

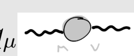
without changing any gauge-invariant physics. This is useful because then

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

$$\begin{aligned} &\overset{(6.33)}{=} \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right)\mathbf{i}\Pi q^2\Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \cdots\right) \\ &= \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \Pi\Delta_T + \Pi^2\Delta_T + \cdots\right) = \frac{-\mathbf{i}\Delta_T}{q^2}\frac{1}{1-\Pi(q^2)}. \end{aligned} \tag{6.34}$$

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 $\Pi(q^2) \sim \frac{A_0}{q^2}$ 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?

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

But here is an identity:

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

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[12], the cutoff dependence looks like[13]:

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

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 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

$$e_0^2 = e^2\left(1 + \frac{\alpha_0}{4\pi}\frac{2}{3}\ln\Lambda^2\right) + \mathcal{O}(\alpha^2). \tag{6.35}$$

$$m_0 = m + \mathcal{O}(\alpha_0) = m + \mathcal{O}(\alpha). \tag{6.36}$$

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.

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

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

[12]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.

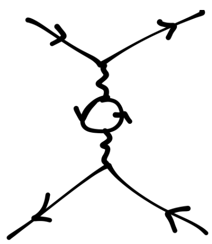
[13]The factor in front of the $\ln\Lambda$ can be made to look like it does in other textbooks using $\alpha = \frac{e^2}{4\pi}$, so that

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

$$= e^2 L_\mu \bar{u}(p') \left[ \gamma^\mu \left( 1 + \frac{\alpha}{2\pi}(A+B+D) \right) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \frac{\alpha}{2\pi} C \right] u(p) + \mathcal{O}(\alpha^2) \tag{6.37}$$

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 (6.35) and (6.36) 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

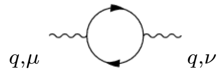


$$\propto e_0 \frac{1}{1-\Pi(0)} e_0 = e^2.$$

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.

### 6.7.1 Under the hood

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



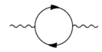

$$= -\int đ^D k \,\mathrm{tr} \left( (\mathbf{i}e\gamma^\mu) \frac{\mathbf{i}(\not{k}+m)}{k^2-m^2} (\mathbf{i}e\gamma^\nu) \frac{\mathbf{i}\left(\not{q}+\not{k}+m\right)}{(q+k)^2-m^2} \right)$$

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

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

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 (this is a distinct issue from the choice of RG scheme). With a euclidean momentum cutoff, the diagram  gives something of the form

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

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.[14]

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

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

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 boiling, the answer is

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

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

[14] Two points from lecture: How could we have predicted that the cutoff on euclidean momentum $\ell_E^2 < \Lambda^2$ would break gauge invariance? I haven't found a more direct argument than its violation of the Ward identity here; let me know if you find one. 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.

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 $\int đ^D p f(p+q) = \int đ^D p f(p)$ [15]
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 $D$:

$$\left(\frac{\pi}{a}\right)^{D/2} = \int d^D x e^{-a\vec{x}^2} = \Omega_{D-1} \int_0^\infty x^{D-1} dx e^{-ax^2} = \frac{1}{2} a^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right) \Omega_{D-1} \ . \tag{6.38}$$

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

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

$$\begin{aligned} \delta\Pi_2(p^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right)\right) \end{aligned} \tag{6.39}$$

where we have introduced the heralded $\mu$:

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

[15] Note that this rule fails for the euclidean momentum cutoff.

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.

[End of Lecture 28]

In the second line of (6.39), 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 demand that the photon propagator is not corrected at $q = 0$, *i.e.* demand $Z_\gamma = 1$. Then the resulting one-loop shift is

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

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

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

---

### 6.7.2 Physics from vacuum polarization

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

$$\tilde{V}(q) = \frac{1}{q^2}\frac{e^2}{1-\Pi(q^2)} \equiv \frac{e^2_{\text{eff}}(q)}{q^2}. \tag{6.40}$$

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

$$\delta\tilde{\Pi}_2(q) = \int dx x(1-x) \ln\left(1 - \frac{q^2}{m^2}x(1-x)\right) \overset{q\ll m}{\simeq} \int dx x(1-x)\left(-\frac{q^2}{m^2}x(1-x)\right) = -\frac{q^2}{30m^2}$$

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 $\Pi_2$ depends. 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

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

Therefore, the effective charge in (6.40) at high momentum exchange is

$$e^2_{\text{eff}}(q^2) = \frac{e^2}{1 - \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right)}.$$

(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. I'll say more about it after we discuss the Wilsonian perspective in §9.

# 7 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'.

## 7.1 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 [16]

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

[16]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:

$$\mathbf{i}\mathcal{D}(x) = \theta(t) \langle 0| e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}\mathcal{O}^\dagger(0) |0\rangle + \theta(-t) \langle 0| \mathcal{O}^\dagger(0)e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x} |0\rangle . \tag{7.1}$$

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 (7.1); sticking the $\mathbb{1}$ in a suitable place:

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

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.

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 $\mathcal{D}(x)$:

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

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

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

$$\left\langle\vec{k}\right| \mathcal{O}(0)^\dagger \left|0\right\rangle = \frac{Z^{\frac{1}{2}}}{\sqrt{(2\pi)^{D-1} 2\omega_{\vec{k}}}} \tag{7.2}$$

where $\omega_{\vec{k}}$ 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. [17] What is $Z$? From (7.2) 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

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

This is a summand in the whole horrible resolution:

$$1\!\!1 = 1\!\!1_1 + \cdots .$$

[17] 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 (7.2), 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 $\sum_n$ which comes from 1-particle states:

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

(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 earlier when we mentioned LSZ.)[18]

[End of Lecture 29]

---

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

$$\operatorname{Im} \frac{1}{Q - \mathbf{i}\epsilon} = \pi\delta(Q), \qquad \text{(for } Q \text{ real).} \tag{7.3}$$

we have

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

More explicitly:

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

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:

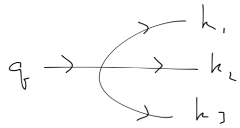
$$\operatorname{Im}\mathcal{D}(q) = ... + \pi Z \delta(q^2 - m^2).$$

[18]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.

---

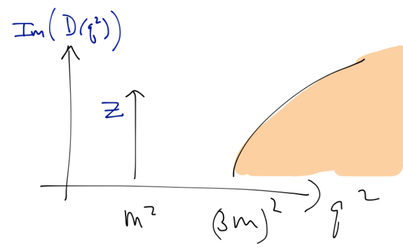
Now what about multiparticle states? The associated sum over such states involves mutliple (spatial) momentum integrals, not fixed by the total momentum *e.g.* in $\phi^4$ theory:




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

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

---

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:

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

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

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

The function $\rho(s)$ is called the spectral density for this Green's function. 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 must go like $\frac{1}{k^2}$ (or larger), since $\rho(s) \geq 0$ means that there cannot be cancellations between each $\frac{1}{k^2-\mu^2}$ 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

$$[\phi(x,t), \partial_t \phi(y,t)] = \mathbf{i}\delta^{(3)}(x-y).$$

This is a statement both about the normalization of the field, and that its canonical momentum is its time derivative. Then[19]

$$1 = \int_0^\infty ds \rho(s). \tag{7.4}$$

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

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

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.

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

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

For an interacting canonical field, we have instead a spectral representation:

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

where $\rho$ is the same spectral density as above. Now take $\partial_{x^0}|_{x^0=y^0}$ of the BHS and use $\partial_t \Delta_+(x-y; \mu^2)|_{x^0=y^0} = -\frac{\mathbf{i}}{2}\delta^3(\vec{x}-\vec{y})$.

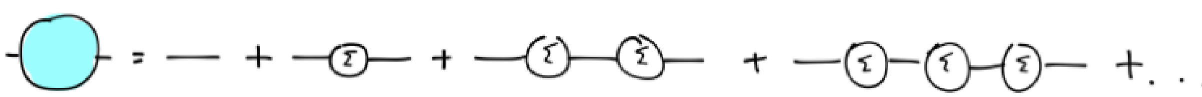
The idea of spectral representation and spectral density is more general than the Lorentz-invariant case. In particular, the spectral density of a Green's function is an important concept in cond-mat. For example, the spectral density for the electron 2-point function is the thing that actually gets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

## 7.2 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:

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

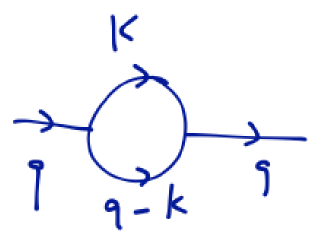
Sum the geometric series of 1PI insertions to get



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

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. 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 $m$ 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 – 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

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

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

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 $d\Phi = đk_1đk_2(2\pi)^D\delta^D(k_1 + k_2 - q)$.

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:

$$\begin{aligned} \mathbf{i}\Sigma(q) &= \int d^dx\; e^{\mathbf{i}qx}\left(\mathbf{i}g\right)^2 \mathbf{i}\mathcal{D}(x)\mathbf{i}\mathcal{D}(x) \\ &= g^2\int d\Phi \frac{1}{k_1^2 - m_1^2 + \mathbf{i}\epsilon}\frac{1}{k_2^2 - m_2^2 + \mathbf{i}\epsilon} \end{aligned} \tag{7.5}$$

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

$$0 = \int d^dx\; e^{\mathbf{i}qx}\left(\mathbf{i}g\right)^2 \mathbf{i}\mathcal{D}_{\text{adv}}(x)\mathbf{i}\mathcal{D}_{\text{ret}}(x)$$

$$= g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 - \sigma_1 \mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \sigma_2 \mathbf{i}\epsilon} \tag{7.6}$$

where $\sigma_{1,2} \equiv \text{sign}(k_{1,2}^0)$. 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.

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:

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

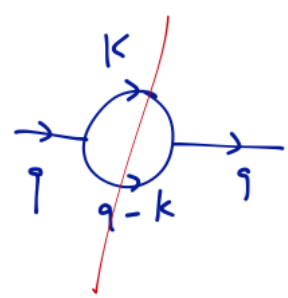
$$= \frac{g^2}{2} \int \frac{đ^{D-1}\vec{k}_1}{2\omega_{\vec{k}_1}} \frac{đ^{D-1}\vec{k}_2}{2\omega_{\vec{k}_2}} (2\pi)^D \delta^D(k_1 + k_2 - q) \; .$$

But this is exactly the density of actual final states into which the thing can decay! In summary:

$$\text{Im}\,\Sigma = \sum_{\substack{\text{actual states } n \text{ of 2 particles} \\ \text{into which } \phi \text{ can decay}}} \| \mathcal{A}_{\phi \to n} \|^2 \tag{7.7}$$

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; re-
place 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.

[End of Lecture 30]

The general form of (7.7) 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 $\sum_n \to \int đ\vec{k}_1\, đ\vec{k}_2$, the two-particle density of states.

A bit more explicitly, introducing a basis of scattering states

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

we have

$$\begin{aligned}\langle F| \mathcal{T}^\dagger \mathbb{1}\mathcal{T} |I\rangle &= \sum_n \langle F| \mathcal{T}^\dagger \sum_n \prod_{f=1}^n \int \frac{đ^3 q_f}{2E_f} |\{q_f\}\rangle \langle \{q_f\}| \mathcal{T} |I\rangle \\ &= \sum_n \prod_{f=1}^n \int \frac{đ^3 q_f}{2E_f} \not{\delta}^4(p_F - \sum_f q_f)\mathcal{M}^\star_{\{q_f\}F}\not{\delta}^4(p_I - \sum_f q_f)\mathcal{M}_{\{q_f\}I}\end{aligned}$$

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

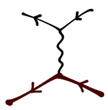
$$\int \frac{đ^3 q_f}{2E_f}\not{\delta}^4(p_F - \sum_f q_f) = \int d\Pi_n$$

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

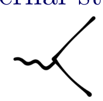
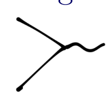
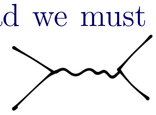
$$\mathbf{i}\left(\mathcal{M}^\star_{IF} - \mathcal{M}_{FI}\right) = \sum_n \int d\Pi_n \mathcal{M}^\star_{\{q_f\}F}\mathcal{M}_{\{q_f\}I}.$$

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

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

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 example (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 $\Big($  $\Big)$.

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

$$\mathcal{M} = -\left(\sqrt{Z}\right)^2 \Sigma = -Z\Sigma \tag{7.8}$$

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

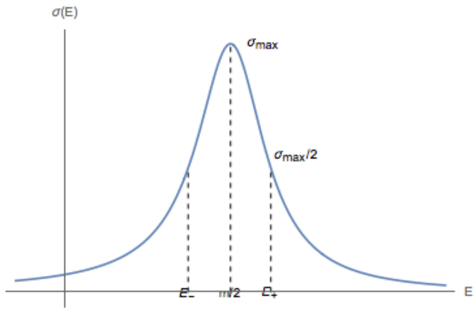
$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{p^2 - m_0^2 - \Sigma(p)} \simeq \frac{\mathbf{i}}{(p^2 - m^2)\underbrace{\left(1 - \partial_{p^2} A|_{m^2}\right)}_{=Z^{-1}} - \mathbf{i}B} = \frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}BZ}.$$

In terms of the *particle width* $\Gamma_w \equiv -ZB/m$, 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

$$\left|\tilde{G}^{(2)}(p)\right|^2 = \left|\frac{1}{(p^2 - m^2) - \mathbf{i} m\Gamma_w}\right|^2 = \frac{1}{(p^2 - m^2)^2 + m^2\Gamma_w^2}$$




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:

$$\Gamma_w = -\frac{BZ}{m} \overset{(7.8)}{=} -\frac{1}{m}\left(-\mathrm{Im}\,\mathcal{M}_{1\to 1}\right) \overset{\text{optical}}{=} \frac{1}{m}\frac{1}{2}\sum_n \int_f d\Pi_n |\mathcal{M}_{\{q_f\}1}|^2 = \Gamma$$

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 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 $E$:

$$\sigma(E \gg ...) \sim G^2 E^{-2-2k}. \tag{7.9}$$

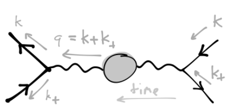
This means that if $k \leq -1$, the cross section grows at high energy. In such a case, something else must happen to 'restore unitarity'. One example is Fermi's theory of Weak interactions, which involves a 4-fermion coupling $G_F \sim M_W^{-2}$. Here we know what happens, namely the electroweak theory,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

## 7.3 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 $s = (k+k_+)^2 \gg m_e^2$:

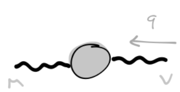
$$\sigma_T^{e^+e^-} \overset{\text{optical thm}}{=} \frac{1}{2s} \mathrm{Im}\, \mathcal{M}(e^+e^- \leftarrow e^+e^-) \tag{7.10}$$

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 (7.10):

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

(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 $M^2 \gg m_e^2$:

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

and

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

with $F(M^2/s) = \sqrt{1 - \frac{4M^2}{s}}\left(1 + \frac{2M^2}{s}\right) = 1 + \mathcal{O}\left(M^2/s\right)$. In perturbative QCD, the leading order result is the same from each quark with small enough mass:

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

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:

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

It looks like we are taking $x \to 0$ and 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. The limit where we can do (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.)

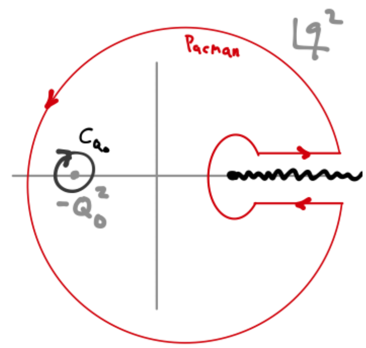
How can we use this knowledge to find the answer in the physical regime of $q^2 > 0$? The fact that $\Pi_h$ is a two-point function means that it has a spectral representation. It is analytic in the complex $q^2$ plane except for a branch cut on the positive real axis coming from production of real intermediate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

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

On the other hand, we know from the (appropriate generalization to currents of the) spectral representation sum rule (7.4) 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 was 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 31]

# 8 A parable on integrating out degrees of freedom

Here's a second parable from QM which gives some useful perspective on renormalization in QFT. It is also a valuable opportunity to understand the differences and connections between euclidean and real-time Green's functions. And it will give me a chance to remind you that you already know all about path integrals in 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 $q^2 Q$ 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 $q$:

$$e^{-S_{\text{eff}}[q]} := \int [dQ] e^{-S[q,Q]}$$
$$= e^{-S_{\omega_0}[q]} \left\langle e^{-S_{\text{int}}[Q,q]} \right\rangle_Q$$

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:

$$\left\langle e^{-S_{\text{int}}[Q,q]} \right\rangle_Q = \int [dQ] e^{-S_\Omega[Q] - \int ds J(s) Q(s)} = \mathcal{N} e^{\frac{1}{4}\int dsdt J(s) G(s,t) J(t)} \quad .$$

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 *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: $G(s) = \int đ\omega e^{-i\omega s} G_\omega$ makes it algebraic:

$$G_\omega = \frac{1}{\omega^2 + \Omega^2}$$

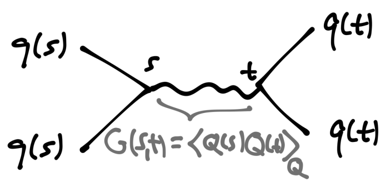
and we have

$$G(s) = \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\omega^2 + \Omega^2}. \tag{8.1}$$

So we have:

$$e^{-S_{\text{eff}}[q]} = e^{-S_{\omega_0}[q]} e^{-\int dtds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2}$$

or taking logs




$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dtds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2 \ . \tag{8.2}$$

$Q$ 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.

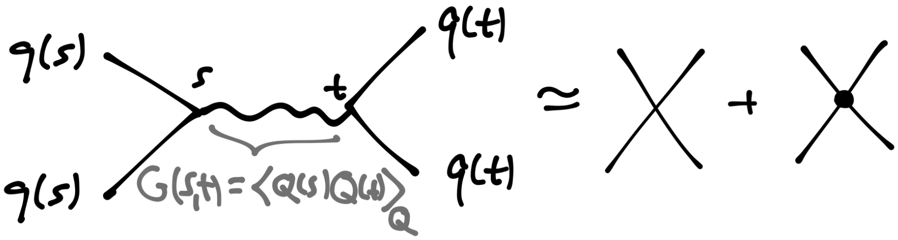
But now suppose we are interested in times much longer than $1/\Omega$, say times comparable to the period of oscillation of the less-stiff spring $2\pi/\omega$. We can accomplish this by Taylor expanding under the integrand in (8.1):

$$G(s) \overset{s \gg 1/\Omega}{\simeq} \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\Omega^2} \underbrace{\frac{1}{1 + \frac{\omega^2}{\Omega^2}}}_{=\sum_n (-1)^n \left(\frac{\omega^2}{\Omega^2}\right)^n} \simeq \frac{1}{\Omega^2}\delta(s) + \frac{1}{\Omega^4}\partial_s^2 \delta(s) + \ldots$$

Plug this back into (8.2):

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt \frac{g^2}{2\Omega^2} q(t)^4 + \int dt \frac{g^2}{2\Omega^4} \dot{q}^2 q^2 + \ldots$$

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 $\Omega$.




$$+\cdots \tag{8.3}$$

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: 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 (7.9). 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.

### 8.0.1 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 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. 8.3.

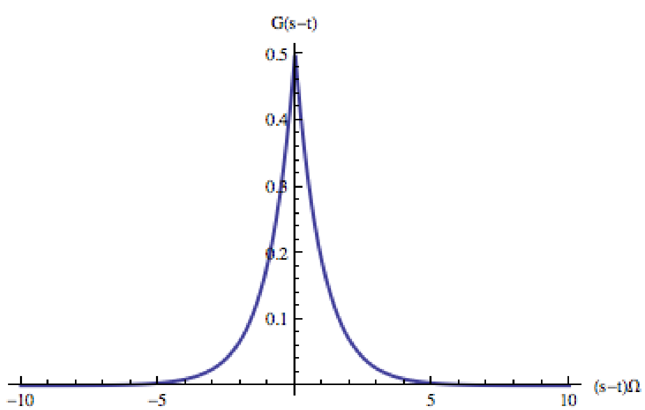
If we have the hubris to care about the exact answer, it's nonlocal in time. But if we want exact answers then we'll have to do the integral over $q$, too. On the other hand, the hierarchy of scales $\omega_0 \ll \Omega$ is useful if we ask questions about energies of order $\omega_0$, e.g.

$$\langle q(t)q(0)\rangle \text{ with } t \sim \frac{1}{\omega_0} \gg \Omega$$

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:




One more attempt to drive home the central message of this discussion: 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. 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:

$$G(s) = \int đ\omega \frac{e^{-\mathbf{i}\omega s}}{\omega^2 + \Omega^2} = e^{-\Omega|s|}\frac{1}{2\Omega}. \tag{8.4}$$

Do it by residues: the integrand has poles at $\omega = \pm\mathbf{i}\Omega$ (see the figure 1 below). 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 $Q$:

$$Z[J] \equiv \left\langle e^{\int dt Q(t)J(t)} \right\rangle_Q \quad , \quad \langle Q(t_1)Q(t_2)...\rangle_Q = \frac{\delta}{\delta J(t_1)}\frac{\delta}{\delta J(t_1)}... \log Z[J]$$

[20] 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. The same is not true of the real-time Green's function, which we discuss next in §8.0.2.

- **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.

  All the complications we'll encounter next (in §8.0.2) with choosing frequency contours are identical in QFT.

### 8.0.2 Wick rotation to real time.

For convenience, I have described this calculation in euclidean time (every $t$ or $s$ or $\tau$ that has appeared so far in this subsection has been a euclidean time). This is nice because the euclidean action is nice and positive, and all the wiggly and ugly configurations are manifestly highly suppressed in the path integral. Also, in real time[21] we have to make statements about states: *i.e.* in what state should we put the heavy mode?

---

[20] Functional derivatives are very useful! A reminder: the definition is

$$\frac{\delta J(s)}{\delta J(t)} = \delta(s-t) \tag{8.5}$$

plus the Liebniz properties (linearity, product rule).

[21] aka Minkowski time aka Lorentzian time

The answer is: in the groundstate – it costs more energy than we have to excite it. I claim that the real-time calculation which keeps the heavy mode in its groundstate is the analytic continuation of the one we did above, where we replace

$$\omega_{\text{Mink}} = e^{-\mathbf{i}(\pi/2-\epsilon)}\omega_{\text{above}} \tag{8.6}$$

where $\epsilon$ is (a familiar) infinitesimal. In the picture of the euclidean frequency plane in Fig. 1, this is a rotation by *nearly* 90 degrees. We don't want to go all the way to 90 degrees, because then we would hit the poles at $\pm\mathbf{i}\Omega$.

The replacement (8.6) just means that if we integrate over *real* $\omega_{\text{Mink}}$, we rotate the contour in the integral over $\omega$ as follows:

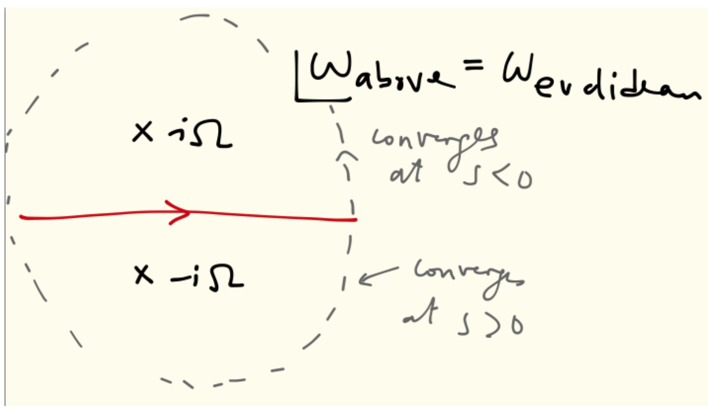
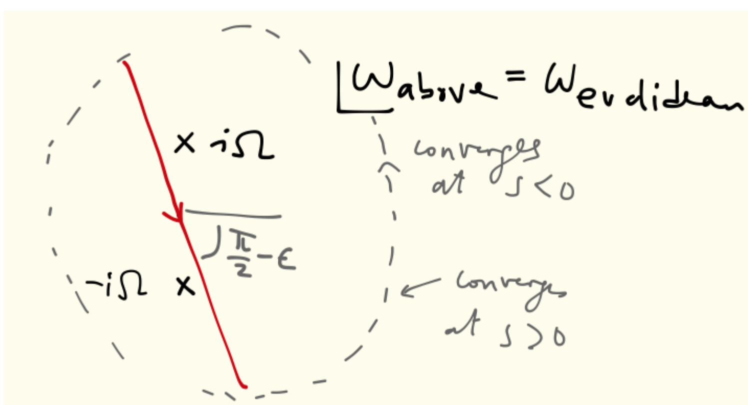



**Figure 1:** Poles of the integrand of the $\omega$ integral in (8.4).




as a result we pick up the same poles at $\omega_{\text{above}} = \pm\mathbf{i}\Omega$ as in the euclidean calculation. Notice that we had better also rotate the argument of the function, $s$, at the same time to maintain convergence, that is:

$$\omega_{\text{eucl}} = -\mathbf{i}\omega_{\text{Mink}}, \quad \omega_{\text{eucl}}t_{\text{eucl}} = \omega_{\text{Mink}}t_{\text{Mink}}, \quad t_{\text{eucl}} = +\mathbf{i}t_{\text{Mink}}. \tag{8.7}$$

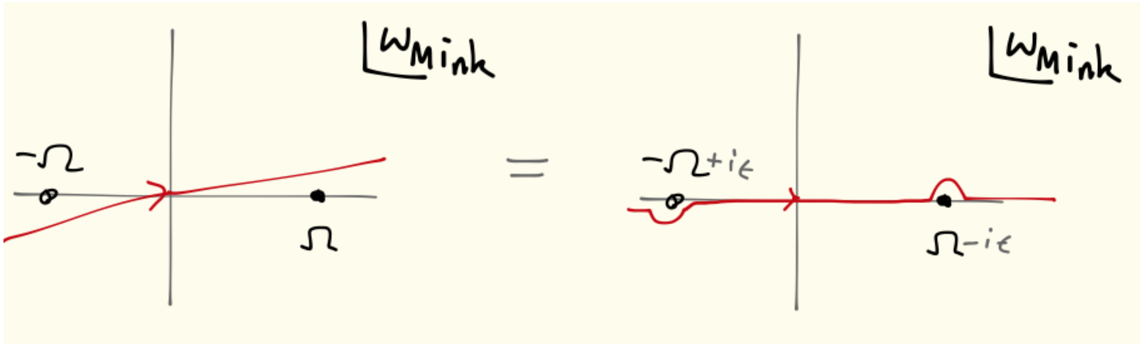



**Figure 2:** The Feynman contour in the $\omega_{\text{Mink}}$ complex plane.

So this is giving us a contour prescription for the real-frequency integral. The result is the *Feynman* propagator, which we've been using all along: depending on the sign of the (real) time separation of the two operators (recall that $t$ is the difference), we close the contour around one pole or the other, giving the *time-ordered* propagator. (It is the same as shifting the heavy frequency by $\Omega \to \Omega - \mathbf{i}\epsilon$, as indicated in the right part of Fig. 2.)

---

Notice for future reference that the euclidean action and real-time action are related by

$$S_{\text{eucl}}[Q] = \int dt_{\text{eucl}} \frac{1}{2}\left(\left(\frac{\partial Q}{\partial t_{\text{eucl}}}\right)^2 + \Omega^2 Q^2\right) = -\mathbf{i}S_{\text{Mink}}[Q] = -\mathbf{i}\int dt_{\text{Mink}} \frac{1}{2}\left(\left(\frac{\partial Q}{\partial t_{\text{Mink}}}\right)^2 - \Omega^2 Q^2\right).$$

because of (8.7). This means the path integrand is $e^{-S_{\text{eucl}}} = e^{\mathbf{i}S_{\text{Mink}}}$.

---

Why does the contour coming from the euclidean path integral put the excited mode into its groundstate? That's the the point in life of the euclidean path integral, to prepare the groundstate from an arbitrary state:

$$\int_{Q_0} [dQ] e^{-S[Q]} = \langle Q_0| e^{-\mathbf{H}T} |...\rangle = \psi_{\text{gs}}(Q_0) \tag{8.8}$$

– the euclidean-time propagator $e^{-\mathbf{H}T}$ beats down the amplitude of any excited state relative to the groundstate, for large enough $T$.

Let me back up one more step and explain (8.8) more. You know a path integral representation for the real-time propagator

$$\langle f| e^{-\mathbf{i}\mathbf{H}t} |i\rangle = \int_{f\leftarrow i} [dq] e^{\mathbf{i}\int_0^t dtL}.$$

On the RHS here, we sum over all paths between $i$ and $f$ in time $t$ (*i.e.* $q(0) = q_i, q(t) = q_f$), weighted by a phase $e^{\mathbf{i}\int dt L}$.

But that means you also know a representation for

$$\sum_f \langle f| e^{-\beta \mathbf{H}} |f\rangle \equiv \mathrm{tr} e^{-\beta \mathbf{H}}$$

– namely, you sum over all *periodic* paths $q_i = q_f$ in *imaginary time* $t = -\mathbf{i}\beta$. So:

$$Z(\beta) = \mathrm{tr} e^{-\beta \mathbf{H}} = \oint [dq] e^{-\int_0^\beta d\tau L} = \oint [dq] e^{-S_{\mathrm{eucl}}[q]}$$

The LHS is the partition function in quantum statistical mechanics. The RHS is the euclidean functional integral we've been using. [For more on this, see Zee §V.2]

The period of imaginary time, $\beta \equiv 1/T$, is the inverse temperature. More accurately, we've been studying the limit as $\beta \to \infty$. Taking $\beta \to \infty$ means $T \to 0$, and you'll agree that at $T = 0$ we project onto the groundstate (if there's more than one groundstate we have to think more).

**Time-ordering.** To summarize the previous discussion: in real time, we must choose a state, and this means that there are many Green's functions, not just one: $\langle \psi| Q(t)Q(s) |\psi\rangle$ depends on $|\psi\rangle$, unsurprisingly.

But we found a special one which arises by analytic continuation from the euclidean Green's function, which is unique[22]. It is

$$G(s,t) = \langle \mathcal{T} (Q(s)Q(t))\rangle_Q \ ,$$

the time-ordered, or Feynman, Green's function, and I write the time-ordering symbol $\mathcal{T}$ to emphasize this. I emphasize that from our starting point above, the time ordering arose because we have to close the contour in the UHP (LHP) for $t < 0$ ($t > 0$).

Let's pursue this one more step. The same argument tells us that the generating functional for real-time correlation functions of $Q$ is

$$Z[J] = \left\langle \mathcal{T} e^{\mathbf{i}\int JQ} \right\rangle = \langle 0| \mathcal{T} e^{\mathbf{i}\int JQ} |0\rangle \,.$$

[22] Another important perspective on the uniqueness of the euclidean Green's function and the non-uniqueness in real time: in euclidean time, we are inverting an operator of the form $-\partial_\tau^2 + \Omega^2$ which is *positive* ($\equiv$ all its eigenvalues are positive) – recall that $-\partial_\tau^2 = \hat{p}^2$ is the square of a hermitian operator. If all the eigenvalues are positive, the operator has no kernel, so it is completely and unambiguously invertible. This is why there are no poles on the axis of the (euclidean) $\omega$ integral in (8.4). In real time, in contrast, we are inverting something like $+\partial_t^2 + \Omega^2$ which annihilates modes with $\partial_t = \mathbf{i}\Omega$ (if we were doing QFT in $d > 0+1$ this equation would be the familiar $p^2 - m^2 = 0$) – on-shell states. So the operator we are trying to invert has a kernel and this is the source of the ambiguity. In frequency space, this is reflected in the presence of poles of the integrand *on* the contour of integration; the choice of how to negotiate them encodes the choice of Green's function.

In the last step I just emphasized that the real time expectation value here is really a *vacuum* expectation value. This quantity has the picturesque interpretation as the *vacuum persistence amplitude*, in the presence of the source $J$.

**Causality.** In other treatments of this subject (such as ours of last quarter), you will see the Feynman contour motivated by ideas about causality. This was not the logic of our discussion but it is reassuring that we end up in the same place. Note that even in 0+1 dimensions there is a useful notion of causality: effects should come after their causes. I will have more to say about this later, when we have reason to discuss other real-time Green's functions. [End of Lecture 32]

# 9 The Wilsonian perspective on renormalization

[Fradkin, 2d edition, chapter 4; Cardy; Zee §VI; Álvarez-Gaumé and Vázquez-Mozo, *An Invitation to QFT*, chapter 8.4-5 ($\simeq$ §7.3-4 of hep-th/0510040)] The following discussion describes a perspective which can be applied to any system of (many) extensive degrees of freedom. This includes many statistical-mechanics systems, condensed-matter systems and also QFTs in high energy physics. The great insight of Kadanoff and Wilson about such systems is that we should organize our thinking about them by length scale. We should think about a *family* of descriptions, labelled by the resolution of our microscope. Before explaining this perspective in detail, let's spend some time addressing the following basic and instructive question:

## 9.1 Where do field theories come from?

### 9.1.1 A model with finitely many degrees of freedom per unit volume

Consider the following system of extensive degrees of freedom – it is an example of a very well-regulated (euclidean) QFT. At each site $i$ of a square lattice we place a two-valued (classical) degree of freedom $s_i = \pm 1$, so that the path 'integral' measure is

$$\int [ds] ... \equiv \sum_{\{s_i\}} ... = \prod_{\text{sites, } i} \sum_{s_i = \pm 1} ... \,.$$

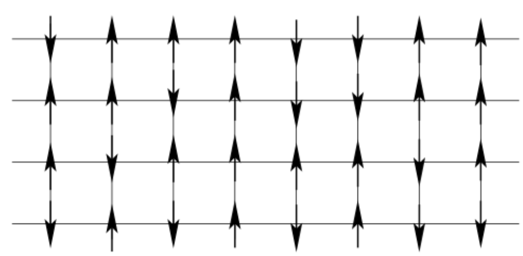


**Figure 3:** A configuration of classical Ising spins on the 2d square lattice. [from Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040]

Let's choose the euclidean action to be

$$S[s] = -\beta J \sum_{\langle i,j \rangle} s_i s_j \ .$$

Here $\beta J$ is some coupling; the notation $\langle i, j \rangle$ means 'sites $i$ and $j$ which are nearest neighbors'. The partition function is

$$Z = \int [ds] e^{-S[s]} = \sum_{\{s_i\}} e^{+\beta J \sum_{\langle i,j \rangle} s_i s_j}. \tag{9.1}$$

(I can't hide the fact that this is the thermal partition function $Z = \text{tr} e^{-\beta H}$ for the *classical Ising model on the square lattice*, with $H = -J \sum_{\langle i,j \rangle} s_i s_j$, and $\beta \equiv 1/T$ is the *coolness*[23], *i.e.* the inverse temperature.)

In the thermodynamic limit (the number of sites goes to infinity), this model has a special value of $\beta J > 0$ above which there is spontaneous breaking of the $\mathbb{Z}_2$ symmetry $s_i \to -s_i$ by a nonzero *magnetization*, $\langle s_i \rangle \neq 0$.

---

**Kramers-Wannier duality.** To see that there is a special value of $\beta J$, we can make the following observation, due to Kramers and Wannier, and generalized by Wegner, which is now a subject of obsession for many theoretical physicists. It is called *duality*. Consider a configuration of the spins. The action $S[s]$ is determined by the number of links across which the spins disagree (positive $\beta J$ favors contributions from spins which agree). It is possible to rewrite the partition sum in terms of these disagreements. (For more on this, see the lecture notes here.) The answer is identical to the original model, except with $\beta J$ replaced by $a(\beta J)^{-1}$ for some number $a$! At high temperature the model is obviously disordered, at low temperature the dual model is obviously disordered, but that means that the original model is ordered. In between something happens. If only *one* something happens, it must happen at the special value $\beta J = a(\beta J)^{-1}$.

For a more complete discussion of this subject of *duality* I recommend this review by Kogut, §4. We might have the opportunity to come back to it later in this course.

**Onsager solution.** Lars Onsager solved the model above exactly (published in 1944) and showed for sure that it has a critical point $(\beta J)_\star = \frac{1}{2} \tanh^{-1}\left(\frac{1}{\sqrt{2}}\right)$. For our present purposes this landmark result is a distraction.

---

**Comment on analyticity in $\beta J$ *versus* the critical point.** [Zee §V.3] The Ising model defined by (9.1) is a model of a magnet (more specifically, when $\beta J > 0$ which

[23]This nomenclature, due to the condensed matter physicist Miles Stoudenmire, does a great job of reminding us that at lower temperatures, quantum mechanics has more dramatic consequences.

makes neighboring spins want to align, a *ferro*magnet). Some basic phenomenology: just below the Curie temperature $T_c$, the magnetization (average magnetic moment per unit volume) behaves like

$$|M| \sim (T_c - T)^\beta$$

where $\beta$ is a pure number (it depends on the number of spatial dimensions)[24]. In terms of the Ising model, the magnetization is[25]

$$\langle M \rangle = \frac{1}{Z} \sum_{\{s_i\}} e^{-H(s)/T} \frac{\sum_i s_i}{\mathcal{V}} \quad . \tag{9.2}$$

($\mathcal{V}$ is the number of sites of the lattice, the volume of space.) How can you get such a *non-analytic* (at $T = T_c \neq 0$) function of $T$ by adding a bunch of terms of the form $e^{-E/T}$? It is clearly impossible if there is only a finite number of terms in the sum, each of which is analytic near $T_c \neq 0$. It *is* actually possible if the number of terms is infinite – finite-temperature phase transitions only happen in the thermodynamic limit.

### 9.1.2 Landau and Ginzburg guess the answer.

Starting from $Z$, even with clever tricks like Kramers-Wannier duality, and even for Onsager, it is pretty hard to figure out what the answer is for the magnetization. But the answer is actually largely determined on general grounds, as follows.

Let's ask what is the free energy $G$ *at fixed magnetization*, $G[M]$. How would we do this in an experiment? We'd apply a uniform magnetic field, and find just the right field to get the desired $M$, and then measure the free energy (with our trusty free-energy-ometer, of course). In more formal terms, we should add a source for the magnetization and compute

$$e^{-\beta F[J]} = \mathrm{tr} e^{-\beta(H + \sum MJ)}.$$

Pick some magnetization $M_c$, and *choose* $J^{[M_c]}$ so that

$$\langle M \rangle = -\frac{\partial F}{\partial J} = M_c.$$

Then $G[M_c] \equiv F[J^{[M_c]}] - \sum M_c J^{[M_c]}$. This is a Legendre transform of the usual $F$ in $Z = e^{-\beta F}$. In this context, the source $J$ is (minus) an external magnetic (Zeeman) field. This $G[M]$ is just the same idea as an object we'll introduce below called the

[24]The name is conventional; don't confuse it with the inverse temperature.

[25]In many real magnets, the magnetization can point in any direction in three-space – it's a vector $\vec{M}$. We are simplifying our lives.

euclidean effective action $\Gamma[\phi_c]$ (up to factors of $\beta$), where the analog of $M$, is called the 'classical field' $\phi_c$. $G$ is the thing we should minimize to find the magnetization in the groundstate.

**LG Effective Potential.** We can even consider a model where the magnetization is a vector. If $\vec{M}$ is independent of position $\vec{x}$ [26] then rotation invariance (or even just $M \to -M$ symmetry) demands that

$$G = V\left(a\vec{M}^2 + b\left(\vec{M}^2\right)^2 + ...\right)$$

where $a, b$[27] are some functions of $T$ that we don't know, and the dots are terms with more $M$s. These functions $a(T)$ and $b(T)$ have no reason not to be smooth functions of $T$. Now suppose there is a value of $T$ for which $a(T)$ vanishes:

$$a(T) = a_1(T - T_c) + ...$$

with $a_1 > 0$ a pure constant. For $T > T_c$, the minimum of $G$ is at $\vec{M} = 0$; for $T < T_c$, the unmagnetized state becomes unstable and new minima emerge at $|\vec{M}| = \sqrt{-\frac{a}{2b}} \sim (T_c - T)^{\frac{1}{2}}$. This is the *mean field theory* description of a second-order phase transition. It's not the right value of $\beta$ (it's about 1/3) for the 3d Curie point, but it shows very simply how to get an answer that is not analytic at $T_c$.

**LG Effective Action.** Landau and Ginzburg can do even better. $G(M)$ with constant $M$ is like the effective *potential*; if we let $M(\vec{x})$ vary in space, we can ask and answer what is the effective *action*, $G[M(\vec{x})]$. The Landau-Ginzburg effective action is

$$G[M] = \int d^d\vec{x}\left(a\vec{M}^2 + b\left(\vec{M}^2\right)^2 + c\partial_i\vec{M}\cdot\partial_i\vec{M} + ...\right) \tag{9.3}$$

– now we are allowed to have gradients. $c$ is a new unknown function of $T$; let's set it to 1 by rescaling $M$. This just a scalar field theory (with several scalars) in euclidean space. Each field has a mass $\sqrt{a}$ (they are all the same as a consequence of the spin rotation symmetry). So $\frac{1}{\sqrt{a}}$ is a length scale, to which we turn next.

**Definition of correlation length.** Suppose we perturb the system by turning on an external (we pick it) magnetic field (source for $\vec{M}$) $\vec{H}$, which adds to the hamiltonian by $-\vec{H}\cdot\vec{M}$. (So far we are doing Euclidean physics, which means equilibrium, no real time dependence.) Pick the field to be small, so its effect is small and we can study the linearized equations (let's do it for $T > T_c$, so we're expanding around $M = 0$):

$$\left(-\partial^2 + a\right)\vec{M} = \vec{H}\ .$$

[26] In (9.2), I've averaged over all space; instead we could have averaged over just a big enough patch to make it look smooth. We'll ask 'how big is big enough?' next – the answer is 'the correlation length'.

[27] Don't confuse $a$ with the lattice spacing; sorry, ran out of letters.

Recall the Green's function $G_2$ of a massive scalar field: $G_2$ solves this equation in the case where $H$ is a delta function. Since the equation is linear, that solution determines the solution for general $H$ (this was why Green introduced Green's functions):

$$M(x) = \int d^3y G_2(x,y)H(y) = \int d^3y \left( \int đ^3k \frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}-\vec{y})}}{\vec{k}^2 + a} \right) H(y)$$
$$= \int d^3y \frac{1}{4\pi|\vec{x}-\vec{y}|} e^{-\sqrt{a}|\vec{x}-\vec{y}|} H(y). \tag{9.4}$$

The Green's function

$$G_2^{IJ}(x) = \left\langle \vec{M}^I(x)\vec{M}^J(0) \right\rangle = \delta^{IJ} \frac{1}{4\pi|\vec{x}|} e^{-\sqrt{a}|\vec{x}|}$$

is diagonal in the vector index $I, J$ so I've suppressed it in (9.4). $G_2$ is the answer to the question: if I perturb the magnetization at the origin, how does it respond at $x$? The answer is that it dies off like

$$\left\langle \vec{M}(x)\vec{M}(0) \right\rangle \sim e^{-|x|/\xi}$$

– this relation defines the *correlation length* $\xi$, which will depend on the parameters. In the LG mean field theory, we find $\xi = \frac{1}{\sqrt{a}}$. The LG theory predicts the behavior of $\xi$ as we approach the phase transition to be $\xi \sim \frac{1}{(T-T_c)^\nu}$ with $\nu = \frac{1}{2}$. Again the exponent is wrong in detail (we'll see why below), but it's a great start.

Now let's return to the microscopic model (9.1). Away from the special value of $\beta J$, the correlation functions behave as

$$\langle s_i s_j \rangle_{\text{connected}} \sim e^{-\frac{r_{ij}}{\xi}}$$

where $r_{ij} \equiv$ distance between sites $i$ and $j$. Notice that the subscript *connected* means that we need not specify whether we are above or below $T_c$, since it subtracts out the disconnected bit $\langle s_i \rangle \langle s_j \rangle$ by which their form differs. From the more microscopic viewpoint, $\xi$ is the length scale over which the values of the spins are highly correlated. This allows us to answer the question of how much coarse-graining we need to do to reach a continuum approximation: The continuum description in terms of

$$M(x) \equiv \frac{\sum_{i \in R_x} \langle s_i \rangle}{\text{Vol}(R_x)} \tag{9.5}$$

is valid if we average over regions $R$ (centered around the point $x$) with linear size bigger than $\xi$.

### 9.1.3 Coarse-graining by block spins.

We want to understand the connection between the microscopic spin model and the macroscopic description of the magnetization better, for example to systematically improve upon the quantitative failures of the LG mean field theory for the critical exponents. Kadanoff's idea is to consider a sequence of *blocking transformations*, whereby we group more and more spins together, to interpolate between the spin at a single site $s_i$, and the magnetization averaged over the whole system, passing through (9.5) on the way.

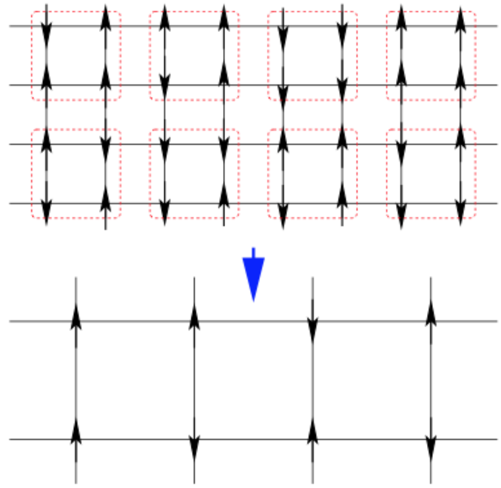


**Figure 4:** A blocking transformation. [from Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040]

The blocking (or 'decimation') transformation can be implemented in more detail for ising spins on the 2d square lattice as follows (Fig. 4). Group the spins into blocks of four as shown; we will construct a new coarser Ising system, where the sites of the new lattice correspond to the blocks of the original one, and the spin at the new site is an average of the four. One way to do this is majority rule:

$$s_{\text{block},\, b} \equiv \text{sign}\left(\sum_{i\in\text{block},b} s_i\right)$$

where we break a tie by defining $\text{sign}(0) = +1$.

We want to write our original partition function in terms of the averaged spins on a lattice with twice the lattice spacing. We'll use the identity

$$1 = \sum_{s_\text{block}} \delta\left(s_\text{block} - \text{sign}(\sum_{i\in\text{block}} s_i)\right) .$$

This is true for each block; we can insert one of these for each block. Split the original sum into nested sums, the outer one over the blocks, and the inner one over the spins within the block:

$$Z = \sum_{\{s\}} e^{-\beta H[s_i]} = \sum_{\{s_{\text{block},\, b}\}} \sum_{s\in\text{block},b} \prod_{\text{blocks}} \delta\left(s_{\text{block},b} - \text{sign}\left(\sum_{i\in\text{block},b} s_i\right)\right) e^{-\beta H^{(a)}[s]} .$$

The superscript $(a)$ on the Hamiltonian is intended to indicate that the lattice spacing is $a$. Now we interpret the inner sum as another example of integrating out stuff we don't care about to generate an effective interaction between the stuff we do care about:

$$\sum_{s\in\text{block},b} \prod_{\text{blocks}} \delta\left(s^{(2a)} - \text{sign}\left(\sum_{i\in\text{block},b} s_i\right)\right) e^{-\beta H^a[s]} \equiv e^{-\beta H^{(2a)}[s^{(2a)}]}$$

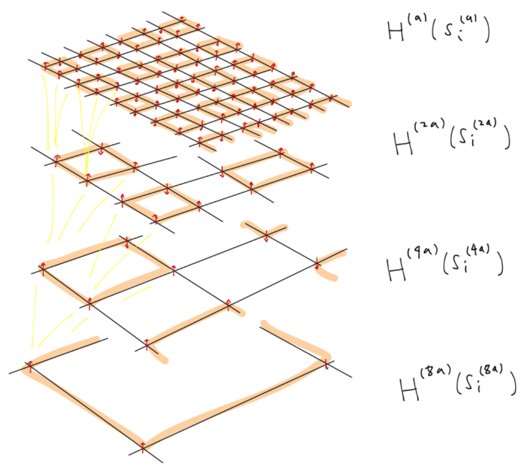
These sums are hard to actually do, except in 1d. But we don't need to do them to understand the form of the result.

As in our QM example from the previous lecture, the new Hamiltonian will be less local than the original one – it won't just be nearest neighbors in general:

$$H^{(2a)}[s^{(2a)}] = -J^{(2a)} \sum_{\langle i,j \rangle} s_i^{(2a)} s_j^{(2a)} + -K^{(2a)} \sum_{\langle\langle i,j \rangle\rangle} s_i^{(2a)} s_j^{(2a)} + ...$$

where $\langle\langle i,j \rangle\rangle$ means next-neighbors. Notice that I've used the same labels $i, j$ for the coarser lattice. We have rewritten the partition function as the same *kind* of model, on a coarser lattice, with different values of the couplings:

$$Z = \sum_{\{s^{(2a)}\}} e^{-\beta H^{(2a)}[s^{(2a)}]}$$




Now we can do it again. The decimation operation defines *a map on the space of (in this case Ising) Hamiltonians*:

$$H^{(a)} \mapsto H^{(2a)} \mapsto H^{(4a)} \mapsto H^{(8a)} \mapsto ...$$

The couplings $J, K...$ are coordinates on the space of Hamiltonians. Each time we do it, we double the lattice spacing; the correlation length in units of the lattice spacing gets halved, $\xi \mapsto \xi/2$. This operation is called a 'renormalization group transformation' but notice that it is very much not invertible; we lose information about the short-distance stuff by integrating it out.

[End of Lecture 33]

**RG fixed points.** Where can it end? One thing that can happen is that the form of the Hamiltonian can stop changing:

$$H^{(a)} \mapsto H^{(2a)} \mapsto H^{(4a)} \mapsto H^{(8a)} \mapsto ... \mapsto H_\star \mapsto H_\star \mapsto H_\star \; ...$$

The fixed point hamiltionian $H_\star$, which is not changed by the rescaling operation, is scale invariant. What can its correlation length be if it is invariant under $\xi \to \xi/2$? Either $\xi = 0$ (the mass of the fields go to infinity and there is nothing left to integrate) or $\xi = \infty$ (the mass goes to zero and we have more to discuss, we can call this a nontrivial fixed point).

Near a nontrivial fixed point, once $\xi \gg a$, the original lattice spacing, we are quite justified in using a continuum description, to which we return in subsection 9.2.

**Perturbations of a fixed point.** Before doing any more work, though, we can examine the possible behaviors of the RG flow near a fixed point. Consider a fixed point Hamiltonian $H_\star$, and move away from it slightly by changing one of the couplings a little bit:

$$H = H_\star + \delta g \mathcal{O}.$$

What does the RG do to this to leading order in $\delta g$? The possibilities are:

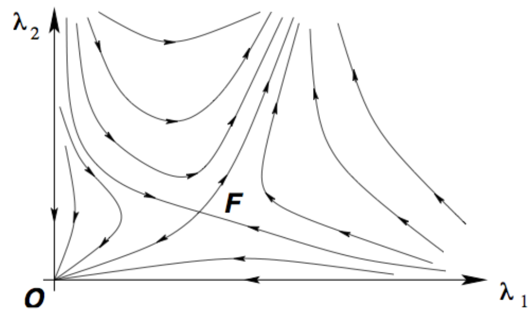



**Figure 5:** A possible set of RG flows for a system with two couplings $\lambda_{1,2}$. [from Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040]

- If the flow takes it back to the original fixed point, $\mathcal{O}$ (and its associated coupling $\delta g$) is called *irrelevant*.
- If the flow takes it away from the original fixed point, $\mathcal{O}$ is called a *relevant* perturbation of $H_\star$.
- The new $H$ might also be a fixed point, at least to this order in $\delta g$. Such a coupling (and the associated operator $\mathcal{O}$) is called *marginal*. If the new $H$ really is a new fixed point, not just to leading order in $\delta g$, then $\mathcal{O}$ is called *exactly marginal*. Usually it goes one way or the other and is called *marginally relevant* or *marginally irrelevant*.

Note the infrared-centric terminology.

---

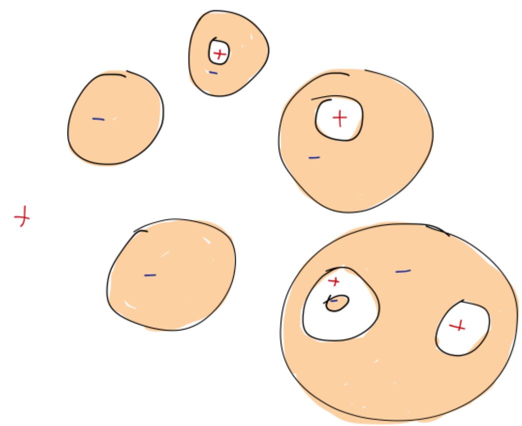
**Comment on *Universality*:** The Ising model is a model of many microscopically-different-looking systems. It can be a model of spins like we imagined above. Or it could be a model of a lattice gas – we say spin up at site $i$ indicates the presence of a gas molecule there, and spin down represents its absence. These different models will naturally have different microscopic interactions. But there will only be so many fixed points of the flow in the space of Hamiltonians on this system of 2-valued variables. This idea of the paucity of fixed points underlies Kadanoff and Wilson's explanation of the experimental phenomenon of *universality*: the same critical exponents arise from very different-seeming systems (*e.g.* the Curie point of a magnet and the liquid-gas critical point).

---

## 9.2 The continuum version of blocking

[Zee, §VI.8 (page 362 of 2d Ed.)]

Here is a very different starting point from which to approach the same critical point as in the previous subsection:



Consider the $\phi^4$ theory in Euclidean space, with negative $m^2$ (and no $\phi^k$ terms with odd $k$). This potential has two minima and a $\mathbb{Z}_2$ symmetry that interchanges them, $\phi \to -\phi$. If we squint at a configuration of $\phi$, we can label regions of space by the sign of $\phi$ (as in the figure at right). The kinetic term for $\phi$ will make nearby regions want to agree, just like the $J\sum_{\langle ij\rangle}\sigma_i\sigma_j$ term in the Ising model. So the critical point described by taking $m^2$ near zero is plausibly the same as the one obtained from the lattice Ising model described above[28].

We will study the integral

$$Z_\Lambda \equiv \int_\Lambda [D\phi] e^{-\int d^Dx \mathcal{L}(\phi)}. \tag{9.6}$$

Here the specification $\int_\Lambda$ says that we integrate over field configurations $\phi(x) = \int đ^D k e^{\mathbf{i}kx}\phi_k$ such that $\phi_k = 0$ for $|k| \equiv \sqrt{\sum_{i=1}^D k_i^2} > \Lambda$. Think of $2\pi/\Lambda$ as the lattice spacing[29] – there just aren't modes of shorter wavelength. We are using (again) a cutoff on the euclidean momenta $k_E^2 \le \Lambda^2$.

We want to understand (9.6) by some coarse-graining procedure. Let us imitate the block spin procedure. Field variations *within* blocks of space of linear size $na$ have wavenumbers greater than $\frac{2\pi}{na}$. (These modes average to zero on larger blocks; modes with larger wavenumber encode the variation *between* these blocks.) So the analog of the partition function after a single blocking step is the following: Break up the configurations into pieces:

$$\phi(x) = \int đk e^{\mathbf{i}kx}\phi_k \equiv \phi^< + \phi^> .$$

[28] For a more sophisticated argument for this equivalence, see page 7-9 of Polyakov, *Gauge Fields and Strings*.

[29] This cutoff is not precisely the same as have a lattice; with a lattice, the momentum space is periodic: $e^{\mathbf{i}kx_n} = e^{\mathbf{i}k(na)} = e^{\mathbf{i}\left(k+\frac{2\pi}{a}\right)(na)}$ for $n \in \mathbb{Z}$. Morally it is the same.

Here $\phi^<$ has nonzero fourier components only for $|k| \leq \Lambda - \delta\Lambda$ and $\phi^>$ has nonzero fourier components only for $\Lambda - \delta\Lambda \leq |k| \leq \Lambda$. Zee calls the two parts 'smooth' and 'wiggly'. They could also be called 'slow' and 'fast' or 'light' and 'heavy'. We want to do the integral over the heavy/wiggly/fast modes to develop an effective action for the light/smooth/slow modes:

$$Z_\Lambda = \int_{\Lambda-\delta\Lambda} [D\phi^<] e^{-\int d^D x \mathcal{L}(\phi^<)} \int [D\phi^>] e^{-\int d^D x \mathcal{L}_1(\phi^<, \phi^>)}$$

where $\mathcal{L}_1$ contains all the dependence on $\phi^>$ (and no other terms).

Just as with the spin sums, these integrals are hard to actually do, except in a gaussian theory. But again we don't need to do them to understand the form of the result. First give it a name:

$$e^{-\int d^D x \delta L(\phi^<)} \equiv \int [D\phi^>] e^{-\int d^D x \mathcal{L}_1(\phi^<, \phi^>)} \tag{9.7}$$

so once we've done the integral we'll find

$$Z_\Lambda = \int_{\Lambda-\delta\Lambda} [D\phi^<] e^{-\int d^D x (\mathcal{L}(\phi^<) + \delta\mathcal{L}(\phi^<))} . \tag{9.8}$$

To get a feeling for the form of $\delta\mathcal{L}$ (and because there is little reason not to) consider the more general Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 + \sum_n g_n \phi^n + ... \tag{9.9}$$

where we include *all* possible terms consistent with the symmetries (rotation invariance, maybe $\phi \to -\phi$...). Then we can find an explicit expression for $\mathcal{L}_1$:

$$\int d^D x \mathcal{L}_1(\phi^<, \phi^>) = \int d^D x \left( \frac{1}{2}(\partial\phi^>)^2 + \frac{1}{2} m^2 \left(\phi^>\right)^2 + ... \right)$$

(I write the integral so that I can ignore terms that integrate to zero such as $\partial\phi^< \partial\phi^>$.) This is the action for a scalar field $\phi^>$ interacting with itself and with a (slowly-varying) background field $\phi^<$. But what can the result $\delta\mathcal{L}$ be but something of the form (9.9) again, with different coefficients? The result is to shift the couplings $g_n \to g_n + \delta g_n$. (This includes the coefficient of the kinetic term and also of the higher-derivative terms which are hidden in the ... in (9.9). You will see in a moment the logic behind which terms I hid.)

Finally, so that we can compare steps of the procedure to each other, we rescale our rulers. We'd like to change units so that $\int_{\Lambda-\delta\Lambda}$ is a $\int_\Lambda$ with different couplings; we accomplish this by defining

$$\Lambda - \delta\Lambda \equiv b\Lambda, \qquad b < 1.$$

In $\int_{\Lambda-\delta\Lambda}$, we integrate over fields with $|k| < b\Lambda$. Change variables: $k = bk'$ so now $|k'| < \Lambda$. So $x = x'/b, \partial' \equiv \partial/\partial x' = \frac{1}{b}\partial_x$ and wavefunctions are preserved $e^{\mathbf{i}kx} = e^{\mathbf{i}k'x'}$. Plug this into the action

$$\int d^D x \mathcal{L}(\phi^<) = \int d^D x' b^{-D} \left( \frac{1}{2} b^2 \left(\partial' \phi^<\right)^2 + \sum_n \left(g_n + \delta g_n\right) \left(\phi^<\right)^n + ... \right)$$

We can make this look like $\mathcal{L}$ again by rescaling the field variable: $b^{2-D} \left(\partial' \phi^<\right)^2 \equiv \left(\partial' \phi'\right)^2$ (*i.e.* $\phi' \equiv b^{\frac{1}{2}(2-D)} \phi^<$):

$$\int d^D x' \mathcal{L}(\phi^<) = \int d^D x' \left( \frac{1}{2} \left(\partial' \phi'\right)^2 + \sum_n \left(g_n + \delta g_n\right) b^{-D + \frac{n(D-2)}{2}} \left(\phi'\right)^n + ... \right)$$

So the end result is that integrating out a momentum shell of thickness $\delta\Lambda \equiv (1-b)\Lambda$ results in a change of the couplings to

$$g_n' = b^{\frac{n(D-2)}{2} - D} \left(g_n + \delta g_n\right) \ .$$

This procedure produces a *flow* on the space of actions.

Ignore the interaction corrections, $\delta g_n$, for a moment. Then, since $b < 1$, the couplings with $\frac{n(D-2)}{2} - D > 0$ get smaller and smaller as we integrate out more shells. If we are interested in only the longest-wavelength modes, we can ignore these terms. They are *irrelevant*. Couplings ('operators') with $\frac{n(D-2)}{2} - D < 0$ get bigger and are *relevant*.

The mass term has $n = 2$ and $(m')^2 = b^{-2} m^2$ is always relevant for any $D < \infty$. So far, the counting is the same as our naive dimensional analysis. That's because we left out the $\delta L$ term! This term can make an important difference, even in perturbation theory, for the fate of marginal operators (such as $\phi^4$ in $D = 4$), where the would-be-big tree-level term is agnostic about whether they grow or shrink in the IR.

---

Notice that starting from (9.6) we are assuming that the system has a rotation invariance in euclidean momentum. If one of those euclidean directions is time, this follows from Lorentz invariance. This simplifies the discussion. But for non-relativistic systems, it is often necessary to scale time differently from space. The relative scaling $z$ in $\vec{x}' = b\vec{x}, t' = b^z t$ is called the *dynamical critical exponent*.

---

The definition of the beta function and of a fixed point theory is just as it was in the first lecture.

At this point we need to pick an example in which to include the interaction term.

## 9.3 An extended example: XY model

[R. Shankar, *Rev. Mod. Phys.* **66** (1994) 129]

Consider complex bosons in $D$ dimensions. I am a little tired of a real scalar field, so instead we will study *two* real scalar fields $\phi = \phi_1 + \mathbf{i}\phi_2$. We can define this model, for example, on a euclidean lattice, by an action of the form

$$S[\phi, \phi^\star] = \frac{1}{2}\sum_{n,i} |\phi(n) - \phi(n+i)|^2 + \sum_n u_0 |\phi(n)|^4 \quad . \tag{9.10}$$

Here $n$ labels sites of some (*e.g.* hypercubic) lattice and $i$ labels the (8 in the 4d hypercubic case) links connecting neighboring sites. We'll call the lattice spacing $2\pi/\Lambda_1$. In terms of Fourier modes, this is

$$S[\phi, \phi^\star] = -\int_{|k|<\Lambda_0} đ^D k \phi^\star(k) J(k)\phi(k) + S_{\text{int}} \ .$$

For the hyper-cubic lattice, we get (the second step is Taylor expansion)

$$J(k) = 2\left(\sum_{\mu=1}^{D} (\cos a k_\mu - 1)\right) \overset{ka \ll 1}{\simeq} \sum_\mu \left(a^2 k_\mu^2 + \frac{a^4}{4\cdot 3} k_\mu^4 ...\right).$$

The energy function $J(k)$ only has the discrete rotation symmetries of the lattice (90° rotations for the hypercubic lattice). But the leading term at small wavenumber has full rotation invariance; in position space, this term is $a^2 \partial_\mu \phi \partial^\mu \phi^\star$. The next term $\int đ^D k \ a^4 k^4 |\phi_k|^2 = \int d^D x \ a^4 \phi^\star \sum_\mu \partial_\mu^4 \phi$, which breaks the rotation group to a discrete subgroup, is irrelevant by the counting we did above: $\int d^D x \partial^4 \phi^2 \sim s^{D-4-2\frac{D-2}{2}} = s^{-2}$. This means that rotation invariance *emerges* on its own.

[30]

[End of Lecture 34]

The path integral is defined by

$$Z \equiv \int \underbrace{[\mathrm{d}\phi^\star \mathrm{d}\phi]_{|k|<\Lambda_0}}_{\equiv \prod_{|k|<\Lambda_0} \frac{\mathrm{dRe}\,\phi(k)\mathrm{dIm}\,\phi(k)}{\pi} = \prod_{|k|<\Lambda_0} \frac{\mathrm{d}\phi^\star(k)\mathrm{d}\phi(k)}{2\pi\mathbf{i}}} e^{-S[\phi,\phi^\star]} \tag{9.11}$$

There is a $U(1)$ global symmetry which acts by

$$\phi(k) \to e^{\mathbf{i}\theta}\phi(k), \phi^\star(k) \to e^{-\mathbf{i}\theta}\phi^\star(k) \ . \tag{9.12}$$

[30] Confession: the restriction on the momenta in the exact lattice model should be to a fundamental domain for the identification $k^\mu \equiv k^\mu + \Lambda_1^\mu$; I am going to replace this right away with a rotation-invariant cutoff on the magnitude $k^2 \equiv k^\mu k_\mu \leq \Lambda_0$ of the euclidean momentum. This is an unimportant lie for our purposes.

In terms of $\phi_{1,2}$, it acts by $\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \to \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$, which we should call $\mathsf{SO}(2)$ [31].

With $u_0 = 0$, this is a bunch of gaussian integrals, and everything can be computed by Wick from the two-point function:

$$\langle \phi^\star(k_1)\phi(k_2)\rangle = (2\pi)^D\, \delta^D(k_1 - k_2)\frac{1}{k_1^2} = (2\pi)^D\, \delta^D(k_1 - k_2) G(k_1).$$

Although this gaussian model is trivial, we can still do the RG to it. (We will turn on the interactions in a moment.) An RG step has three ingredients, of which I've emphasized only two so far:

1. Integrate out the fast modes, *i.e.* $\phi^>$, with $|k| \in (\Lambda - \delta\Lambda, \Lambda)$. I will call $\Lambda - \delta\Lambda \equiv \Lambda/s$, and[32] $s > 1$, we will regard $s$ as close to 1: $s - 1 \ll 1$.

$$\begin{aligned} Z &= \int \prod_{0\le|k|\le\Lambda/s} \mathrm{d}\phi_<(k) \left( \int \prod_{\Lambda/s\le|k|\le\Lambda} \mathrm{d}\phi_>(k) e^{-\left(\underbrace{S_0[\phi^<] + S_0[\phi^>]}_{\text{quadratic}} + \underbrace{S_{\text{int}}[\phi^<, \phi^>]}_{\text{mixes fast and slow}}\right)} \right) \\ &= \int [\mathrm{d}\phi^<] e^{-S_0[\phi^<]} \qquad \underbrace{\left\langle e^{-S_{\text{int}}[\phi^<, \phi^\cdot]} \right\rangle_{0,>}}_{\text{average over } \phi^>,\text{ with gaussian measure}} \qquad Z_{0,>} \end{aligned} \tag{9.13}$$

The factor of $Z_{0,>}$ is independent of $\phi^<$ and can be ignored.

2. Rescale momenta so that we may compare successive steps: $\tilde{k} \equiv sk$ lies in the same interval $|\tilde{k}| \in (0, \Lambda)$.

3. Are the actions $s(\phi) = r\phi^2 + u\phi^4$ and $\tilde{s}(\psi) = 4r\psi^2 + 16u\psi^4$ different? No: let $2\psi \equiv \phi$. We can resacle the field variable at each step:

$$\tilde{\phi}(\tilde{k}) \equiv \zeta^{-1}\phi_<(\tilde{k}/s).$$

We will choose the 'wavefunction renormalization' factor $\zeta$ so that the kinetic terms are fixed.

### RG for free field

[31] Actually, the symmetry of (9.11) is $\mathsf{O}(2)$, since $(\phi_1, \phi_2) \to (-\phi_1, \phi_2)$ is also a symmetry and has determinant minus one.

[32] I note that $s = 1/b$ from the previous subsection; sorry.

If $S_{\text{int}} = 0$, then (9.13) gives

$$\tilde{S}[\phi_<] = \int_{|k|<\Lambda/s} đ^D k \phi^\star_<(k) k^2 \phi_>(k) \overset{\text{steps 2 and 3}}{=} s^{-D-2}\zeta^2 \int_{|\tilde{k}|<\Lambda} \tilde{\phi}^\star(\tilde{k})\tilde{k}^2\tilde{\phi}(\tilde{k})đ^D\tilde{k} \ .$$

With $\zeta \equiv s^{\frac{D+2}{2}}$, the Gaussian action is a fixed point of the RG step:

$$\tilde{S}[\tilde{\phi}] = S[\phi] = S^\star.$$

---

Warning: the field $\phi(k)$ is the Fourier transform of the field $\phi(x)$ that we considered above. They are different by an integral over space or momenta: $\phi(x) = \int đ^D k \phi(k) e^{\mathbf{i}kx}$. So they scale differently. The result that $\zeta = s^{\frac{D+2}{2}}$ is perfectly consistent with our earlier result that $\phi(x)$ scales like $s^{\frac{2-D}{2}}$.

---

Now we consider perturbations. We'll only study those that preserve the symmetry (9.12). We can order them by their degree in $\phi$. The first nontrivial case preserving the symmetry is

$$\delta S_2[\phi] = \int_{|k|<\Lambda} đ^D k \phi^\star(k)\phi(k) r(k) \ .$$

Here $r(k)$ is a coupling *function*. If its position-space representation is local, it has a nice Taylor expansion about $k = 0$:

$$r(k) = \underbrace{r_0}_{\equiv m_0^2} + k^2 r_2 + \ldots$$

(I also assumed rotation invariance.) The same manipulation as above gives

$$\widetilde{\delta S_2}[\tilde{\phi}(\tilde{k})] = s^{-D+\frac{D+2}{2}2=2} \int_{|\tilde{k}|<\Lambda} \tilde{\phi}^\star(\tilde{k})\tilde{r}(\tilde{k})\tilde{\phi}(\tilde{k})đ^D\tilde{k}$$

with $\tilde{r}(\tilde{k}) = s^2 r(\tilde{k}/s)$, so that

$$\underbrace{\tilde{r}_0 = s^2 r_0}_{\text{relevant}} \ , \quad \underbrace{\tilde{r}_2 = s^0 r_2}_{\text{marginal by design}} \ , \quad \underbrace{\tilde{r}_4 = s^{-2} r_4}_{\text{irrelevant}} \ldots$$

**Quartic perturbation**

$$\delta S_4 = S_{\text{int}} = \int_\Lambda \phi^\star(4)\phi^\star(3)\phi(2)\phi(1)u(4321)$$

This is some shorthand notation for

$$\delta S_4 = S_{\text{int}} = \frac{1}{(2!)^2}\int \prod_{i=1}^{4} đ^D k_i (2\pi)^D \delta^D(k_4+k_3-k_2-k_1)\phi^\star(k_4)\phi^\star(k_3)\phi(k_2)\phi(k_1)u(k_4k_3k_2k_1) \ .$$

The delta function maintains translation invariance in real space. Here $u(4321)$ is some general function, but only the bit with $u(4321) = u(3421) = u(4312)$ matters. This interaction couples the fast and slow modes. We need to evaluate

$$e^{-\tilde{S}[\phi_<]} = e^{-S_0[\phi_<]} \left\langle e^{-\delta S[\phi_<, \phi_>]} \right\rangle_{0,>} .$$

A tool at our disposal is the *cumulant expansion*, aka the exponentiation of the disconnected diagrams:

$$\left\langle e^{-\Omega} \right\rangle = e^{-\langle \Omega \rangle + \frac{1}{2}\left(\left\langle \Omega^2 \right\rangle - \langle \Omega \rangle^2\right) + \ldots}$$

So

$$\widetilde{\delta S} = \underbrace{\langle \delta S \rangle_{>,0}}_{\sim u_0} - \frac{1}{2} \underbrace{\left( \left\langle \delta S^2 \right\rangle_{>,0} - \langle \delta S \rangle_{>,0}^2 \right)}_{\sim u_0^2} + \ldots$$

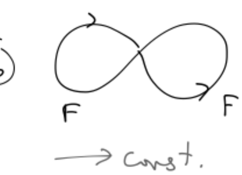
So this expansion is a perturbative expansion in $u_0$.

First the first term ($\sim u_0$):

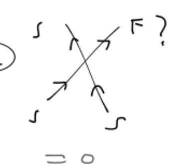
$$\langle \delta S \rangle_{>,0} = \frac{u_0}{(2!)^2} \int_{|k|<\Lambda} \langle (\phi_< + \phi_>)_4^\star (\phi_< + \phi_>)_3^\star (\phi_< + \phi_>)_2 (\phi_< + \phi_>)_1 u(4321) \rangle_{>,0}$$

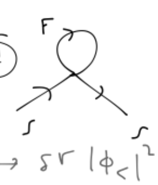
This is made of 16 terms which can be decomposed as follows, and illustrated by the Feynman diagrams at right. These Feynman diagrams are just like the usual ones with the important difference that the loop momenta only run over the shell from $|k| = \Lambda/s$ to $|k| = \Lambda$. The only allowed external lines are the slow modes. The ones that contribute to the $\mathcal{O}(u_0)$ term all have a single 4-point vertex.

a 


b 


(a) 1 diagram with all external lines being slow modes. This gives the tree level interaction term for the slow modes.

c 


d 


(b) 1 diagram with only fast modes involved in the vertex. This contributes to the irrelevant constant $Z_{0,>}$.

(c) 8 diagrams with an odd number of fast modes; these all vanish by the usual Wick business.

(d) 6 diagrams with 2 slow 2 fast. The fast modes must be contracted and this makes a loop. The arrows (representing the flow of the U(1) charge) must work out to allow nonzero contractions (recall that $\langle \phi\phi \rangle = 0$ by charge conservation).

So the only interesting ones are diagrams of type (d), which give

$$\begin{aligned}
\widetilde{\delta S_2}(\phi_<) &= \frac{u_0}{(2!)^2} \int_{|k|<\Lambda} \langle (\phi^\star_>(4)\phi^\star_<(3) + \phi^\star_>(3)\phi^\star_<(4))(\phi_>(2)\phi_<(1) + \phi_>(1)\phi_<(2))\rangle_{0,>} \\
&= u_0 \int_{|k|<\Lambda/s} đ^D k \phi^\star_<(k)\phi_<(k) \quad \cdot \underbrace{\int_{\Lambda/s}^{\Lambda} đ^D p \frac{1}{p^2}}_{\substack{= \frac{\Omega_{D-1}}{(2\pi)^D}\int_{\Lambda/s}^{\Lambda} k^{D-3}dk \\ \overset{D=4}{=} \frac{2\pi^2}{(2\pi)^4}\frac{\Lambda^2}{2}(1-s^{-2})\,.}}
\end{aligned} \tag{9.14}$$

$$\widetilde{\delta S}_2[\tilde\phi_<(\tilde k)] = u_0 s^2 \int_{|\tilde k|<\Lambda} đ^4 k \tilde\phi^\star(\tilde k)\tilde\phi(\tilde k)\frac{\Lambda^2}{16\pi^2}(1-s^{-2}).$$

$$\boxed{\delta r_0 = \frac{u_0 \Lambda^2}{16\pi^2}(s^2-1)}.$$

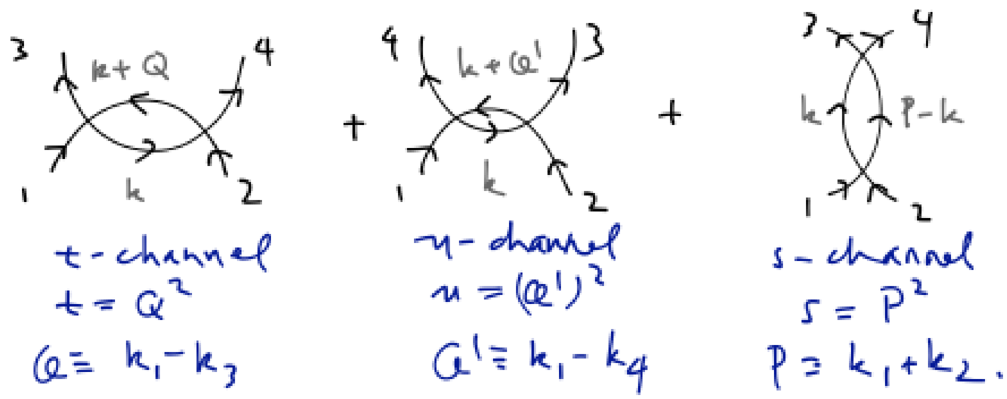
The correction to the mass is of order the cutoff.

In $D$ dimensions, we get instead

$$\delta r_0 = \frac{\Omega_{D-1}}{(2\pi)^D} u_0 \Lambda^{D-2}(s^2 - s^{4-D}).$$

**The next term in the cumulant expansion**

Now for the $\mathcal{O}(u_0^2)$ term in $\widetilde{\delta S}$. The diagrammatic representation of $\frac{1}{2}\left(\langle \delta S^2 \rangle - \langle \delta S \rangle^2\right)$ is: all connected diagrams containing *two* 4-point vertices, with only external slow lines. The second term cancels all disconnected diagrams. Diagrammatically, these are (we are in Euclidean spacetime here, so I don't mind violating my rule that time goes to the left):




These correct the quartic coupling $u = u_0 + u_1 k^2 + ....$ We care about the sign of $\delta u_0$, because in $D = 4$ it is marginal. Even small corrections will make a big difference.

$$\tilde{u}(\tilde{k}_4, ...\tilde{k}_1) = u_0 - u_0^2 \underbrace{\int_{\Lambda/s}^{\Lambda} đ^D k}_{\equiv \int_{d\Lambda}} \left( \frac{1}{k^2 |k - (\tilde{k}_3 - \tilde{k}_1)/s|^2} + \frac{1}{k^2 |k - (\tilde{k}_4 - \tilde{k}_1)/s|^2} + \frac{1}{2} \frac{1}{k^2 |-k - (\tilde{k}_1 + \tilde{k}_2)/s|^2} \right)$$

Note the symmetry factor in the $s$-channel diagram, which you can see directly from the cumulant expression.

[End of Lecture 35]

The most interesting part of this expression is the correction to $u_0$, which is what we get when we set the external momenta to zero:

$$\tilde{u}(k = 0) = \tilde{u}_0 = u_0 - u_0^2 \frac{5}{2} \underbrace{\int_{d\Lambda} \frac{k^3 dk}{k^4}}_{= \log s} \cdot \underbrace{\frac{\Omega_3}{(2\pi)^4}}_{= \frac{1}{16\pi^2}}.$$

Let $\Lambda(s) \equiv \Lambda_0/s \equiv \Lambda_0 e^{-\ell}$ so $s = e^\ell, \ell = \log \Lambda_0/\Lambda$ and $\Lambda \frac{d}{d\Lambda} = s\partial_s = \partial_\ell$. Large $\ell$ is the IR.

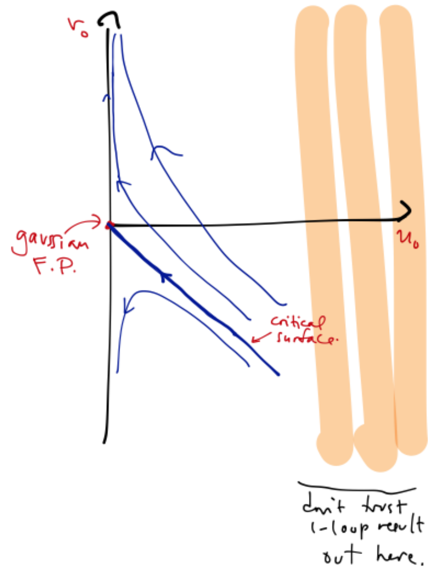
$$\begin{cases} \frac{du_0}{d\ell} = -\frac{5}{16\pi^2} u_0^2 \equiv -b u_0^2 \\ \frac{d\hat{r}_0}{d\ell} = 2\hat{r}_0 + \frac{u_0}{16\pi^2} = 2r_0 + a u_0 \end{cases} . \tag{9.15}$$

Here $a, b > 0$ are constants, and $\hat{r}_0 \equiv r_0 \Lambda^2$ is the mass$^2$ in units of the cutoff. (Note that the usual high-energy definition of the beta function has the opposite sign, $\frac{dg}{d\ell} = -\beta_g$.)

These equations can be solved in terms of two initial conditions:

$$u_0(\ell) = \frac{u_0(0)}{1+bu_0(0)\ell} \overset{\ell\to\infty, u_0(0)>0}{\sim} \frac{1}{\ell} = \frac{1}{\log \Lambda_0/\Lambda} \to 0.$$

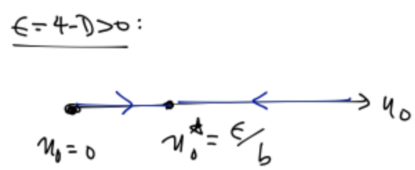
$u_0$ is a marginally irrelevant perturbation of the gaussian fixed point. This theory is not asymptotically free[33] The phase diagram is at right. There's just the one fixed Gaussian point. Notice that it's not true that an arbitrary small $u_0$ added to the gaussian FP runs back to the gaussian FP. $r_0$ runs too:




$$r_0(\ell) = e^{2\ell}\left[r_0(0) + \int_0^\ell e^{-2\ell'}\frac{au_0(0)}{1+bu_0(0)\ell'}\mathrm{d}\ell'\right].$$

There is a curve of choices of initial data in $(u_0(0), r_0(0))$ which ends up at the origin – it's when the thing in brackets vanishes; for small $u_0$, this is the line $r_0(0) = -\frac{a}{2}u_0(0)$.

Following Wilson and Fisher, it is an extremely good idea to consider dimensions other than 4, $D \equiv 4-\epsilon$. We've already been willing to do this as a regulator of short-distance physics; it turns out that it also resolves some short-distance physics in the phase diagram. If $D \neq 4$, the quartic interaction is no longer marginal at tree level, but rather scales like $s^\epsilon$. The RG equation is modified to




$$\frac{du_0}{dt} = \epsilon u_0 - bu_0^2 \quad . \tag{9.16}$$

For $\epsilon > 0$ ($D < 4$) there is another fixed point at $u_0^\star = \epsilon/b > 0$. And in fact the Gaussian FP is unstable, and this *Wilson-Fisher fixed point* is the stable one in the IR (see fig at right, which is drawn along the critical surface leading to $r_0(\infty) = 0$.). This situation allows one to calculate (universal) critical exponents at the fixed point in an expansion in $\epsilon$.

As $\epsilon \to 0$, the two fixed points coalesce.

The W-F fixed point describes a continuous phase transition between ordered and disordered phases. An external variable (roughly $r_0$) must be tuned to reach the phase

[33]This statement was for $u_0(0) > 0$. For $u_0(0) < 0$, it is AF (this was an observation of Symanzik, before the study of Yang-Mills), but seems likely to be unstable. For an interesting claim to the contrary, see here if you are feeling brave. It would be nice to know for sure.

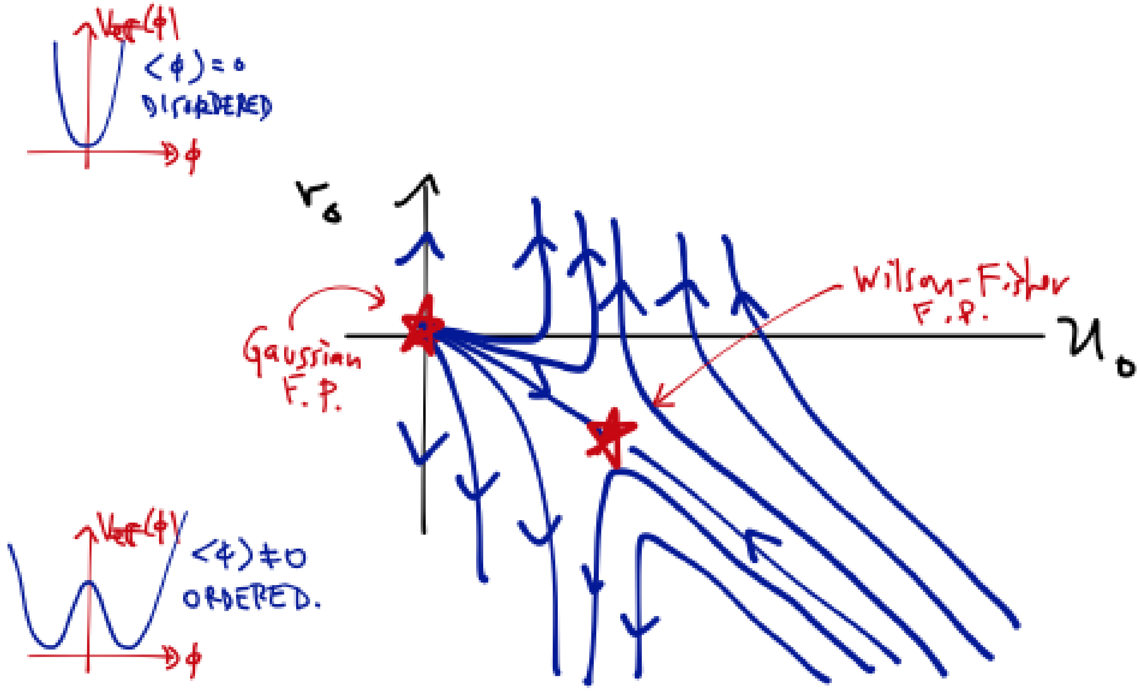



**Figure 6:** The $\phi^4$ phase diagram. If $r_0(\ell = \infty) > 0$, the effective potential for the uniform 'magnetization' has a minimum at the origin; this is the disordered phase, where there is no magnetization. If $r_0(\ell = \infty) = V''_{\text{eff}} < 0$, the effective potential has minima away from the origin, and the groundstate breaks the symmetry (here $\phi \to e^{\mathbf{i}\theta}\phi$); this is the ordered phase.

transition. A physical realization of this is the following: think of our euclidean path integral as a thermal partition function at temperature $1/\beta$:

$$Z = \int [D\phi] e^{-\beta H[\phi]} \ ;$$

here we are integrating over thermal fluctuations of classical fields. Above we've studied the case with $\mathsf{O}(2)$ symmetry (called the XY model). WLOG, we can choose normalize our fields so that the coefficient $\beta$ determines $r_0$. The critical value of $r_0$ then realizes the critical temperature at which this system goes from a high-temperature disordered phase to a low-temperature ordered phase. For this kind of application, $D \leq 3$ is most interesting physically. We will see that the $\epsilon$ expansion about $D = 4$ is nevertheless quite useful.

You could ask me what it means for the number of dimensions $D$ to be not an integer. One correct answer is that we have constructed various well-defined functions of continuous $D$ simply by keeping $D$ arbitrary; basically all we need to know is the volume of a $D$-sphere for continuous $D$, (6.38). An also-correct answer that some people (*e.g.* me) find more satisfying is is the following. Suppose we can define our QFT by a discrete model, defined on a discretized space (like in (9.10)). Then we can also put the model on a graph whose *fractal dimension* is not an integer. Evidence that this is a physical realization of QFT in non-integer dimensions is given in [Gefen-Meir-Mandelbrot-Aharony] and [Gefen-Mandelbrot-Aharony]. Some subtle and interesting issues about uniqueness and unitarity of the field theories so defined are raised here and here.

**Important lessons.**

- Elimination of modes does not introduce new singularities into the couplings. At each step of the RG, we integrate out a finite-width shell in momentum space – we are doing integrals which are convergent in the infrared and ultraviolet.

- The RG plays nicely with symmetries. In particular any symmetry *of the regulated model* is a symmetry of the long-wavelength effective action.[34]

- Some people conclude from the field theory calculation of the $\phi^4$ beta function that $\phi^4$ theory "does not exist" or "is trivial", in the sense that *if we demand that this description is valid up to arbitrarily short distances*, we would need to pick $u(\Lambda = \infty) = \infty$ in order to get a finite interaction strength at long wavelengths. You can now see that this is a ridiculous conclusion. Obviously the theory exists in a useful sense. It can easily be defined at short distances (for example) in terms of the lattice model we wrote at the beginning of this subsection. Similar statements apply to QED.

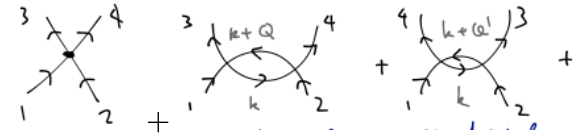
- The corrections to the mass of the scalar field are of order of the cutoff. This makes it hard to understand how you could arrive in the IR and find that an interacting scalar field has a mass which is much smaller than the cutoff. Yet, there seems to be a Higgs boson with $m \simeq 125$ GeV, and no cutoff on the Standard Model in sight. This is a mystery.

- As Tony Zee says, a more accurate (if less catchy) name than 'renormalization group' would be 'the trick of doing the path integral a little at a time'.

[34]The extra qualifier about the regulated model is important because some symmetries of continuum classical field theories cannot be realized as symmetries of well-defined quantum field theories. We will discuss this phenomenon, called *anomalies*, in the near future.

### 9.3.1 Comparison with renormalization by counterterms

Is this procedure the same as 'renormalization' in the high-energy physics sense of sweeping divergences under the rug of bare couplings? Suppose we impose the renormalization condition that $\Gamma_4(k_4...k_1) \equiv \Gamma(4321)$, the 1PI 4-point vertex, is cutoff independent. Its leading contributions come from the diagrams:




(where now the diagrams denote amputated amplitudes, the arrows indicate flow of scalar charge (since we're studying the case with $\mathsf{O}(2)$ symmetry) and also momentum, and the integrals run over all momenta up to the cutoff). Clearly there is already a big similarity. In more detail, this is

$$\Gamma(4321) = u_0 - u_0^2 \int_0^\Lambda đ^D k$$
$$\left( \frac{1}{(k^2+r_0)(|k+k_3-k_1|^2+r_0)} + \frac{1}{(k^2+r_0)(|k+k_4-k_1|^2+r_0)} + \frac{1}{2}\frac{1}{(k^2+r_0)(|-k+k_1+k_2|^2+r_0)} \right)$$

And in particular, the bit that matters is

$$\Gamma(0000) = u_0 - u_0^2 \frac{5}{32\pi^2} \log \frac{\Lambda^2}{r_0}.$$

Demanding that this be independent of the cutoff $\Lambda = e^{-\ell}\Lambda_0$,

$$0 = \partial_\ell \left( \Gamma(0000) \right) = -\Lambda \frac{d}{d\Lambda} \Gamma(0000)$$

gives

$$0 = \frac{du_0}{d\ell} + \frac{5}{16\pi^2} u_0^2 + \mathcal{O}(u_0^3)$$
$$\implies \boxed{\beta_{u_0} = -\frac{5}{16\pi^2} u_0^2} + \mathcal{O}(u_0^3)$$

as before. (The bit that would come from $\partial_\ell u_0^2$ in the second term is of order $u_0^3$ and so of the order of things we are already neglecting.)

I leave it to you to show that the flow for $r_0$ that results from demanding that $\langle \phi(k)\phi^\star(k) \rangle$ have a pole at $k^2 = -m^2$ (with $m$ independent of the cutoff) gives the same flow we found above.

It is worth noting that although the continuum field theory perspective with counterterms is less philosophically satisfying, it is often easier for actual calculations than integrating momentum shells, mainly because we can use a convenient regulator like dim reg.

[End of Lecture 36]

### 9.3.2 Comment on critical exponents

[Zinn-Justin, chapter 25, Peskin, chapter 12.5, Stone, chapter 16, Cardy, and the classic Kogut-Wilson]

Recall that the Landau-Ginzburg mean field theory made a (wrong) prediction for the critical exponents at the Ising transition:

$$\langle M\rangle \sim (T_c - T)^{\beta} \qquad \text{for } T < T_c, \quad \xi \sim (T_c - T)^{-\nu}$$

with $\beta_{MFT} = \frac{1}{2}$, $\nu_{MFT} = \frac{1}{2}$. This answer was wrong (*e.g.* for the Ising transition in (euclidean) $D = 3$, which describes uniaxial magnets (spin is $\pm 1$) or the liquid-gas critical point) because it simply ignored the effects of fluctuations of the modes of nonzero wavelength, *i.e.* the $\delta L$ bit in (9.8). I emphasize that these numbers are worth getting right because they are universal – they are properties of a fixed point, which are completely independent of any microscopic details.

Now that we have learned to include the effects of fluctuations at all length scales on the long-wavelength physics, we can do better. We've done a calculation which includes fluctuations at the transition for an XY magnet (the spin has two components, and a U(1) symmetry that rotates them into each other), and is also relevant to certain systems of bosons with conserved particle number. The mean field theory prediction for the exponents is the same as for the Ising case (recall that we did the calculation for a magnetization field with an arbitrary number $N$ of components, and in fact the mean field theory prediction is independent of $N \geq 1$; we'll say more about general $N$-component magnets below).

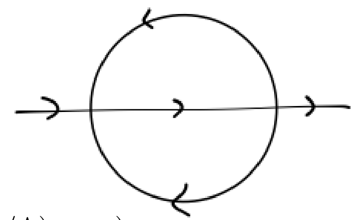
In general there are many *scaling relations* between various critical exponents, which can be understood beginning from the effective action, and were understood before the correct calculation of the exponents. So not all of them are independent. For illustration, we will briefly discuss two independent exponents.

**Order parameter exponent, $\eta$.** The simplest critical exponent to understand from what we've done so far is $\eta$, the exponent associated with the anomalous dimension of the field $\phi$ itself. (It is not the easiest to actually calculate, however.) This can be defined in terms of the (momentum-space) amputated two-point function of $\phi$ (that is, $\Gamma_2(p) = 1/\tilde{G}(p)$) as

$$\Gamma_2(p) \overset{\xi^{-1} \ll p \ll \Lambda}{\simeq} \left(\frac{p}{\Lambda}\right)^{2-\eta}$$

where $\xi$ is the correlation length and $\Lambda$ is the UV cutoff. This looks a bit crazy – at nonzero $\eta$, the full propagator has a weird power-law singularity instead of a $\frac{1}{p^2 - m^2}$, and in position space it is a power law $G_2(x) \sim \frac{1}{|x|^{D-2+\eta}}$, instead of an exponential decay. An example where all the details can be understood is the operator $e^{\mathbf{i}\alpha X}$ the massless scalar field $X$ in 1+1 dimensions (see the homework).

But how can this happen in perturbation theory? Consider physics near the gaussian fixed point, where $\eta$ must be small, in which case we can expand:



$$\Gamma_2(p) \overset{\xi^{-1}\ll p\ll\Lambda,\eta\ll 1}{\simeq} \left(\frac{p}{\Lambda}\right)^2 \left(e^{-\eta\log(p/\Lambda)}\right) = \left(\frac{p}{\Lambda}\right)^2 (1-\eta\log(p/\Lambda)+...)$$

In the $\phi^4$ theory, $\eta = 0$ at one loop. The leading correction to $\eta$ comes from the 'sunrise' (or 'eyeball') diagram at right, at two loops. So in this model, $\eta \sim g_\star^2 \sim \epsilon^2$. $\Gamma_2(p)$ is the 1PI momentum space 2-point vertex, *i.e.* the kinetic operator. We can interpret a nonzero $\eta$ as saying that the dimension of $\phi$, which in the free theory was $\Delta_0 = \frac{2-D}{2}$, has been modified by the interactions to $\Delta = \frac{2-D}{2} - \eta/2$. $\eta/2$ is the *anomalous dimension* of $\phi$. Quantum mechanics violates (naive) dimensional analysis; it must, since it violates classical scale invariance. Of course (slightly more sophisticated) dimensional analysis is still true – the extra length scale is the UV cutoff, or some other scale involved in the renormalization procedure.

**Correlation length exponent, $\nu$.** Returning to the correlation length exponent $\nu$, we can proceed as follows. First we relate the scaling of the correlation length to the scaling behavior of the relevant perturbation that takes us away from from the fixed point. The latter we will evaluate subsequently in our example. (There is actually an easier way to do this, discussed in §9.3.3, but this will be instructive.)

The correlation length is the length scale above which the relevant perturbation gets big and cuts off the critical fluctuations of the fixed point. As the actual fixed point is approached, this happens at longer and longer scales: $\xi$ diverges at a rate determined by the exponent $\nu$.

Suppose we begin our RG procedure with a perturbation of a fixed-point Hamiltonian by a relevant operator $\mathcal{O}$:

$$H(\xi_1) = H_\star + \delta_1 \mathcal{O} \quad .$$

Under a step of the RG, $\xi_1 \to s^{-1}\xi_1$, $\delta_1 \to s^{\Delta}\delta_1$, where I have defined $\Delta$ to be the scaling dimension of the operator $\mathcal{O}$. Then after $N$ steps, $\delta = s^{N\Delta}\delta_1, \xi = s^{-N}\xi_1$. Eliminating $s^N$ from these equations we get the relation

$$\xi = \xi_1 \left(\frac{\delta}{\delta_1}\right)^{-\frac{1}{\Delta}} \tag{9.17}$$

which is the definition of the correlation length exponent $\nu$, and we conclude that $\nu = \frac{1}{\Delta}$.

Here is a better way to think about this. At the critical point, the two-point function of the order parameter $G(x) \equiv \langle\phi(x)\phi(0)\rangle$ is a power law in $x$, specified by $\eta$. Away from the critical point, there is another scale, namely the size of the perturbation – the deviation of the microscopic knob $\delta_0$ from its critical value, such as $T - T_c$. Therefore, dimensional analysis says that $G(x)$ takes the form

$$G(x) = \frac{1}{|x|^{D-2}} \left(\frac{1}{|x|/a}\right)^{\eta} \Phi\left(|x|\delta_0^{1/\Delta}\right)$$

where the argument of the *scaling function* $\Phi$ is dimensionless. (I emphasized that some length scale $a$, such as the lattice spacing, must make up the extra engineering dimensions to allow for an anomalous dimension of the field at the critical point.) When $x \gg$ all other length scales, $G(x)$ should decay exponentially, and the decay length must then be $\xi \sim \delta_0^{-\frac{1}{\Delta}}$ which says $\nu = \frac{1}{\Delta}$.

In the case of $\phi^4$ theory, $r_0$ is the parameter that an experimentalist must carefully tune to access the critical point (what I just called $\delta_0$) – it is the coefficient of the relevant operator $\mathcal{O} = |\phi|^2$ which takes us away from the critical point; it plays the role of $T - T_c$.

At the free fixed point the dimension of $|\phi|^2$ is just twice that of $\phi$, and we get $\nu^{-1} = \Delta^{(0)}_{|\phi|^2} = 2\frac{D-2}{2} = D - 2$. At the nontrivial fixed point, however, notice that $|\phi|^2$ is a composite operator in an interacting field theory. In particular, its scaling dimension is not just twice that of $\phi$! This requires a bit of a digression.

---

**Renormalization of composite operators.**

[Peskin §12.4] Perturbing the Wilson-Fisher fixed point by this seemingly-innocuous quadratic operator, is then no longer quite so innocent. In particular, we must define what we mean by the operator $|\phi|^2$! One way to define it (from the counterterms point of view, now, following Peskin and Zinn-Justin) is by adding an extra renormalization condition[35]. We can define the normalization of the composite operator $\mathcal{O}(k) \equiv |\phi|^2(k)$ by the condition that its (amputated) 3-point function gives

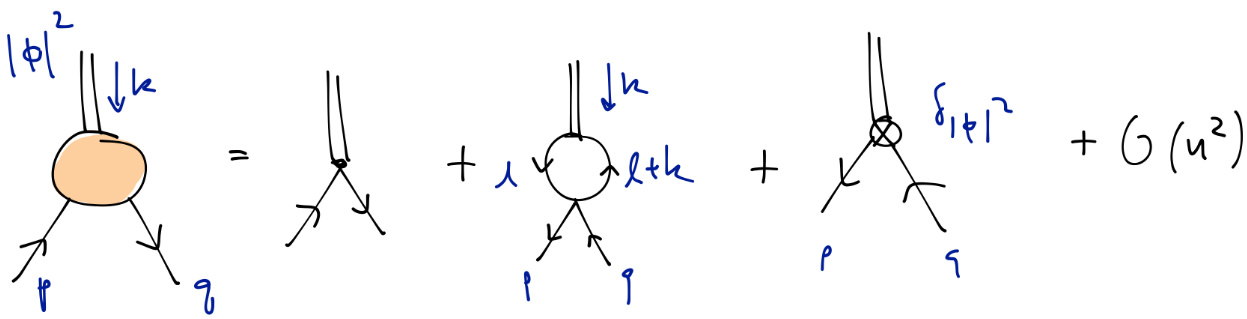
$$\langle \mathcal{O}_\Lambda(k)\phi(p)\phi^\star(q)\rangle = 1 \text{ at } p^2 = q^2 = k^2 = -\Lambda^2 \ .$$

The subscript on $\mathcal{O}_\Lambda(k)$ is to emphasize that its (multiplicative) normalization is defined by a renormalization condition at scale (spacelike momentum) $\Lambda$. Just like for the 'elementary fields', we can define a wavefunction renormalization factor:

$$\mathcal{O}_\Lambda \equiv Z_{\mathcal{O}}^{-1}(\Lambda)\mathcal{O}_\infty$$

where $\mathcal{O}_\infty \equiv \phi^\star\phi$ is the bare product of fields.

[35] Note that various factors differ from Peskin's discussion in §12.4 because I am discussing a complex field $\phi \neq \phi^\star$; this changes the symmetry factors.



We can represent the implementation of this prescription diagramatically. In the diagram above, the double line is a new kind of thing – it represents the insertion of $\mathcal{O}_\Lambda$. The vertex where it meets the two $\phi$ lines is *not* the 4-point vertex associated with the interaction – two $\phi$s can turn into two $\phi$s even in the free theory. The one-loop, 1PI correction to this correlator is (the second diagram on the RHS of the figure)[36]

$$(-u_0)\int_0^\infty đ^D\ell \frac{1}{\ell^2}\frac{1}{(k+\ell)^2} = -u_0 \frac{c}{k^{4-D}}$$

where $c$ is a number (I think it is $c = \frac{\Gamma\left(2-\frac{D}{2}\right)}{(4\pi)^2}$) and we know the $k$ dependence of the integral by scaling. If you like, I am using dimensional regularization here, thinking of the answer as an analytic function of $D$.

Imposing the renormalization condition requires us to add a counterterm diagram (part of the definition of $|\phi|^2$, indicated by the $\otimes$ in the diagrams above) which adds

$$Z_{\mathcal{O}}^{-1}(\Lambda) - 1 \equiv \delta_{|\phi|^2} = \frac{u_0 c}{\Lambda^{4-D}} \ .$$

We can infer the dimension of (the well-defined) $|\phi|^2_\Lambda$ by writing a renormalization group equation for our 3-point function

$$G^{(2;1)} \equiv \left\langle |\phi|^2_\Lambda(k)\phi(p)\phi^\star(q)\right\rangle .$$

$$0 \overset{!}{=} \Lambda\frac{d}{d\Lambda}G^{(n;1)} = \left(\Lambda\frac{\partial}{\partial\Lambda} + \beta(u)\frac{\partial}{\partial u} + n\gamma_\phi + \gamma_{\mathcal{O}}\right)G^{(n;1)} \ .$$

This (Callan-Symanzik equation) is the demand that physics is independent of the cutoff. $\gamma_{\mathcal{O}} \equiv \Lambda\frac{\partial}{\partial\Lambda}\log Z_{\mathcal{O}}(\Lambda)$ is the *anomalous dimension* of the operator $\mathcal{O}$, roughly the addition to its engineering dimension coming from the interactions (similarly $\gamma_\phi \equiv \Lambda\frac{\partial}{\partial\Lambda}\log Z_\phi(\Lambda)$). To leading order in $u_0$, we learn that

$$\gamma_{\mathcal{O}} = \Lambda\frac{\partial}{\partial\Lambda}\left(-\delta_{\mathcal{O}} + \frac{n}{2}\delta_Z\right)$$

[36]At higher order in $u_0$, the wavefunction renormalization of $\phi$ will also contribute to the renormalization of $|\phi|^2$.

which for our example with $n = 2$ gives the anomalous dimension of $|\phi|^2$ to be (just the first term to this order since $\delta_Z$ is the wavefunction renormalization of $\phi$, which as we discussed first happens at $\mathcal{O}(u_0^2)$)

$$\gamma_{|\phi|^2} = \frac{2u_0}{16\pi^2}.$$

---

Plugging in numbers, we get, at the $N = 2$ (XY) Wilson-Fisher fixed point at $u_0^\star = \epsilon/b$,

$$\nu = \frac{1}{\Delta_{|\phi|^2}} = \frac{1}{2 - \gamma_{|\phi|^2}} \overset{D=4-\epsilon}{=} \frac{1}{2 - \frac{2u_0^\star}{16\pi^2}} = \frac{1}{2 - 2\frac{16\pi^2}{5}\frac{\epsilon}{16\pi^2}} = \frac{1}{2 - \frac{2\epsilon}{5}} \; .$$

(for the Ising fixed point the $5/2$ would be replaced by $\frac{N+8}{N+2}|_{N=1} = 3$).

It is rather amazing how well one can do at estimating the answers for $D = 3$ by expanding in $\epsilon = 4 - D$, keeping the leading order correction, and setting $\epsilon = 1$. The answer from experiment and the lattice is $\nu_{D=3,N=2} \simeq 0.67$, while we find $\nu_{\epsilon=1,N=2} \simeq 0.63$. It is better than mean field theory for sure. You can do even better by Padé approximating the $\epsilon$ expansion. Currently (and for the foreseeable future) the best answer comes from the conformal bootstrap.

---

One final comment about defining and renormalizing composite operators: if there are multiple operators with the same quantum numbers and the same scaling dimension, they will *mix* under renormalization. That is, in order to obtain cutoff-independent correlators of these operators, their definition must be of the form

$$\mathcal{O}^i_\Lambda = \left(Z^{-1}(\Lambda)\right)_{ij} \mathcal{O}^j_\infty$$

– there is a wavefunction renormalization *matrix*, and a matrix of anomalous dimensions

$$\gamma_{ij} = -\Lambda \partial_\Lambda \log \left(Z^{-1}(\Lambda)\right)_{ij} .$$

'Operator mixing' is really just the statement that correlation functions like $\langle \mathcal{O}^i \mathcal{O}^j \rangle$ are nonzero.

---

### 9.3.3 Once more, with feeling (and an arbitrary number of components)

I've decided to skip this subsection in lecture. You may find it useful for the homework.

[Kardar, *Fields*, §5.5, 5.6] Let's derive the RG for $\phi^4$ theory again, with a number of improvements:

- Instead of two components, we'll do $N$ component fields, with $\mathcal{U} = \int d^D x u_0 \left(\phi^a \phi^a\right)^2$ (repeated indices are summed, $a = 1..N$).
- We'll show that it's not actually necessary to ever do any momentum integrals to derive the RG equations.
- We'll keep the mass perturbation in the discussion at each step; this lets us do the following:
- We'll show how to get the correlation length exponent without that annoying discussion of composite operators. (Which was still worth doing because in other contexts it is not avoidable.)

We'll now assume O($N$) symmetry, $\phi^a \to R^a_b \phi^b$, with $R^t R = \mathbb{1}_{N\times N}$, and perturb about the gaussian fixed point with (euclidean) action

$$S_0[\phi] = \int_0^\Lambda đ^D k \underbrace{\phi^a(k)\phi^a(-k)}_{\equiv |\phi|^2(k)} \frac{1}{2}\left(r_0 + r_2 k^2\right).$$

The coefficient $r_2$ of the kinetic term is a book-keeping device that we may set to 1 if we choose. Again we break up our fields into slow and fast, and integrate out the fast modes:

$$Z_\Lambda = \int [D\phi_<] e^{-\int_0^{\Lambda/s} đ^D k |\phi_<(k)|^2 \left(\frac{r_0 + r_2 k^2}{2}\right)} Z_{0,>} \left\langle e^{-\mathcal{U}[\phi_<,\phi_>]}\right\rangle_{0,>} .$$

Again the $\langle ...\rangle_{0,>}$ means averaging over the fast modes with their Gaussian measure, and $Z_{0,>}$ is an irrelevant normalization factor, independent of the objects of our fascination, the slow modes $\phi_<$. With $N$ components we do Wick contractions using

$$\left\langle \phi^a_>(q_1)\phi^b_>(q_2)\right\rangle_{0,>} = \frac{\delta^{ab} \not{\delta}(q_1+q_2)}{r_0 + q_1^2 r_2}.$$

I've defined $\not{\delta}(q) \equiv (2\pi)^D \delta^D(q)$. Notice that we are now going to keep the mass perturbation $r_0$ in the discussion at each step. Again

$$\log\left\langle e^{-\mathcal{U}}\right\rangle_{0,>} = -\underbrace{\langle \mathcal{U}\rangle_{0,>}}_{1} + \underbrace{\frac{1}{2}\left(\left\langle \mathcal{U}^2\right\rangle_{0,>} - \langle \mathcal{U}\rangle^2_{0,>}\right)}_{2}$$

$$1 = \langle \mathcal{U}[\phi_<, \phi_>] \rangle_{0,>} = u_0 \int \prod_{i=1}^{4} \bar{d}^D k_i \delta\!\!\!/(\sum_i k_i) \left\langle \prod_i (\phi_< + \phi_>)_i \right\rangle_{0,>}$$

Diagramatically, these 16 terms decompose as in Fig. 7.

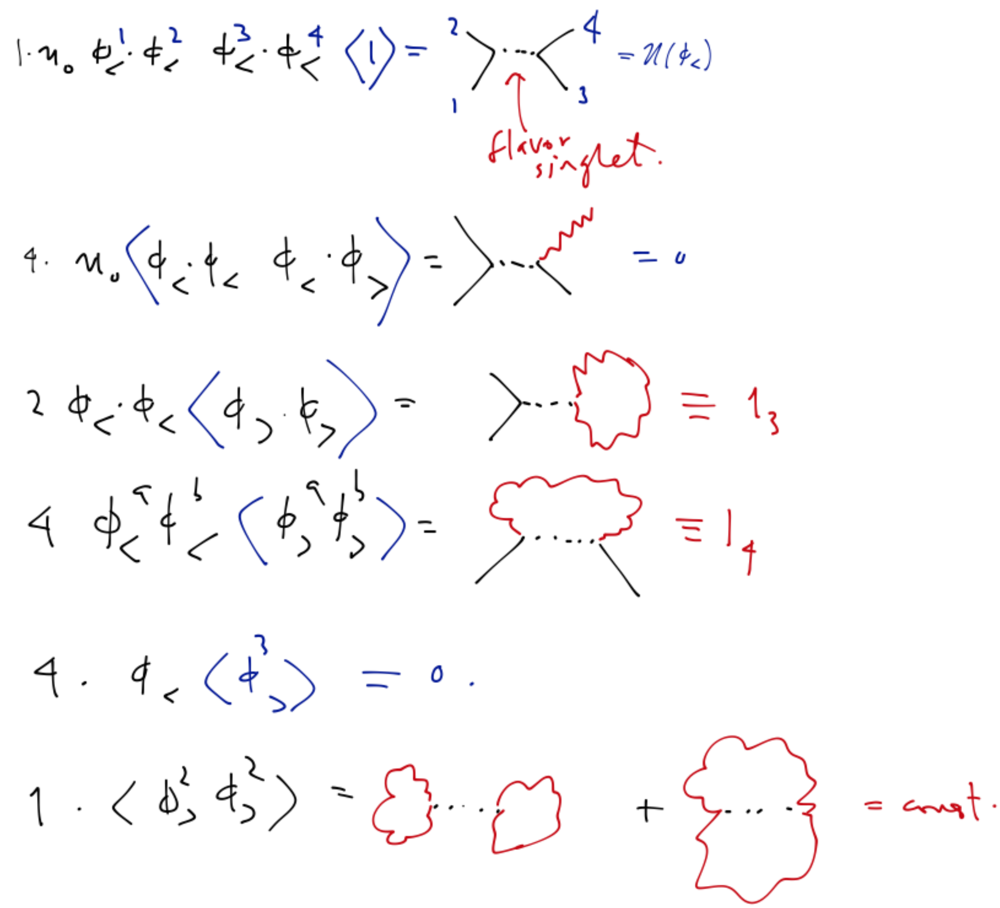



**Figure 7:** 1st order corrections from the quartic perturbation of the Gaussian fixed point of the O($N$) model. Wiggly lines denote propagation of fast modes $\phi_>$, straight lines denote (external) slow modes $\phi_<$. A further refinement of the notation is that we split apart the 4-point vertex to indicate how the flavor indices are contracted; the dotted line denotes a direction in which no flavor flows, i.e. it represents a coupling between the two flavor singlets, $\phi^a\phi^a$ and $\phi^b\phi^b$. The numbers at left are multiplicities with which these diagrams appear. (The relative factor of 2 between $1_3$ and $1_4$ can be understood as arising from the fact that $1_3$ has a symmetry which exchanges the fast lines but not the slow lines, while $1_4$ does not.) Notice that closed loops of the wiggly lines represent factors of $N$, since we must sum over which flavor is propagating in the loop – the flavor of a field running in a closed loop is not determined by the external lines, just like the momentum.

The interesting terms are

$$1_3 = -u_0 \underbrace{2}_{\text{symmetry}} \underbrace{N}_{=\delta^{aa}} \int_0^{\Lambda/s} \bar{d}^D k |\phi_<(k)|^2 \int_{\Lambda/s}^{\Lambda} \bar{d}^D q \frac{1}{r_0 + r_2 q^2}$$

$$1_4 = \frac{4 \cdot 1}{2 \cdot N} 1_3$$

has a bigger symmetry factor but no closed flavor index loop. The result through $\mathcal{O}(u)$ is then

$$\boxed{r_0 \to r_0 + \delta r_0 = r_0 + 4u_0(N+2) \int_{\Lambda/s}^{\Lambda} \bar{d}^D q \frac{1}{r_0 + r_2 q^2} + \mathcal{O}(u_0^2)} \, .$$

$r_2$ and $u$ are unchanged. RG step ingredients 2 (rescaling: $\tilde{q} \equiv sq$) and 3 (renormalizing: $\tilde{\phi} \equiv \zeta^{-1}\phi_<$) allow us to restore the original action; we can choose $\zeta = s^{1+D/2}$ to keep $\tilde{r}_2 = r_2$.

The second-order-in-$u_0$ terms are displayed in Fig. 8. The interesting part of the

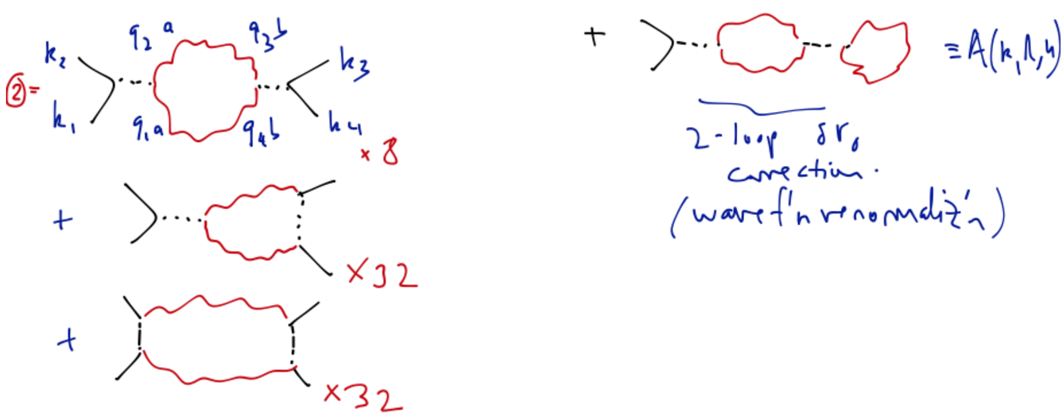



**Figure 8:** 2nd order corrections from the quartic perturbation of the Gaussian fixed point of the O($N$) model. Notice that the diagram at right has two closed flavor loops, and hence goes like $N^2$, and it comes with two powers of $u_0$. You can convince yourself by drawing some diagrams that this pattern continues at higher orders. If you wanted to define a model with large $N$ you should therefore consider taking a limit where $N \to \infty$, $u_0 \to 0$, holding $u_0 N$ fixed. The quantity $u_0 N$ is often called the 't Hooft coupling.

second order bit

$$2 = \frac{1}{2} \left\langle \mathcal{U}[\phi_<, \phi_>]^2 \right\rangle_{0,>,\text{connected}}$$

is the correction to $\mathcal{U}[\phi_<]$. There are less interesting bits which are zero or constant or two-loop corrections to the quadratic term. The correction to the quartic term at

2nd order is

$$\delta_2 S_4[\phi_<] = u_0^2(4N+32)\int_0^{\Lambda/s}\prod_i^4\left(đ^D k_i\phi_<(k_i)\right)\delta\!\!\!/(\sum k_i)f(k_1+k_2)$$

with

$$f(k_1+k_2) = \int đ^D q\frac{1}{(r_0+r_2q^2)(r_0+r_2(k_1+k_2-q)^2)} \simeq \int đ^D q\frac{1}{(r_0+r_2q^2)^2}\left(1+\mathcal{O}(k_1+k_2)\right)$$

– the bits that depend on the external momenta give irrelevant derivative corrections, like $\phi_<^2\partial^2\phi_<^2$ . We ignore them.

The full result through $\mathcal{O}(u_0^2)$ is then the original action, with the parameter replacement

$$\begin{pmatrix} r_2\\ r_0\\ u_0\end{pmatrix} \mapsto \begin{pmatrix} \tilde{r}_2\\ \tilde{r}_0\\ \tilde{u}_0\end{pmatrix} = \begin{pmatrix} s^{-D-2}\zeta^2(r_2+\delta r_2)\\ s^{-D}\zeta^2(r_0+\delta r_0)\\ s^{-3D}\zeta^4\left(u_0+\delta u_0\right)\end{pmatrix} + \mathcal{O}(u_0^3).$$

The shifts are:

$$\begin{cases} \delta r_2 = u_0^2\frac{\partial_k^2 A(0)}{r_2}\\ \delta r_0 = 4u_0(N+2)\int_{\Lambda/s}^{\Lambda}đ^D q\frac{1}{r_0+r_2q^2} - A(0)u_0^2\\ \delta u_0 = -\frac{1}{2}u_0^2(8N+64)\int_{\Lambda/s}^{\Lambda}đ^D q\frac{1}{(r_0+r_2q^2)^2}\end{cases}.$$

Here $A$ is the two-loop $\phi^2$ correction that we didn't compute (it contains the leading contribution to the wavefunction renormalization, $A(k) = A(0)+\frac{1}{2}k^2\partial_k^2A(0)+...$). We can choose to keep $\tilde{r}_2 = r_2$ by setting

$$\zeta^2 = \frac{s^{D+2}}{1+u_0^2\partial_k^2A(0)/r_2} = s^{D+2}\left(1+\mathcal{O}(u_0^2)\right).$$

Now let's make the RG step infinitesimal:

$$s = e^\ell \simeq 1+\delta\ell$$

$$\begin{cases} \frac{dr_0}{d\ell} = 2r_0 + \frac{4(N+2)K_D\Lambda^D}{r_0+r_2\Lambda^2}u_0 - Au_0^2 + \mathcal{O}(u_0^3)\\ \frac{du_0}{d\ell} = (4-D)u_0 - \frac{4(N+8)K_D\Lambda^D}{(r_0+r_2\Lambda^2)^2}u_0^2 + \mathcal{O}(u_0^3)\end{cases} \tag{9.18}$$

I defined $K_D \equiv \frac{\Omega_{D-1}}{(2\pi)^D}$.

---

To see how the previous thing arises, and how the integrals *all went away*, let's consider just the $\mathcal{O}(u_0)$ correction to the mass:

$$\tilde{r}_0 = r_0 + \delta\ell\frac{dr_0}{d\ell} = s^2\left(r_0 + 4u(N+2)\int_{\Lambda/s}^{\Lambda}\frac{đ^D q}{r_0+r_2q^2} + \mathcal{O}(u_0^2)\right)$$

$$= (1 + 2\delta\ell)\left(r_0 + 4u_0(N+2)\frac{\Omega_{D-1}}{(2\pi)^D}\Lambda^D\frac{1}{r_0 + r_2\Lambda^2}\delta\ell + \mathcal{O}(u_0^2)\right)$$
$$= \left(2r_0 + \frac{4u_0(N+2)}{r_0 + r_2\Lambda^2}K_D\Lambda^D\right)\delta\ell + \mathcal{O}(u_0^2). \qquad (9.19)$$

---

Now we are home. (9.18) has two fixed points. One is the free fixed point at the origin where nothing happens. The other (Wilson-Fisher) fixed point is at

$$\begin{cases} r_0^\star = -\frac{2u_0^\star(N+2)K_D\Lambda^D}{r_0^\star + r_2\Lambda^2} \overset{D=4-\epsilon}{=} -\frac{1}{2}\frac{N+2}{N+8}r_2\Lambda^2\epsilon + \mathcal{O}(\epsilon^2) \\ u_0^\star = \frac{(r^\star + r_2\Lambda^2)^2}{4(N+8)K_D\Lambda^D}\epsilon \overset{D=4-\epsilon}{=} \frac{1}{4}\frac{r_2^2}{(N+8)K_4}\epsilon + \mathcal{O}(\epsilon^2) \end{cases}$$

which is at positive $u_0^\star$ if $\epsilon > 0$. In the second step we keep only leading order in $\epsilon = 4 - D$.

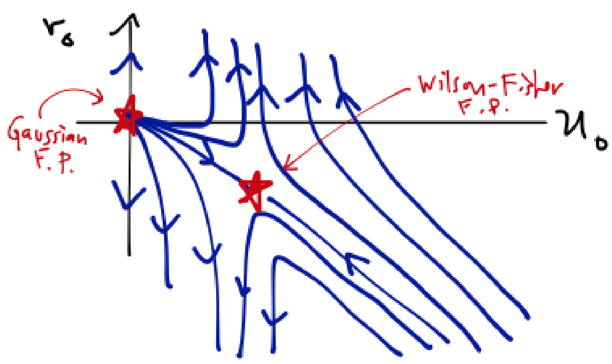



**Figure 9:** The $\phi^4$ phase diagram, for $\epsilon > 0$.

Now we follow useful strategies for dynamical systems and *linearize* near the W-F fixed point:

$$\frac{d}{d\ell}\begin{pmatrix}\delta r_0 \\ \delta u_0\end{pmatrix} = M\begin{pmatrix}\delta r_0 \\ \delta u_0\end{pmatrix}$$

The matrix $M$ is a 2x2 matrix whose eigenvalues describe the flows near the fixed point. It looks like

$$M = \begin{pmatrix} 2 - \frac{N+2}{N+8}\epsilon & \dots \\ \mathcal{O}(\epsilon^2) & -\epsilon \end{pmatrix}$$

Its eigenvalues (which don't care about the off-diagonal terms because the lower left entry is $\mathcal{O}(\epsilon^2)$ are

$$y_r = 2 - \frac{N+2}{N+8}\epsilon + \mathcal{O}(\epsilon^2) > 0$$

which determines the instability of the fixed point and

$$y_u = -\epsilon + \mathcal{O}(\epsilon^2) < 0 \quad \text{for } D < 4$$

which is a stable direction.

So $y_r$ determines the correlation length exponent. Its eigenvector is $\delta r_0$ to $\mathcal{O}(\epsilon^2)$. This makes sense: $r_0$ is the relevant coupling which must be tuned to stay at the critical point. The correlation length can be found as follows (as we did around Eq. (9.17)). $\xi$ is the value of $s = s_1$ at which the relevant operator has turned on by an order-1 amount, *i.e.* by setting $\xi \sim s_1$ when $1 \sim \delta r_0(s_1)$. According to the linearized RG equation, close to the fixed point, we have $\delta r_0(s) = s^{y_r}\delta r_0(0)$. Therefore

$$\xi \sim s_1^{-\frac{1}{y_r}} = (\delta r_0(0))^{-\nu} \ .$$

This last equality is the definition of the correlation length exponent (how does the correlation length scale with our deviation from the critical point $\delta r_0(0)$). Therefore

$$\nu = \frac{1}{y_r} = \left(2\left(1 - \frac{1}{2}\frac{N+2}{N+8}\epsilon\right)\right)^{-1} + \mathcal{O}(\epsilon^2) \simeq \frac{1}{2}\left(1 + \frac{N+2}{2(N+8)}\epsilon\right) + \mathcal{O}(\epsilon^2).$$

The remarkable success of setting $\epsilon = 1$ in this expansion to get answers for $D = 3$ continues. See the references for more details on this; for refinements of this estimate, see Zinn-Justin's book.

## 9.4 Which bits of the beta function are universal?

[Cardy, chapter 5] Some of the information in the beta functions depends on our choice of renormalization scheme and on our choice of regulator. Some of it does not: for example, the topology of the fixed points, and the critical exponents associated with them. Here is a way to see that some of the data in the beta functions is also universal. It also gives a more general point of view on the epsilon expansion and why it works.

**Operator product expansion (OPE).** Suppose we want to understand a (vacuum) correlation function of local operators like

$$\langle \phi_i(x_1)\phi_j(x_2)\Phi\rangle$$

where $\{\Phi\}$ is a collection of other local operators at $\{r_l\}$; suppose that the two operators we've picked out are closer to each other than to any of the others:

$$|r_1 - r_2| \ll |r_{1,2} - r_l|, \ \forall l.$$

Then from the point of view of the collection $\Phi$, $\phi_i\phi_j$ looks like a single local operator. But which one? Well, it looks like some sum over all of them:

$$\langle \phi_i(x_1)\phi_j(x_2)\Phi\rangle = \sum_k C_{ijk}(x_1 - x_2) \langle \phi_k(x_1)\Phi\rangle$$

where $\{\phi_k\}$ is some basis of local operators. By Taylor expanding we can move all the space-dependence of the operators to one point:

$$\phi(x_2) = e^{(x_2-x_1)^\mu \frac{\partial}{\partial x_1^\mu}}\phi(x_1) = \phi(x_1) + (x_2 - x_1)^\mu \partial_\mu \phi(x_1) + \cdots .$$

A shorthand for this collection of statements (for any $\Phi$) is the OPE

$$\phi_i(x_1)\phi_j(x_2) \sim \sum_k C_{ijk}(x_1 - x_2)\phi_k(x_1)$$

which is to be understand as an operator equation: true for all states, but only up to collisions with other operator insertions (hence the $\sim$ rather than $=$).

This is an attractive concept, but is useless unless we can find a good basis of local operators. At a fixed point of the RG, it becomes much more useful, because of scale invariance. This means that we can organize our operators according to their scaling dimension. Roughly it means two wonderful simplifications:

- We can find a basis (here, for the simple case of scalar operators)

$$\langle \phi_i(x)\phi_j(0)\rangle = \frac{\delta_{ij}}{r^{2\Delta_i}} \tag{9.20}$$

  where $\Delta_i$ is the scaling dimension of $\phi_i$. Then we can order the contributions to $\sum_k$ by increasing $\Delta_k$, which means smaller contributions to $\langle \phi\phi\Phi\rangle$.

- Further, the form of $C_{ijk}$ is fixed up to a number. Again for scalar operators,

$$\phi_i(x_1)\phi_j(x_2) \sim \sum_k \frac{c_{ijk}}{|x_1 - x_2|^{\Delta_i+\Delta_j-\Delta_k}}\phi_k(x_1) \tag{9.21}$$

  where $c_{ijk}$ is now a set of pure numbers, the *OPE coefficients* (or *structure constants*).

  The structure constants are universal data about the fixed point: they transcend perturbation theory. How do I know this? Because they can be computed from correlation functions of scaling operators *at the fixed point*: multiply the BHS of (9.21) by $\phi_k(x_3)$ and take the expectation value at the fixed point:

$$\begin{aligned}\langle\phi_i(x_1)\phi_j(x_2)\phi_k(x_3)\rangle_\star &= \sum_{k'} \frac{c_{ijk'}}{|x_1 - x_2|^{\Delta_i+\Delta_j-\Delta_k}}\langle\phi_{k'}(x_1)\phi_k(x_3)\rangle \\ &\overset{(9.20)}{=} \frac{c_{ijk}}{|x_1 - x_2|^{\Delta_i+\Delta_j-\Delta_k}}\frac{1}{|x_1 - x_3|^{2\Delta_k}}\end{aligned} \tag{9.22}$$

  (There is a better way to organize the RHS here, but let me not worry about that here.) The point here is that by evaluating the LHS at the fixed point, with some known positions $x_{1,2,3}$, we can extract $c_{ijk}$.

Confession: I (and Cardy) have used a tiny little extra assumption of *conformal invariance* to help constrain the situation here. It is difficult to have scale invariance without conformal invariance, so this is not a big loss of generality. We can say more about this next quarter but for now it is a distraction. [End of Lecture 37]

**Conformal perturbation theory.** I'll make this discussion in the Euclidean setting and we'll think about the equilibrium partition function

$$Z = \text{tr}e^{-H}$$

– we set the temperature equal to 1 and include it in the couplings.

Suppose we find a fixed point of the RG, $H_\star$. (For example, it could be the gaussian fixed point of $N$ scalar fields.) Let us study its neighborhood. (For example, we could seek out the nearby interacting Wilson-Fisher fixed point in $D < 4$ in this way.) Then

$$H = H_\star + \sum_x \sum_i g_i a^{\Delta_i}\phi_i(x)$$

where $a$ is the short distance cutoff (*e.g.* the lattice spacing), and $\phi_i$ has dimensions of length$^{-\Delta_i}$ as you can check from (9.20). So $g_i$ are de-dimensionalized couplings which we will treat as small and expand in. Then

$$Z = \underbrace{Z_\star}_{\equiv \text{tr}e^{-H^\star}} \left\langle e^{-\sum_x\sum_i g_i a^{\Delta_i}\phi_i(x)}\right\rangle_\star$$

$$\overset{\sum_x \simeq \frac{1}{a^D}\int d^D r}{\simeq} Z_\star \Bigg( 1 - \sum_i g_i \int \langle \phi_i(x) \rangle_\star \frac{d^D x}{a^{D-\Delta_i}}$$
$$+ \frac{1}{2} \sum_{ij} g_i g_j \int \frac{d^D x_1 d^D x_2}{a^{2D-\Delta_i-\Delta_j}} \langle \phi_i(x_1)\phi_j(x_2) \rangle_\star$$
$$- \frac{1}{3!} \sum_{ijk} g_i g_j g_k \int\int\int \frac{\prod_{a=1}^3 d^D x_a}{a^{3D-\Delta_i-\Delta_j-\Delta_k}} \langle \phi_i(x_1)\phi_j(x_2)\phi_k(x_3) \rangle_\star + ... \Bigg) .$$

Comments:

- We used the fact that near the fixed point, the correlation length is much larger than the lattice spacing to replace $\sum_x \simeq \frac{1}{a^D}\int d^D r$.
- There is still a UV cutoff on all the integrals – the operators can't get within a lattice spacing of each other: $|r_i - r_j| > a$.
- The integrals over space are also IR divergent; we cut this off by putting the whole story in a big box of size $L$. This is a physical size which should be RG-independent.
- The structure of this expansion does *not* require the initial fixed point to be a free fixed point; it merely requires us to be able to say something about the correlation functions. As we will see, the OPE structure constants $c_{ijk}$ are quite enough to learn something.

Now let's do the RG dance. While preserving $Z$, we make an infinitesimal change of the cutoff:

$$a \to sa = (1+\delta\ell)a, \delta l \ll 1 .$$

The price for preserving $Z$ is letting the couplings run $g_i = g_i(s)$. Where does $a$ appear?
(1) in the integration measure factors $a^{D-\Delta_i}$.
(2) in the cutoffs on $\int dx_1 dx_2$ which enforce $|x_1 - x_2| > a$.
(3) *not* in the IR cutoff.
The leading-in-$\delta\ell$ effects of (1) and (2) are additive and so may be considered separately:

$$(1) \qquad \tilde{g}_i = (1+\delta\ell)^{D-\Delta_i} g_i \simeq g_i + (D-\Delta_i) g_i \delta\ell \equiv g_i + \delta_1 g_i$$

The effect of (2) first appears in the $\mathcal{O}(g^2)$ term, the *change* in which is

$$(2) \qquad \sum_{i,j} g_i g_j \int_{|x_1-x_2|\in(a(1+\delta\ell),a)} \int \frac{d^D x_1 d^D x_2}{a^{2D-\Delta_i-\Delta_j}} \underbrace{\langle \phi_i(x_1)\phi_j(x_2) \rangle_\star}_{=\sum_k c_{ijk}|x_1-x_2|^{\Delta_k-\Delta_i-\Delta_j}\langle \phi_k \rangle_\star}$$
$$= \sum_{ij} g_i g_j c_{ijk} \Omega_{D-1} a^{-2D+\Delta_k} \langle \phi_k \rangle_\star$$

So this correction can be absorbed by a change in $g_k$ according to

$$\delta_2 g_k = -\frac{1}{2}\Omega_{D-1}\sum_{ij} c_{ijk} g_i g_j + \mathcal{O}(g^3)$$

where the $\mathcal{O}(g^3)$ term comes from triple collisions which we haven't considered here. Therefore we arrive at the following expression for evolution of couplings: $\frac{dg}{d\ell} = (\delta_1 g + \delta_2 g)/\delta\ell$

$$\frac{dg}{d\ell} = (D - \Delta_k) g_k - \frac{1}{2}\Omega_d \sum_{ij} c_{ijk} g_i g_j + \mathcal{O}(g^3) \ . \tag{9.23}$$

[37] At $g = 0$, the linearized solution is $dg_k/g_k = (D - \Delta_k)d\ell \implies g_k \sim e^{(D-\Delta_k)\ell}$ which reproduces our understanding of relevant and irrelevant at the initial fixed point.

Let's consider the Ising model.

$$\begin{aligned} H &= -\frac{1}{2}\sum_{x,x'} J(x-x')S(x)S(x') - h\sum_x S(x) \\ &\simeq -\frac{1}{2}\sum_{x,x'} J(x-x')S(x)S(x') - h\sum_x \phi(x) + \lambda \sum_x \left(S(x)^2 - 1\right)^2 \\ &\simeq \int d^D x \left( \frac{1}{2}\left(\vec{\nabla}\phi\right)^2 + r_0 a^{-2}\phi^2 + u_0 a^{D-4}\phi^4 + h a^{-1-D/2}\phi \right) \end{aligned} \tag{9.24}$$

In the first step I wrote a lattice model of spins $S = \pm 1$; in the second step I used the freedom imparted by universality to relax the $S = \pm 1$ constraint, and replace it with a potential which merely discourages other values of $S$; in the final step we took a continuum limit.

In (9.24) I've temporarily included a Zeeman-field term $hS$ which breaks the $\phi \to -\phi$ symmetry. Setting it to zero it stays zero (*i.e.* it will not be generated by the RG) because of the symmetry. This situation is called *technically natural.*

Now, consider for example as our starting fixed point the Gaussian fixed point, with

$$H_{\star,0} \propto \int d^D x \frac{1}{2}\left(\vec{\nabla}\phi\right)^2 \ .$$

[37] To make the preceding discussion we considered the partition function $Z$. If you look carefully you will see that in fact it was not really necessary to take the expectation values $\langle\rangle_\star$ to obtain the result (9.23). Because the OPE is an operator equation, we can just consider the running of the operator $e^{-H}$ and the calculation is identical. A reason you might consider doing this instead is that expectation values of scaling operators on the plane actually vanish $\langle\phi_i(x)\rangle_\star = 0$. However, if we consider the partition function in finite volume (say on a torus of side length $L$), then the expectation values of scaling operators are not zero. You can check these statements explicitly for the normal-ordered operators at the gaussian fixed point introduced below. Thanks to Sridip Pal for bringing these issues to my attention.

Since this is quadratic in $\phi$, all the correlation functions (and hence the OPEs, which we'll write below) are determined by Wick contractions using

$$\langle \phi(x_1)\phi(x_2)\rangle_{\star,0} = \frac{\mathcal{N}}{|x_1 - x_2|^{D-2}}.$$

It is convenient to rescale the couplings of the perturbing operators by $g_i \to \frac{2}{\Omega_{D-1}} g_i$ to remove the annoying $\Omega_{D-1}/2$ factor from the beta function equation. Then the RG equations (9.23) say

$$\begin{cases} \frac{dh}{d\ell} = & (1 + D/2) - \sum_{ij} c_{ijh} g_i g_j \\ \frac{dr_0}{d\ell} = & 2r_0 - \sum_{ij} c_{ijr_0} g_i g_j \\ \frac{du_0}{d\ell} = & \epsilon u_0 - \sum_{ij} c_{iju_0} g_i g_j \end{cases}$$

So we just need to know a few numbers, which we can compute by doing Wick contractions with free fields. That is: to find the beta function for $g_k$, we look at all the OPEs between operators in the perturbed hamiltonian (9.24) which produce $g_k$.

---

**Algebra of scaling operators at the Gaussian fixed point.** It is convenient to choose a basis of *normal-ordered* operators, which are defined by subtracting out their self-contractions. That is

$$\phi_n \equiv : \phi^n := \phi^n - \text{(self-contractions)}$$

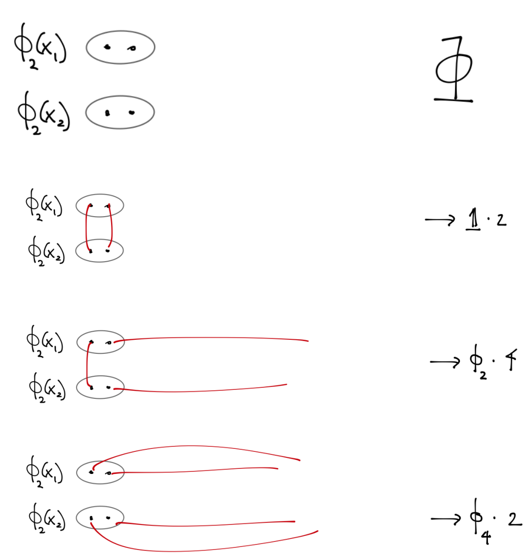
so that $\langle : \phi^n : \rangle = 0$, and specifically

$$\phi_2 = \phi^2 - \left\langle \phi^2 \right\rangle , \quad \phi_4 = \phi^4 - 3 \left\langle \phi^2 \right\rangle \phi^2 \ .$$

This amounts to a shift in couplings $r_0 \to r_0 + 3u \left\langle \phi^2 \right\rangle_\star$. The benefit of this choice of basis is that we can ignore any diagram where an operator is contracted with itself. Note that the contractions $\langle \phi^2 \rangle$ discussed here are defined on the plane. They are in fact quite UV sensitive and require some short-distance cutoff.

To compute their OPEs, we consider a correlator of the form above:s

$$\langle \phi_n(x_1)\phi_m(x_2)\Phi\rangle$$



We do wick contractions with the free propagator, but the form of the propagator doesn't matter for the beta function, only the combinatorial factors. If we can contract all the operators making up $\phi_n$

with those of $\phi_m$, then what's left looks like the identity operator to $\Phi$; that's the leading term, if it's there, since the identity has dimension 0, the lowest possible. More generally, some number of $\phi$s will be left over and will need to be contracted with bits of $\Phi$ to get a nonzero correlation function. For example, the contributions to $\phi_2 \cdot \phi_2$ are depicted at right.

The part of the result we'll need (if we set $h = 0$) can be written as (omitting the implied factors of $|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}$ necessary to restore dimensions):

$$\begin{cases} \phi_2\phi_2 & \sim 2\mathbb{1} + 4\phi_2 + \phi_4 + \cdots \\ \phi_2\phi_4 & \sim 12\phi_2 + 8\phi_4 + \cdots \\ \phi_4\phi_4 & \sim 24\mathbb{1} + 96\phi_2 + 72\phi_4 + \cdots \end{cases}$$

---

At $h = 0$, the result is (the $N = 1$ case of the result in §9.3.3)

$$\begin{cases} \frac{dr_0}{d\ell} = & 2r_0 - 4r_0^2 - 2 \cdot 12 r_0 u_0 - 96 u_0^2 \\ \frac{du_0}{d\ell} = & \epsilon u_0 - r_0^2 - 2 \cdot 8 r_0 u_0 - 72 u_0^2 \end{cases}$$

and so the ($N = 1$) WF fixed point occurs at $u_0 = u_0^\star = \epsilon/72, r_0 = \mathcal{O}(\epsilon^2)$.

Linearizing the RG flow about the *new* fixed point,

$$\frac{dr_0}{d\ell} = 2r_0 - 24u_0^\star r_0 + \cdots$$

gives

$$\frac{dr_0}{r_0} = (2 - \frac{24}{72}\epsilon) d\ell \quad \Longrightarrow \quad r_0 \sim e^{(2 - \frac{24}{72}\epsilon)\ell} \equiv \left(e^\ell\right)^{\frac{1}{\nu}}$$

which gives $\nu = \frac{1}{2} + \frac{1}{12}\epsilon + \mathcal{O}(\epsilon^2)$.

# 10 Gauge theory

## 10.1 Massive vector fields as gauge fields

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

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

where $(dB)_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$. 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. Obviously, any functional of $B$ is now invariant under the transformation

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

Notice that the fake new field $\theta$ transforms non-linearly (instead the 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. 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 $A_\mu$. One nice thing about this reshuffling is that the $m \to 0$ limit decouples the longitudinal bits.

Who is $\theta$? Our previous point of view was that it is fake and we can just choose the gauge parameter $\lambda$ 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:

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

where $\phi$ is a scalar field of charge $e$ whose covariant derivative is $D_\mu \phi = (\partial_\mu - \mathbf{i} e A_\mu)\,\phi$, 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

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

If we set $e = 1$, this differs from the action for $B$ written in terms of $A, \theta$ only in the addition of the *Higgs mode* $\rho$, whose mass (expanding $V(\rho)$ about $\rho = v$) is

$$m_\rho^2 = 4\kappa^2 v^2 \overset{\kappa \gg 1}{\gg} m_A^2 = \langle\rho\rangle^2 = v^2.$$

That is: in the limit of large $\kappa$, the excitations of $\rho$ are hard to make.

The description in terms of $\mathcal{L}_h$ 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). It is also a description of what happens to the EM field in a superconductor: it 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 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 consequence of the charge of $\phi$ for the flux quantization of vortices. We will say more about its origins in terms of electrons next quarter.

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) in a superconductor. If we try to insert it into a superconductor, it will trail behind it a vortex string. If we make the superconducting region larger and larger, its energy 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.

## 10.2 Festival of gauge invariance

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

$$\mathcal{L} = \sum_{\alpha=1}^{N} \partial_\mu \phi_\alpha^\star \partial^\mu \phi_\alpha - V(\phi_\alpha^\star \phi_\alpha) \tag{10.1}$$

(or $\mathcal{L} = \bar{\Psi}_\alpha \partial_\mu \Psi_\alpha$). The first term is just like the $\mathsf{O}(2N)$ generalization of the Wilson-Fisher theory, except that for kicks I grouped the scalars into pairs, and made the potential of the combination $\sum_{\alpha=1}^{N} \phi_\alpha^\star \phi_\alpha$.

**Lighting review of Lie groups and Lie algebras.** (10.1) is invariant under the $\mathsf{U}(N)$ transformation

$$\phi_\alpha \mapsto \Lambda_{\alpha\beta}\phi_\beta, \quad \Lambda^\dagger \Lambda = 1\!\!1. \tag{10.2}$$

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 $\mathrm{tr}\Lambda(\lambda^0 = 0) = 0$. Here the 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

$$[T^A, T^B] = \mathbf{i} f_{ABC} T^C \tag{10.3}$$

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 (10.2), with $\Lambda$ close to the identity, is

$$\phi_\alpha \mapsto \phi_\alpha + \mathbf{i}\lambda^A T^A_{ab} \phi_B.$$

Other representations of the group come from other sets of $T^A_R$s which satisfy the same algebra (10.3), but can have other dimensions. For example, the structure constants themselves $\left(T^B_{\mathrm{adj}}\right)_{AC} \equiv -\mathbf{i} f_{ABC}$ 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 (10.1). Let's consider how we might change the model in (10.1) to make it invariant under a *local* transformation, with $\lambda = \lambda(x)$. In the Abelian case, we have learned

$$\phi \mapsto e^{\mathbf{i}\lambda(x)}\phi(x), A_\mu \mapsto A_\mu + \partial_\mu \lambda, \quad \partial_\mu \phi \rightsquigarrow D_\mu \phi = (\partial_\mu - \mathbf{i}A_\mu)\phi \mapsto e^{\mathbf{i}\lambda(x)} D_\mu \phi.$$

By replacing partial derivatives with covariant derivatives, we can make gauge-invariant Lagrangians. The same thing works in the non-abelian case:

$$(D_\mu \phi)_\alpha \equiv \partial_\mu \phi_\alpha - \mathbf{i}A^A T^A_{\alpha\beta}\phi_\beta$$

$$\phi \mapsto \phi + \mathbf{i}\lambda^A(x) T^A \phi, \quad A^A_\mu \mapsto A^A_\mu + \partial_\mu \lambda^A - f_{ABC}\lambda^B A^C_\mu(x).$$

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:

$$\mathcal{L}_{YM} = -\frac{1}{4g^2}\sum_A \left( \underbrace{\partial_\mu A^A_\nu - \partial_\nu A^A_\mu + f_{ABC} A^B_\mu A^C_\nu}_{=F^A_{\mu\nu} = -F^A_{\nu\mu}} \right)^2 = -\frac{1}{4g^2}\mathrm{tr} F_{\mu\nu}F^{\mu\nu}. \qquad (10.4)$$

The field strength

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} + f_{ABC}\lambda^B F^C_{\mu\nu} = F^A_{\mu\nu} + \mathbf{i}\lambda^B \left(T^B_{\text{adj}}\right)_{AC} F^C_{\mu\nu}$$

is designed so that it transforms in the adjoint representation, and therefore $S_{YM}$ is gauge-invariant.

[End of Lecture 38]

**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 $W_{xy} \mapsto \Lambda(x)^\dagger W_{xy}\Lambda(y)$, 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

$$D_\mu \phi(x) \equiv \lim_{\Delta x \to 0} \frac{W(x, x+\Delta x)\phi(x+\Delta x) - \phi(x)}{\Delta x^\mu} \mapsto U(x) D_\mu \phi(x) \ . \qquad (10.5)$$

Expanding near $\Delta x \to 0$, we can let

$$W(x, x+\Delta x) = \mathbb{1} - \mathbf{i}e\Delta x^\mu A_\mu(x) + \mathcal{O}(\Delta x^2) \tag{10.6}$$

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 $D_\mu^{A^\Lambda}(\Lambda\phi) \stackrel{!}{=} \Lambda\left(D_\mu^A\phi\right)$ 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 (10.6) 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.

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. [38] 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 $[D_\mu, D_\nu] = -\mathbf{i}eF_{\mu\nu}$. 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: $F_{\mu\nu} = F_{\mu\nu}^A T^A$, so more explicitly,

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - \mathbf{i}ef_{ABC}A_\mu^B A_\nu^C.$$

[38] Which 2d surface? Let me speak about the abelian case for the rest of this footnote. The difference in phase between two possible choices is $e^{\mathbf{i}e\int_{R-R'} F} \stackrel{\text{Stokes}}{=} e^{\mathbf{i}e\int_V dF}$ 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 an identity if $F = dA$. You might think this prevents magnetic sources. But actually $\int_V dF$ only appears in $e^{\int_V dF}$, 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.

Since it is made from products of covariant derivatives, $[D, D]\phi \mapsto \Lambda[D, D]\phi$, it must transform in the adjoint representation, $F \mapsto \Lambda^\dagger F \Lambda$, or in infinitesimal form

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} - f^{ABC}\lambda^B F^C_{\mu\nu}.$$

**Actions for gauge fields.** The Yang-Mills (YM) action (10.4) 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 $f_{ABC}$.

In even spacetime dimensions, another gauge invariant, Lorentz invariant local functional of $A$ is the total-derivative term $S_\theta = \theta \int \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $D/2$ factors of $F$. (By $\wedge$ I mean antisymmetrize all the indices.) This doesn't affect the equations of motion or perturbation theory (*e.g.* in $D = 4$, $\epsilon^{\mu\nu\rho\sigma} F^A_{\mu\nu} F^A_{\rho\sigma} = 2\partial_\mu \left(\epsilon^{\mu\nu\rho\sigma} A^A_\nu F^A_{\rho\sigma}\right)$) but does matter non-perturbatively. We'll see next quarter 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.

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 $F$. 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 $D_\mu \psi = \left(\partial_\mu - \mathbf{i} T^A_R A^A_\mu\right)\psi$ also transforms in representation $R$, so

$$\bar{\psi} D_\mu \psi + V(\bar{\psi}\psi)$$

is a gauge-invariant lagrangian density. The lowest-dimension couplings of $A$ to matter are determined by the representation matrices $T^A_R$, 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. In the small time left in this quarter, we will instead think about how we might define the thing non-perturbatively and see what we learn from that.

## 10.3 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 $U_{xy} \mapsto g_x^\dagger U_{xy} g_y$, 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 $\int_0^{2\pi} d\varphi$; 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-invariant 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:

$$S[U] = \frac{1}{2f^2} \sum_{\Box} S_\Box, \quad S_\Box = \mathrm{Re}\,\mathrm{tr} \prod_{\ell \in \partial \Box} U = \mathrm{Re}\,\mathrm{tr} \left( U_{x,x+dx} U_{x+dx,x+dx+dy} U_{x+dx+dy,x+dy} U_{x+dy,x} \right).$$

Now let's think seriously about the $\mathsf{G} = \mathsf{SU}(N)$ 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[39] 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 $a$,

$$\begin{aligned} S_{\Box_{\mu\nu}} &\overset{\mathrm{CBH}}{=} \mathrm{Re}\,\mathrm{tr} \left( e^{\mathbf{i} \frac{a^2}{\sqrt{2} g^2} F_{\mu\nu}} + \mathcal{O}(a^3) \right) \\ &= \mathrm{Re}\,\mathrm{tr} \left( 1\!\!1 + \mathbf{i} \frac{a^2}{\sqrt{2} g^2} F_{\mu\nu} - \frac{1}{2} \frac{a^4}{4 g^2} F_{\mu\nu} F_{\mu\nu} + \mathcal{O}(a^5) \right) \\ &= \mathrm{tr} 1\!\!1 - \frac{a^4}{4 g^4} \mathrm{tr} F_{\mu\nu} F_{\mu\nu} + \ldots = \mathcal{L}_{YM}(\Box) + \mathrm{const.} \end{aligned}$$

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

$$S_q = \frac{1}{a^\#} \sum_{x,\ell} q_x^\dagger U_{x,x+\ell} q_{x+\ell} \simeq \int d^D x \; q^\dagger(x) \left( \not{D} - m \right) q(x) + \ldots$$

where $D_\mu = \partial_\mu - \mathbf{i} A_\mu$ is the covariant derivative, and we used its definition (10.5). 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

[39] Charlie-Baker-Hotel

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:

$$\lim_{T\to\infty} Z^{-1} \int DA\ e^{\mathbf{i}S_{\mathrm{Maxwell}}[A] + \mathbf{i}\int A_\mu J^\mu} = e^{-\mathbf{i}V(R)T}.$$

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

$$\langle W(C_{R\times T})\rangle \overset{T \gg R}{\simeq} e^{-\mathbf{i}V(R)T},$$

where the LHS is 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 $\log\left\langle e^{\mathbf{i}\oint_{C_{R\times T}} A}\right\rangle \simeq -\mathbf{i}E(R)T + f(T)R$ 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

$$\langle W(C_{R\times T})\rangle = Z^{-1} \int \prod dU\ e^{-\frac{1}{2f^2}\sum_\Box \mathrm{Re}\, S_\Box} W(C_{R\times T}) \overset{T \gg R}{\simeq} e^{-\mathbf{i}V(R)T}$$

with

$$V(R) = \sigma R \quad \Longrightarrow \quad F = -\frac{\partial V}{\partial R} = -\sigma\ .$$

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 abelian 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 over 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.[40]

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

[40]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.

structure (though different large-$\beta$ (weak coupling) physics). So take $U_\ell = e^{\mathbf{i}\theta_\ell} \in \mathsf{U}(1)$, in which case

$$S_{\Box_{\mu\nu}}[U] = -\left(1 - \cos\theta_{\mu\nu}\right), \ \ \theta_{\mu\nu}(x) = \theta_\mu(x+\nu) - \theta_\mu(x) - \theta_\nu(x+\mu) + \theta_\nu(x) \equiv \Delta_\nu\theta_\mu - \Delta_\mu\theta_\nu(x).$$

First let's consider the case where the world is a single plaquette. Then

$$\begin{aligned} \langle W(\Box)\rangle &= \int \prod_\ell dU_\ell \ U_1U_2U_3U_4 \left(1 + \beta(S_\Box + S_\Box^\dagger) + \frac{1}{2}\beta^2\left(S + S^\dagger\right)^2 + \frac{1}{3!}\beta^3\left(S + S^\dagger\right)^3 + \cdots\right) \\ &= \beta \underbrace{\langle S_\Box S_{-\Box}\rangle}_{=1} + \frac{\beta^3}{2}\langle S_{2\Box}S_{-2\Box}\rangle + \mathcal{O}(\beta^5) = \beta^{A(\Box)}\left(1 + \mathcal{O}(\beta^2)\right) = e^{-f(\beta)\text{Area}} \end{aligned} \tag{10.7}$$

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.

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-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 $4t+2$, 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:

$$\partial^\mu F_{\mu\nu} = J_\nu^{(e)}, \partial^\mu \tilde{F}_{\mu\nu} = J_\nu^{(m)} \tag{10.8}$$

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 (10.8) by introducing a smooth vector potential $A_\mu$ via

$$F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu + \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\int d^4y J^{(m)}(y)^\sigma f^\rho(x-y)$$

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 an long and infinitely thin solenoid, which is invisible classically. Quantumly, it could be detected by Aharonov-Bohm effect of a charged particle going around it $e^{\mathbf{i}\int B} = e^{\mathbf{i}\oint A} = e^{\mathbf{i}eg}$ unless $eg \in 2\pi\mathbb{Z}$, Dirac quantization again. (For particles with electric and magnetic charge (dyons), the condition is $q_1m_2 - q_2m_1 \in 2\pi\mathbb{Z}$.)

So, the duality interchanges $e$ and $m$. 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 our friend the XY model. I hope we will come back to this next quarter.