

Physics 215B: Particles and Fields

Winter 2020

Lecturer: McGreevy

These lecture notes live [here](#). Please email corrections to mcgreevy at physics dot
ucsd dot edu.

Last updated: 2020/03/18, 14:52:58

Contents

0.1	Introductory remarks	3
0.2	Sources and acknowledgement	6
0.3	Conventions	7
1	To infinity and beyond	8
1.1	A parable from quantum mechanics on the breaking of scale invariance	8
1.2	Casimir effect: vacuum energy is real	16
1.3	A simple example of perturbative renormalization in QFT	20
1.4	Classical interlude: Mott formula	23
1.5	Electron self-energy in QED	28
1.6	Big picture interlude	34
1.7	Vertex correction in QED	43
1.8	Vacuum polarization	55
2	Consequences of unitarity	65
2.1	Spectral density	65
2.2	Cutting rules and optical theorem	72
2.3	How to study hadrons with perturbative QCD	79
3	Wilsonian perspective on renormalization	82
3.1	A parable on integrating out degrees of freedom	82
3.2	Where do field theories come from?	88
3.3	The continuum version of blocking	96
3.4	An extended example: XY model	99
3.5	The operator product expansion and conformal perturbation theory . .	123
4	Effective field theory	130
4.1	Introduction to effective field theory	130
4.2	The color of the sky	135
4.3	Fermi theory of Weak Interactions	137
4.4	Loops in EFT	138
4.5	The Anderson-Higgs Mechanism and Superconductors	144
4.6	Effective field theory of Fermi surfaces	155
4.7	The Standard Model as an EFT.	166
5	Gauge theory	170

0.1 Introductory remarks

Quantum field theory (QFT) is the quantum mechanics of *extensive degrees of freedom*. What I mean by this is that at each point of space, there's some stuff that can wiggle.

'Extensive degrees of freedom' are those which, if we like, we can sprinkle over vast tracts of land, like *sod* (depicted in the figure at right). And also like sod, each little patch of degrees of freedom only interacts with its neighboring patches: this property of sod and of QFT is called *locality*.



More precisely, in a quantum mechanical system, we specify the degrees of freedom by their Hilbert space; by an extensive system, I'll mean one in which the Hilbert space is of the form $\mathcal{H} = \otimes_{\text{patches of space}} \mathcal{H}_{\text{patch}}$ and the interactions are local $\mathbf{H} = \sum_{\text{patches}} \mathbf{H}(\text{nearby patches})$.¹

It's not surprising that QFT is so useful, since this situation happens all over the place. Some examples of 'stuff' are: the atoms in a solid, or the electrons in those atoms, or the spins of those electrons. A less obvious, but more visible, example is the electromagnetic field, even in vacuum. More examples are provided by other excitations of the vacuum, and it will be our job here to understand those very electrons and atoms that make up a solid in these terms. The vacuum has other less-long-lasting excitations which are described by the Standard Model of particle physics.

Some examples of QFT are Lorentz invariant ('relativistic'). That's a nice simplification when it happens. Indeed this seems to happen in particle physics. We're going to focus on this case for much of this quarter. Still I would like to emphasize: though some of the most successful applications of QFT are in the domain of high energy particle physics, this is not a class on that subject, and I will look for opportunities to emphasize the universality of QFT.

Last quarter you saw that the low-energy excitations of weakly-coupled fields are *particles*. A consequence of relativity is that the number of particles isn't fixed. That is: there are processes where the number of particles changes in time. Sometimes this is used to motivate the study of QFT. It's a necessary consequence of Lorentz symmetry, but the converse is false: particle production and destruction can happen without relativity.

¹Actually, the Hilbert space of a gauge theory is *not* of this form; rather, it is a subspace of such a space which satisfies the Gauss law. This is a source of a lot of confusion, which I hope to dispel.

‘Divergences’. Another intrinsic and famous feature of QFT discernible from the definition I gave above is its flirtation with infinity. I said that there is ‘stuff at each point of space’; how much stuff is that? Well, there are two senses in which ‘the number of points of space’ is infinite: (1) space can go on forever (the infrared (IR)), and (2) in the continuum, in between any two points of space are more points (the ultraviolet (UV)). The former may be familiar from statistical mechanics, where it is associated with the thermodynamic limit, which is where interesting things happen. For our own safety, our discussion will take place in a padded room, protected on both sides from the terrors of the infinite.

Last quarter 215A ended right when you learned to compute amplitudes for simple processes in QED. There are many measurable quantities that can be computed using the formalism you developed, and the success of leading-order QED is a real high point of physics. I’ll have to say a few words about that success.

Lurking behind that success, however, is a dark cloud. The leading order of perturbation theory is given by tree diagrams; more complicated diagrams should be suppressed by more powers of $\frac{e^2}{4\pi} \equiv \alpha$. You might think that if the leading calculation worked so well, we should do even better by looking at the next term. But there is a surprise: naively calculating the next term gives an infinite correction.

So an important job will be to explain the non-naive point of view on this calculation which allows us to extract finite, meaningful answers from perturbation theory.

Even more important will be to give a better point of view, from which we never would have encountered divergences in the first place. The (silly) name for this point of view is the renormalization group.

So here are some goals for the quarter, both practical and philosophical:

- There is more to QFT than the S-matrix. In a particle-physics QFT course (like 215 so far) you learn that the purpose in life of correlation functions or green’s functions or off-shell amplitudes is that they have *poles* (at $p^\mu p_\mu - m^2 = 0$) whose residues are the S-matrix elements, which are what you measure (or better, are the distribution you sample) when you scatter the particles which are the quanta of the fields of the QFT. I want to make two extended points about this:
 1. In many physical contexts where QFT is relevant, you can actually measure the off-shell stuff. This is yet another reason why including condensed matter in our field of view will deepen our understanding of QFT.
 2. This is good, because the Green’s functions don’t always have simple poles! There are lots of interesting field theories where the Green’s functions instead have power-law singularities, like $G(p) \sim \frac{1}{p^{2\Delta}}$. If you Fourier trans-

form this, you don't get an exponentially-localized packet. The elementary excitations created by a field whose two point function does this *are not particles*. (Any conformal field theory (CFT) is an example of this.) The theory of particles (and their dance of creation and annihilation and so on) is an important but proper subset of QFT.

- In addition to its importance in high energy physics, I want to emphasize that QFT is also quite central in many aspects of condensed matter physics. From the point of view of someone interested in QFT, high energy particle physics has the severe drawback that it offers only one example! (OK, for some purposes we can think about QCD and the electroweak theory separately...)

From the high-energy physics point of view, we could call this the study of *regulated* QFT, with a particular kind of lattice regulator. Why make a big deal about 'regulated'? Besides the fact that this is how QFT comes to us (when it does) in condensed matter physics, such a description is required if we want to know what we're talking about. For example, we need it if we want to know what we're talking about well enough to explain it to a computer. Many QFT problems are too hard for our brains. A related but less precise point is that I would like to do what I can to erase the problematic, theorist-centered perspective on QFT which 'begins from a classical lagrangian and quantizes it'.

- A central theme this quarter will be coarse-graining in quantum systems with extensive degrees of freedom, aka the renormalization group (RG) in QFT.

By 'coarse-graining' I mean ignoring things we don't care about, or rather only paying attention to them to the extent that they affect the things we do care about. And the things we should care about the most are the biggest ones – the modes with the longest wavelength. So the 'better perspective' alluded to above is: we should try to understand QFT scale by scale.

To continue the sod example in 2+1 dimensions, a person laying the sod in the picture above cares that the sod doesn't fall apart, and rolls nicely onto the ground (as long as we don't do high-energy probes like bending it violently or trying to lay it down too quickly). These *long-wavelength* properties of *rigidity* and *elasticity* are collective, emergent properties of the microscopic constituents (sod molecules) – we can describe the dynamics involved in covering the Earth with sod (never mind whether this is a good idea in a desert climate) without knowing the microscopic theory of the sod molecules ('grass'). Our job is to think about the relationship between the microscopic model (grassodynamics) and its macroscopic counterpart (in this case, suburban landscaping). In my experience, learning to do this is approximately synonymous with *understanding*.

- I would like to convince you that “non-renormalizable” does not mean “not worth your attention,” and explain the incredibly useful notion of an Effective Field Theory.
- There is more to QFT than perturbation theory about free fields in a Fock vacuum. In particular, it is worthwhile to spend some time thinking about non-perturbative physics, effects of topology, solitons. Topology is one tool for making precise statements without perturbation theory (the basic idea: if we know something is an integer, it is easy to get many digits of precision!). Maybe this will wait until 215C.
- Given time, I would like to show that many fancy phenomena precious to particle physicists can emerge from humble origins in the kinds of (completely well-defined) local quantum lattice models we will study. Here I have in mind: fermions, gauge theory, photons, anyons, strings, topological solitons, CFT, and many other sources of wonder I’m forgetting right now.

0.2 Sources and acknowledgement

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.

Xiao-Gang Wen, *Quantum Field Theory of Many-Body Systems*

Sidney Coleman, *Aspects of Symmetry*

Alexander Polyakov, *Gauge Fields and Strings*

Eduardo Fradkin, *Field Theories of Condensed Matter Systems*

0.3 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 $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = (\partial_t, \vec{\nabla}_x)^\mu$, and $\partial^\mu \equiv \eta^{\mu\nu} \partial_\nu$. I'll use μ, ν, \dots for Lorentz indices, and i, j, k, \dots for spatial indices.

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

d is the number of space dimensions, D is the number of spacetime dimensions (it's bigger!).

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 $d\tilde{k} \equiv \frac{d^d k}{(2\pi)^d}$. I will also write $\delta^d(q) \equiv (2\pi)^d \delta^{(d)}(q)$. I will try to be consistent about writing Fourier transforms as

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

IFF \equiv if and only if.

RHS \equiv right-hand side. LHS \equiv left-hand side. BHS \equiv both-hand side.

IBP \equiv integration by parts. WLOG \equiv without loss of generality.

$+\mathcal{O}(x^n) \equiv$ plus terms which go like x^n (and higher powers) when x is small.

$+h.c. \equiv$ plus hermitian conjugate.

$\mathcal{L} \ni \mathcal{O}$ means the object \mathcal{L} contains the term \mathcal{O} .

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

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

1 To infinity and beyond

At this point I believe you are able to use QED to compute the amplitudes and cross-sections for many physical processes involving electrons, photons and positrons. More precisely, you know how to compute the leading-order-in- α amplitudes, using Feynman diagrams which are trees – diagrams without loops. The natural next step is to look at the next terms in the perturbation expansion in α , which come from diagrams with one loop. This innocent-seeming step opens a big can of worms. The reason this is a big step is that tree-level processes are actually classical.

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. Surely we have no information yet about these modes from our piddling low-energy experiments. (Perhaps this is an opportunity to learn about them?) What do we do?

In order to put ourselves in the right frame of mind to think about this stuff, we'll make a brief retreat to a parable about a system with finitely many degrees of freedom in §1.1. A second useful parable in §1.2 will come from the Casimir effect – the additive constant in the energy of free fields raises the same issues. Then we'll apply these lessons to a simple field theory example, namely scalar field theory. Then we'll come back to perturbation theory in QED. Reading assignment for this chapter: Zee §III.

1.1 A parable from quantum mechanics on the breaking of scale invariance

Recall that the coupling constant in ϕ^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)² with such (classical) scale invariance. It exhibits many interesting features that can happen in strongly interacting quantum

²I learned this example from Marty Halpern.

field theory – asymptotic freedom, dimensional transmutation. Because the model is simple, we can understand these phenomena without resort to perturbation theory. They will nevertheless illuminate some ways of thinking which we’ll need in examples where perturbing 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 , $q \rightarrow \sqrt{m}q$.)

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

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

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

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

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

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

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

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 when $d = 2$ 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

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

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

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

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:

$$\psi(\vec{q}) \equiv \int \frac{d^2p}{(2\pi\hbar)^2} e^{i\vec{p}\cdot\vec{q}/\hbar} \varphi(\vec{p}) \equiv \int d^2p e^{i\vec{p}\cdot\vec{q}} \varphi(\vec{p}) \quad (1.1)$$

where I used some notation from §0.3 and set $\hbar = 1$. The delta function is

$$\delta^{(2)}(q) = \int \frac{d^2p}{(2\pi\hbar)^2} e^{i\vec{p}\cdot\vec{q}/\hbar}. \quad (1.2)$$

So the Schrödinger equation says

$$\begin{aligned} \left(-\frac{1}{2}\nabla^2 - E\right)\psi(q) &= -V(q)\psi(q) \\ \int d^2p e^{i\vec{p}\cdot\vec{q}} \left(\frac{p^2}{2} - E\right)\varphi(p) &= +g_0\delta^2(q)\psi(0) \\ &\stackrel{(1.2)}{=} +g_0 \left(\int d^2p e^{i\vec{p}\cdot\vec{q}}\right)\psi(0) \end{aligned} \quad (1.3)$$

where to get to the second line, we just plugged in (1.1). Integrating the both-hand side of (1.3) over q : $(\int d^2q e^{-i\vec{p}\cdot\vec{q}} ((1.3)))$ 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:

- $\psi(\vec{q} = 0) = \int d^2p \varphi(\vec{p}) = 0$. 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.* $\psi \propto \sin \frac{p_x x}{\hbar} e^{\pm i p_y y / \hbar}$. 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 :

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

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

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

What happens if we integrate this $\int \mathrm{d}^2 p$ to check self-consistency – the LHS should give α again:

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

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

is a condition on the energy ϵ_B of possible boundstates.

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

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

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

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

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

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

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

More precisely:

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

So in our cutoff theory, the boundstate condition is:

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

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 ϵ_B .

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} . \tag{1.4}$$

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

Suppose we insist that the boundstate energy ϵ_B is a fixed thing – imagine we've *measured* it to be 200 MeV³. We should express everything in terms of the measured quantity. Then, given some cutoff Λ , 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 ϵ_B). This phenomenon is called *dimensional transmutation* (d.t.). The cutoff is supposed to go away in observables, which depend on ϵ_B instead.

In QCD we expect that in an identical way, an arbitrary scale Λ_{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, Λ_{QCD} . Nature chooses $\Lambda_{QCD} \simeq 200$ MeV.

³Spoiler 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 D* **7** (1973) 1898. Maybe we'll come back to their example.]

There are more lessons in this example. Go back to (1.4):

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

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 ϵ_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.

Notice furthermore that even for moderately weak coupling, $\epsilon_B \ll \Lambda^2$. For example, when $g_0 = .1$, $\epsilon_B/\Lambda^2 = 10^{-28}$! Dimensional transmutation generates a hierarchy of scales. This phenomenon explains why the critical temperature below which metals go superconducting is much less than their intrinsic energy scale (the Fermi energy).

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

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

– the bare coupling vanishes in this limit, since we are insisting that the parameter ϵ_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.

[End of Lecture 1]

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

$$\text{Def:} \quad \beta(g_0) \equiv \Lambda \frac{\partial}{\partial \Lambda} g_0(\Lambda) .$$

For this theory

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

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

we can get good results near the fixed point from the perturbative part of β . That is: we can reliably 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.

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.

1.2 Casimir effect: vacuum energy is real

[A. Zee, *Quantum Field Theory in a Nutshell*, §I.9]

Our success in the last subsection relied on our ability to completely solve the theory, which we could do because there was just a single degree of freedom. Now we will be brave and think about a system with extensive degrees of freedom.

This subsection has two purposes. The main purpose is to give an object lesson in asking the right questions. In physics, the right question is often a question which can be answered by an experiment, at least in principle. The answers to such questions are less sensitive to our silly theoretical prejudices, *e.g.* about what happens to physics at very short distances.

A second purpose is to show that the $\frac{1}{2}\hbar\omega$ energy of the vacuum of the quantum harmonic oscillator is real. Sometimes we can get rid of it by choosing the zero of energy (which doesn't matter unless we are studying dynamical gravity). But it is meaningful if we can *vary* ω (or the collection of ω s if we have many oscillators as for the radiation field) and compare the difference.

In the context of the bunch of oscillators making up the radiation field, we can change the spectrum of frequencies of these oscillators $\{\omega_k\}$ by putting it in a box and varying the size of the box. In particular, two parallel conducting plates separated by some distance d experience an attractive force from the change in the vacuum energy of the EM field resulting from their presence. The plates put boundary conditions on the field, and therefore on which normal modes are present.

To avoid some complications of E&M which are not essential for our point here, we're going to make two simplifications:

- we're going to solve the problem in 1+1 dimensions
- and we're going to solve it for a scalar field.

To avoid the problem of changing the boundary conditions *outside* the plates we use the following device with *three* plates:

$$| \leftarrow d \rightarrow | \leftarrow \quad L - d \quad \longrightarrow |$$

(We will consider $L \gg d$, so we don't really care about the far right plate.) The 'perfectly conducting' plates impose the boundary condition that our scalar field $q(x)$ vanishes there. The normal modes of the scalar field $q(x)$ in the left cavity are then

$$q_j = \sin(j\pi x/d), \quad j = 1, 2, \dots$$

with frequencies $\omega_j = \frac{\pi|j|}{d}c$. There is a similar expression for the modes in the right cavity which we won't need. We're going to add up all the $\frac{1}{2}\hbar\omega_s$ for all the modes in both cavities to get the vacuum energy $E_0(d)$; the force on the middle plate is then $-\partial_d E_0$.

The vacuum energy in the whole region of interest between the outer plates is the sum of the vacuum energies of the two cavities

$$E_0(d) = f(d) + f(L - d)$$

where

$$f(d) = \frac{1}{2}\hbar \sum_{j=1}^{\infty} \omega_j = \hbar c \frac{\pi}{2d} \sum_{j=1}^{\infty} j \stackrel{?!?!}{=} \infty.$$

We have done something wrong. What?

Our crime is hubris: we assumed that we knew what the modes of arbitrarily large mode number k (arbitrarily short wavelength, arbitrarily high frequency) are doing, and in particular we assumed that they cared about our silly plates. In fact, no metal in existence can put boundary conditions on the modes of large enough frequency – those modes don't care about d . The reason a conductor puts boundary conditions on the EM field is that the electrons move around to compensate for an applied field, but there is a limit on how fast the electrons can move (*e.g.* the speed of light). The resulting cutoff frequency is called the *plasma frequency* but we don't actually need to know about all these details. To parametrize our ignorance of what the high-frequency modes do, we must *cut off* (or *regularize*) the contribution of the high-frequency modes. Let's call modes with $\omega_j \gg \pi/a$ high frequency, where a is some short time⁵, $a \ll d$. Replace

$$\begin{aligned} f(d) \rightsquigarrow f(a, d) &= \hbar \frac{\pi}{2d} \sum_{j=1}^{\infty} e^{-a\omega_j/\pi} j \\ &= -\frac{\pi\hbar}{2} \partial_a \left(\underbrace{\sum_{j=1}^{\infty} e^{-aj/d}}_{= \frac{1}{1-e^{-a/d}} - 1} \right) \\ &= +\frac{\pi\hbar}{2d} \frac{e^{a/d}}{(e^{a/d} - 1)^2} \\ &\stackrel{a \ll d}{\approx} \hbar \left(\underbrace{\frac{\pi d}{2a^2}}_{\rightarrow \infty \text{ as } a \rightarrow 0} - \frac{\pi}{24d} + \frac{\pi a^2}{480d^3} + \dots \right) \end{aligned} \quad (1.5)$$

⁵You can think of a as the time it takes the waves to move by one lattice spacing. If we work in units where the velocity is $c = 1$, this is just the lattice spacing. I will do so for the rest of this discussion.

Answers which don't depend on a have a chance of being meaningful. The thing we can measure is the force:

$$\begin{aligned}
 F &= -\partial_d E_0 = -(f'(d) - f'(L-d)) \\
 &= -\hbar \left(\left(\frac{\pi}{2a^2} + \frac{\pi}{24d^2} + \mathcal{O}(a^2) \right) - \left(\frac{\pi}{2a^2} + \frac{\pi}{24(L-d)^2} + \mathcal{O}(a^2) \right) \right) \\
 &\stackrel{a \rightarrow 0}{=} -\frac{\pi\hbar}{24} \left(\frac{1}{d^2} - \frac{1}{(L-d)^2} \right) \\
 &\stackrel{L \gg d}{=} -\frac{\pi\hbar c}{24d^2} (1 + \mathcal{O}(d/L)) . \tag{1.6}
 \end{aligned}$$

This is an attractive force between the plates. (I put the c back in the last line by dimensional analysis.)

The analogous force between real conducting plates, caused by the change of boundary conditions on the electromagnetic field, has been measured.

The string theorists will tell you that $\sum_{j=1}^{\infty} j = -\frac{1}{12}$, and our calculation above agrees with them in some sense. But what this foolishness means is that if we compute something which is not dependent on the cutoff we have to get the same answer no matter what cutoff we use. Notice that it is crucial to ask the right questions.

An important question is to what extent could we have picked a different cutoff function (instead of $e^{-\pi\omega/a}$) and gotten the same answer for the physics. This interesting question is answered affirmatively in Zee's wonderful book, 2d edition, section I.9 Appendix (available electronically [here!](#)). Go study this appendix now.

A comment about possible physical applications of the calculation we actually did: you could ask me whether there is such a thing as a Casimir force due to the vacuum fluctuations of *phonons*. Certainly it's true that the boundary of a chunk of solid puts boundary conditions on the phonon modes, which change when we change the size of the solid. The problem with the idea that this might produce a measurable force (which would lead the solid to want to shrink) is that it is hard to distinguish the 'phonon vacuum energy' from the rest of the *energy of formation* of the solid, that is, the energy difference between the crystalline configuration of the atoms and the configuration when they are all infinitely separated. Certainly the latter is not well-described in the harmonic approximation.

Lest this seem too disconnected from the other 'divergences' we will encounter in QFT associated with loop amplitude, let me emphasize that the vacuum energy *is* a loop amplitude in this free field theory (the only one). Specifically

$$E_0 = \langle 0|H_0|0\rangle = \langle 0| \int d^d x \phi(x) (-\partial_t^2 + \nabla^2 + m^2) \phi(x) |0\rangle = \text{[loop diagram]} .$$

A few comments about the 3+1 dimensional case of E&M. Assume the size of the plates is much larger than their separation L . Dimensional analysis shows that the force per unit area from vacuum fluctuations must be of the form

$$P = A \frac{\hbar c}{L^4}$$

where A is a numerical number. A is not zero!

Use periodic boundary conditions in the xy planes (along the plates). The allowed wave vectors are then

$$\vec{k} = \left(\frac{2\pi n_x}{L_x}, \frac{2\pi n_y}{L_y} \right)$$

with n_x, n_y integers.

We have to do a bit of E&M here. Assume the plates are perfect conductors (this where the hubris about the high-frequency modes enters). This means that the transverse component of the electric field must vanish at the surface. Instead of plane waves in z , we get standing waves: $\phi(z) \propto \sin(n\pi z/L)$.

The frequencies of the associated standing waves are then

$$\omega_n(\vec{k}) = c \sqrt{\frac{\pi^2 n^2}{L^2} + \vec{k}^2}, n = 0, 1, 2$$

Also, there is only one polarization state for $n = 0$.

So the zero-point energy is

$$E_0(L) = \frac{\hbar}{2} \left(2 \sum'_{n, \vec{k}} \omega_n(\vec{k}) \right)$$

where it's useful to define

$$\sum'_{n, \vec{k}} \equiv \frac{1}{2} \sum_{n=0, \vec{k}} + \sum_{n \geq 1, \vec{k}}$$

Now you can imagine introducing a regulator like the one we used above, and replacing

$$\sum'_{n, \vec{k}} \cdot \rightsquigarrow \sum'_{n, \vec{k}} e^{-a\omega_n(\vec{k})/\pi}.$$

and doing the sums and integrals and extracting the small- a behavior.

1.3 A simple example of perturbative renormalization in QFT

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

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{1}{2}m^2\phi^2 - \frac{g}{4!}\phi^4. \quad (1.7)$$

Demanding that the action is dimensionless means 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 ϕ particles.

$$\begin{aligned} i\mathcal{M}_{2\leftarrow 2} &= \text{tree} + \text{1-loop } s\text{-channel} + \text{1-loop } t\text{-channel} + \text{1-loop } u\text{-channel} + \mathcal{O}(g^3) \\ &= -ig + i\mathcal{M}_s + i\mathcal{M}_t + i\mathcal{M}_u + \mathcal{O}(g^3) \end{aligned}$$

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

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

Parametrizing ignorance. What is a scalar field? One way to discover scalar field theory is to start with a chain of masses connected by springs, like a mattress, and look at the long-wavelength (small-wavenumber) modes. So the coherent excitations of such a field are sound waves and the quanta of the field are called phonons. 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 over k 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 §1.2: in that calculation, different choices of regulators for the mode sum corresponded (for example) to different material properties of the conducting plates. The 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.

If we had an actual lattice (like the chain of springs), we would replace the inverse propagator $p^2 - m^2 = \omega^2 - \vec{p}^2 - m^2$ with $\omega^2 - \omega_p^2 - m^2$, where ω_p is the dispersion

relation (e.g. $\omega_p = 2t \sum_{i=1}^d (1 - \cos p_i a)$ for nearest-neighbor hopping on the cubic lattice), and p is restricted to the Brillouin zone ($-\pi/a \leq p_i < \pi/a$ for the cubic lattice). Instead, for simplicity, let's keep just impose a hard cutoff on the euclidean momentum $\sum_{i=0}^d p^2 \leq \Lambda^2$.

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

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

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

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

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

[End of Lecture 2]

Observables can be predicted from other observables. Again, the boldface statement might sound like some content-free tweet from a 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 ϕ particles at some given $(s, t, u) = (s_0, t_0, u_0)$, and finds for the amplitude $\mathbf{iM}(s_0, t_0, u_0) = -\mathbf{i}g_P$ where P is for ‘physical’.⁶ *This* we can relate to our theory letters:

$$-\mathbf{i}g_P = \mathbf{iM}(s_0, t_0, u_0) = -\mathbf{i}g + \mathbf{i}Cg^2 L_0 + \mathcal{O}(g^3) \tag{1.8}$$

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.) I emphasize that this much we would have to do to make useful predictions, even if there were no specter of infinity or dependence on a fictitious cutoff.

⁶You might hesitate here about my referring to the amplitude \mathcal{M} as an ‘observable’. The difficult and interesting question of what can actually be measured in experiments can be decoupled a bit from this discussion. If you want to worry about this, see the beginning of Schwartz, chapter 18.

Renormalization. Now here comes the big gestalt shift: Solve this equation (1.8) for the stupid letter g

$$\begin{aligned} -\mathbf{i}g &= -\mathbf{i}g_P - \mathbf{i}Cg^2L_0 + \mathcal{O}(g^3) \\ &= -\mathbf{i}g_P - \mathbf{i}Cg_P^2L_0 + \mathcal{O}(g_P^3). \end{aligned} \quad (1.9)$$

and eliminate g from the discussion:

$$\begin{aligned} \mathbf{i}\mathcal{M}(s, t, u) &= -\mathbf{i}g + \mathbf{i}Cg^2L + \mathcal{O}(g^3) \\ &\stackrel{(1.9)}{=} -\mathbf{i}g_P - \mathbf{i}Cg_P^2L_0 + \mathbf{i}Cg_P^2L + \mathcal{O}(g_P^3) \\ &= -\mathbf{i}g_P + \mathbf{i}Cg_P^2 \left(\log \frac{s_0}{s} + \log \frac{t_0}{t} + \log \frac{u_0}{u} \right) + \mathcal{O}(g_P^3). \end{aligned} \quad (1.10)$$

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 Λ is gone! Just like in our parable in §1.1, it was eliminated by letting the coupling vary with it, $g = g(\Lambda)$, according to (1.9). 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 (1.7), with $m = 0$ for simplicity):

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 \quad (1.11)$$

but written in terms of the measured coupling g_P , and some as-yet-undetermined ‘counterterm’ δ_g . Then (with some foresight, we treat $\delta_g \sim g_P^2$, since its job is to cancel a term of this order)

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

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

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

then we find

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

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

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

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

1.4 Classical interlude: Mott formula

As a prelude to studying loops in QED, and to make clear what is at stake, let me fill in some of the details of the leading-order calculation. 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). This will emphasize the fact that the tree-level process is classical. 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:

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

(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:

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

– 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.

$\mu^+\mu^- \leftarrow e^+e^-$. For example, consider the process $\mu^+\mu^- \leftarrow e^+e^-$. 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. The amplitude is

$$i\mathcal{M}_{\mu^+\mu^-\leftarrow e^+e^-} = \left(-ie\bar{u}^s(k)\gamma^\mu v^{s'}(k') \right)_{\text{muons}} \frac{-i\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu q_\nu}{q^2} \right)}{q^2} \left(-ie\bar{v}^{r'}(p')\gamma^\nu u^r(p) \right)_{\text{electrons}} \quad (1.12)$$

(with $q \equiv p + p' = k + k'$). If we don't keep track of the spins, then we must average over initial states and sum over final states, so the unpolarized scattering probability density is

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

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

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

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

Spinor trace ninjutsu. In case you missed it, here is how to evaluate the spinor traces:

$$\text{The trace is cyclic: } \text{tr}(AB \cdots C) = \text{tr}(CAB \cdots). \quad (1.13)$$

Our gamma matrices are 4×4 , so $\text{tr} \mathbb{1} = 4$.

$$\text{tr} \gamma^\mu = \text{tr} (\gamma^5)^2 \gamma^\mu \stackrel{(1.13)}{=} \text{tr} \gamma^5 \gamma^\mu \gamma^5 \stackrel{\{\gamma^5, \gamma^\mu\}=0}{=} -\text{tr} \gamma^\mu = 0. \quad (1.14)$$

The same trick shows that the trace of any odd number of gammas vanishes. The idea is that an odd number of gammas is a map between the L and R subspaces, so it has only off-diagonal terms in the Weyl basis.

$$\text{tr} \gamma^\mu \gamma^\nu \stackrel{\text{clifford}}{=} -\text{tr} \gamma^\nu \gamma^\mu + 2\eta^{\mu\nu} \text{tr} \mathbb{1} \stackrel{(1.13)}{=} -\text{tr} \gamma^\mu \gamma^\nu + 8\eta^{\mu\nu} \implies \text{tr} \gamma^\mu \gamma^\nu = 4\eta^{\mu\nu}. \quad (1.15)$$

$$\text{tr} \gamma^\mu \gamma^\nu \gamma^\rho \gamma^\sigma = 4(\eta^{\mu\nu} \eta^{\rho\sigma} + \eta^{\sigma\mu} \eta^{\nu\rho} - \eta^{\mu\rho} \eta^{\nu\sigma}). \quad (1.16)$$

Why is this? The completely antisymmetric bit vanishes because it is proportional to γ^5 which is traceless (by the same argument as (1.14)). If any pair of indices is the same then the other two must be too by (1.15). If adjacent pairs are the same they can just square to one and we get +1; if alternating pairs are the same (and different from each other) then we must move them through each other with the anticommutator. If they are all the same we get 4.

$$\text{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^5 = -4i\epsilon^{\mu\nu\rho\sigma}.$$

$e^- \mu^- \leftarrow e^- \mu^-$. To get from our previous work 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 σ will be $(-1)^{\text{number of fermions shuffled between in and out}} = (-1)^2 = 1$). Relative to the amplitude for $\mu^+ \mu^- \leftarrow e^+ e^-$ (1.12), we must replace the relevant vs with us for the initial/final antiparticles that were moved into final/initial particles, and we must replace $p' \rightarrow -p', k' \rightarrow -k'$:

$$i\mathcal{M} = \begin{array}{c} \text{Diagram: } e^- \mu^- \leftarrow e^- \mu^- \text{ with } q_t = p - p' = k - k' \end{array} = (-ie\bar{u}(p')\gamma^\mu u(p))_{\text{electrons}} \frac{-i\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu^t q_\nu^t}{q_t^2}\right)}{q_t^2} (-ie\bar{u}(k)\gamma^\nu u(k'))_{\text{muons}} \quad (1.17)$$

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

$$\frac{1}{4} \sum_{s,s',r,r'} |\mathcal{M}|^2 = 4 \frac{e^4}{t^2} (-p^\mu p'^\nu - p'^\mu p^\nu - \eta^{\mu\nu}(-p \cdot p' + m_e^2)) \cdot (-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k \cdot k' + m_\mu^2)) \quad (1.18)$$

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

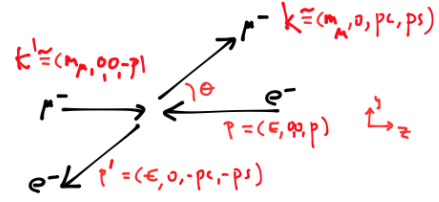
Payoff: the Mott formula. Recall other ways of figuring out the scattering cross

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 $k'_0 = m_\mu$, $k_0 = \sqrt{m_\mu^2 + \vec{k}^2} = m_\mu + \frac{1}{2}\vec{k}^2/m_\mu + \dots \simeq m_\mu$. Also, this means the collision is approximately elastic. In the diagram of the kinematics at right, $c \equiv \cos \theta$, $s \equiv \sin \theta$.



This means that the muon-line tensor factor $M_{\mu\nu}$ in (1.18) simplifies dramatically:

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

In the electron line, we get

$$-p \cdot p' + m_e^2 = -E^2 + \vec{p}^2 \cos \theta + m_e^2 = -\vec{p}^2(1 - \cos \theta). \quad (1.19)$$

So

$$\begin{aligned} E^{\mu\nu} M_{\mu\nu} &= 32m_\mu^2 E^{00} = 32m_\mu^2 (2E^2 + \eta^{00}(p \cdot p' - m_e^2)) \\ &\stackrel{(1.19)}{=} 32m_\mu^2 (2E^2 - \vec{p}^2(1 - \cos \theta)) \\ &\stackrel{\text{trig}}{=} 32m_\mu^2 2(E^2 - \vec{p}^2 \sin^2 \theta/2) \stackrel{\beta^2 \equiv \vec{p}^2/E^2}{=} 64m_\mu^2 E^2 (1 - \beta^2 \sin^2 \theta/2). \end{aligned} \quad (1.20)$$

Note that $t = (p - p')^2 = -2\vec{p}^2(1 - \cos \theta)$.

Now we can use the formula for the cross section (for the derivation, see [these notes](#), §4.7):

$$d\sigma = \frac{1}{v_{\text{rel}}} \frac{1}{2E_A 2E_B} \left(\frac{1}{4} \sum_{\text{spins}} |\mathcal{M}|^2 \right) d\Pi_{LI}. \quad (1.21)$$

So the differential cross section is

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

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 $\theta \rightarrow 0$. 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):

$$i\mathcal{M}_{e\mu \leftarrow e\mu} = \text{tree} + \left(\text{self-energy} + \text{vertex} \right) + \left(\text{muon self-energy} + \text{muon vertex} + \text{muon exchange} + \text{muon exchange} \right) + \mathcal{O}(e^6)$$

- What do the one-loop diagrams in the second line have in common? They have an internal muon line. Why does this matter? When the energy going through the line is much smaller than the muon mass, then the propagator is $\frac{i(k+m_\mu)}{k^2-m_\mu^2} \sim \frac{1}{m_\mu}$ 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

$$S_{e\mu \leftarrow e\mu} = \sqrt{Z_e}^2 \sqrt{Z_\mu}^2 \left(\text{tree} + \left(\text{self-energy} + \text{vertex} \right) + \dots \right)_{\text{amputated, on-shell}}$$




and in determining the locations of the poles whose residues are the S -matrix elements. We'll take care of these when we talk about the electron self-energy.

- 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

$$\sigma \sim \left| \text{tree} + \left(\text{self-energy} \cdot \text{tree} + \text{tree} \cdot \text{vertex} \right) + \dots \right|^2 \sim \sigma_{\text{tree}} + \mathcal{O}(\alpha^3)$$

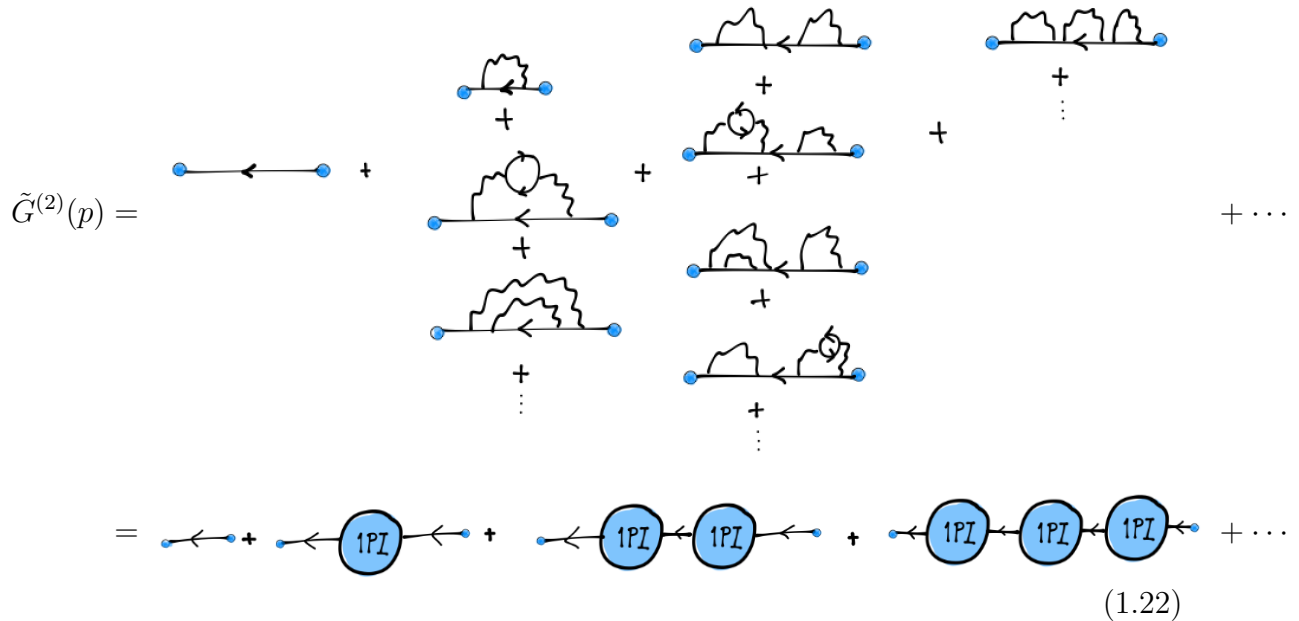
from the cross term between the tree and one-loop amplitudes.

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


- electron self-energy: 
- vertex correction: 
- vacuum polarization (photon self-energy): 

1.5 Electron self-energy in QED

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

$$\begin{aligned}
 \tilde{G}^{(2)}(p) = & \text{[tree diagram]} + \text{[self-energy loop]} + \text{[vacuum polarization]} + \text{[vertex correction]} + \dots \\
 = & \text{[tree]} + \text{[1PI blob]} + \text{[1PI blob]} \text{[1PI blob]} + \text{[1PI blob]} \text{[1PI blob]} \text{[1PI blob]} + \dots
 \end{aligned}
 \tag{1.22}$$


We’ve grouped the diagrams according to their behavior when we divide input and output by cutting a single line. A diagram which cannot be divided by cutting a single line is called *one-particle irreducible* (1PI). The blue blob is defined to be the sum of all 1PI diagrams. We will denote the 1PI two-point function by

$$-i\Sigma(p) \equiv \text{[1PI blob]}$$


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:

$$\begin{aligned} \tilde{G}^{(2)}(p) &= \mathbf{i}S + \mathbf{i}S(-\mathbf{i}\Sigma(p))\mathbf{i}S + \mathbf{i}S(-\mathbf{i}\Sigma(p))\mathbf{i}S(-\mathbf{i}\Sigma(p))\mathbf{i}S + \dots \\ &= \mathbf{i}S(1 + \Sigma S + \Sigma S \Sigma S + \dots) = \mathbf{i}S \frac{1}{1 - \Sigma S} \\ &= \frac{\mathbf{i}}{\not{p} - m_0} \frac{1}{1 - \Sigma \frac{1}{\not{p} - m_0}} = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)}. \end{aligned} \tag{1.23}$$

A comment about summing this infinite series:

$$1 + x + x^2 + \dots = \frac{1}{1 - x}. \tag{1.24}$$

You probably know that a geometric series has a radius of convergence of 1. This is because the function to which it sums has a pole at $x = 1$, and the radius of convergence is at most the distance to the nearest singularity.

On the other hand, there is a theorem of complex analysis that if two functions analytic in an open set D agree in D then they are the same. This is the basis for analytic continuation.

The LHS of (1.24) is not a priori defined when $|x| > 1$. The relation (1.24) therefore gives a useful meaning to it in this regime.

[End of Lecture 3]

Are you worried about these manipulations because Σ and S are matrices in the spinor indices? Don't be: they are both made entirely from the matrix \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

$$\not{p} = m \equiv m_0 + \Sigma(m). \tag{1.25}$$

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 Σ .

In QED we must study Σ in perturbation theory. As you can see from (1.22), the leading (one-loop) contribution is

$$-i\Sigma_2(p) = \text{diagram} = (-ie)^2 \int d^4k \gamma^\mu \frac{i(\not{k} + m_0)}{k^2 - m_0^2 + i\epsilon} \gamma^\nu \frac{-i\eta_{\mu\nu}}{(p-k)^2 - \mu^2 + i\epsilon} .$$

Notice that I am relying on the Ward identity to enforce the fact that only the transverse bit of the photon propagator matters. Also, I added a mass μ 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

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

This allows us to combine the denominators into one:

$$\mathcal{I} = \frac{1}{\underbrace{k^2 - m_0^2 + i\epsilon}_B} \frac{1}{\underbrace{(p-k)^2 - \mu^2 + i\epsilon}_A} = \int_0^1 dx \frac{1}{(x((p^2 - 2pk + k^2) - \mu^2 + i\epsilon) + (1-x)(k^2 - m_0^2 + i\epsilon))^2}$$

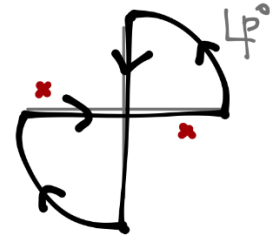
Step 2: Now we can complete the square

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

with

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

Step 3: Wick rotate. Because of the $i\epsilon$ 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: $\ell^0 \equiv i\ell^4$. 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

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

where the numerator is

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

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 ℓ

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

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

$$\begin{aligned} \Sigma_2(p) &= e^2 \int_0^1 dx \int \frac{\ell^2 d\ell^2}{2} \frac{(2\pi^2)}{(2\pi)^4} \frac{2(2m_0 - x\not{p})}{(\ell^2 + \Delta)^2} \\ &= \frac{e^2}{8\pi^2} \int_0^1 dx (2m_0 - x\not{p}) \mathcal{J} \end{aligned} \quad (1.26)$$

with

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

In the large ℓ part of the integrand this is

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

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

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

Recall that

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

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

$$\begin{aligned} \text{wiggly line with } \vec{k} &\rightsquigarrow -i\eta_{\mu\nu} \left(\frac{1}{k^2 - \mu^2 + i\epsilon} - \frac{1}{k^2 - \Lambda^2 + i\epsilon} \right) \\ &= -i\eta_{\mu\nu} \left(\frac{\mu^2 - \Lambda^2}{(k^2 - \mu^2 + i\epsilon)(k^2 - \Lambda^2 + i\epsilon)} \right) \end{aligned}$$

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

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

Remembering that the residue of the pole in the propagator is the probability for the field operator to create a particle from the vacuum, you might worry that this is a negative probability, and unitarity isn't manifest. This particle is a *ghost*. However, we will choose Λ so large that the pole in the propagator at $k^2 = \Lambda^2$ will never be accessed and we'll never have external Pauli-Villars particles. We are using this as a device to define the theory in a regime of energies much less than Λ . 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. Nice things about PV are that it is Lorentz invariant and gauge invariant; the bad thing is it's not unitary. 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

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

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

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

Finally then

$$\Sigma_2(p)_{PV} = \frac{\alpha}{2\pi} \int_0^1 dx (2m_0 - x\not{p}) \ln \frac{x\Lambda^2}{x\mu^2 + (1-x)m_0^2 - x(1-x)p^2}. \quad (1.27)$$

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

$$\begin{aligned}
\delta m &\equiv m - m_0 = \Sigma_2(\not{p} = m) + \mathcal{O}(e^4) = \Sigma_2(\not{p} = m_0) + \mathcal{O}(e^4) \\
&= \frac{\alpha}{2\pi} \int_0^1 dx (2-x)m_0 \ln \frac{x\Lambda^2}{\underbrace{x\mu^2 + (1-x)m_0^2 + x(1-x)m_0^2}_{=x\mu^2 + (1-x^2)m_0^2 \equiv f(x, m_0, \mu)}} \\
&= \frac{\alpha}{2\pi} \int_0^1 dx (2-x)m_0 \left(\underbrace{\ln \frac{\Lambda^2}{m_0^2}}_{\text{divergent}} + \underbrace{\ln \frac{xm_0^2}{f(x, m_0, \mu)}}_{\text{relatively small}} \right) \\
&\approx \frac{\alpha}{2\pi} \left(2 - \frac{1}{2} \right) m_0 \ln \frac{\Lambda^2}{m_0^2} = \frac{3\alpha}{4\pi} m_0 \ln \frac{\Lambda^2}{m_0^2}. \tag{1.28}
\end{aligned}$$

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 $\ln \frac{\Lambda^2}{m_0^2}$ 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 α 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 Λ and the order of perturbation theory to which we are working and other theorist bookkeeping garbage; m_e does not. At each order in perturbation theory, we eliminate m_0 and write our predictions in terms of m_e . It is not too surprising that the mass of the electron includes such contributions: it must be difficult to travel through space if you are constantly emitting and re-absorbing photons.

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

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)} \stackrel{p \sim m}{\simeq} \frac{\mathbf{i}Z}{\not{p} - m} + \text{regular terms.} \tag{1.29}$$

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

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

So then (1.29) becomes

$$\tilde{G}^{(2)}(p) \stackrel{p \sim m}{\sim} \frac{\mathbf{i}}{\not{p} - m - \frac{\partial \Sigma}{\partial \not{p}} \Big|_{m_0} (\not{p} - m)} = \frac{\mathbf{i}}{(\not{p} - m) \left(1 - \frac{\partial \Sigma}{\partial \not{p}} \Big|_{m_0}\right)} \quad (1.30)$$

So that

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

and at leading nontrivial order

$$\begin{aligned}\delta Z &= \frac{\partial \Sigma_2}{\partial \not{p}} \Big|_{m_0} \stackrel{(1.27)}{=} \frac{\alpha}{2\pi} \int_0^1 dx \left(-x \ln \frac{x\Lambda^2}{f(x, m_0, \mu)} + (2m_0 - xm_0) \frac{-2x(1-x)}{f(x, m_0, \mu)} \right) \\ &= -\frac{\alpha}{4\pi} \left(\ln \frac{\Lambda^2}{m_0^2} + \text{finite} \right).\end{aligned} \quad (1.31)$$

Here $f = f(x, m_0, \mu)$ is the same quantity defined in the second line of (1.28). We'll see below that the cutoff-dependence in δZ plays a crucial role in making the S matrix (for example for the $e\mu \rightarrow e\mu$ process we've been discussing) cutoff-independent and finite, when written in terms of physical variables.

1.6 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 sensitivity in (relativistic) QFT, before returning to the QED vertex correction and vacuum polarization.

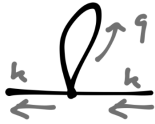
1.6.1 Self-energy in ϕ^4 theory

[Zee §III.3] Let's return to the ϕ^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 ϕ^4 term in the Lagrangian. Now we will see that we need counterterms for everybody:

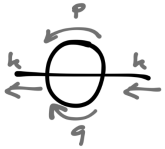
$$\mathcal{L} = -\frac{1}{2} (\phi \square \phi + m^2 \phi^2) - \frac{g_P}{4!} \phi^4 - \frac{\delta_g}{4!} \phi^4 - \frac{1}{2} \delta Z \phi \square \phi - \frac{1}{2} \delta_{m^2} \phi^2.$$

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

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


which is certainly quadratically divergent, but totally independent of the external momentum. This means that when we Taylor expand in k (as we just did in (1.30)), this diagram only contributes to the mass renormalization. Demanding that the pole in the propagator occurs at $p^2 = m^2$, we must set $\delta_{m^2} = -\delta\Sigma_1$.

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

$$\delta\Sigma_2(k) = \text{diagram} = \frac{(-ig)^2}{3!} \int d^4p \int d^4q \mathbf{i}D_0(p)\mathbf{i}D_0(q)\mathbf{i}D_0(k-p-q) \equiv I(k^2, m, \Lambda).$$


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

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

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

If instead the physical pole were at a nonzero value of the mass, we should Taylor expand about $k^2 = m_P^2$ instead:

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m_0^2 - \underbrace{(\delta\Sigma_1(m_P^2) + A_0)}_{\equiv a \sim \Lambda^2} - (k^2 - m_P^2)A_1 - (k^2 - m_P^2)^2 A_2 + \dots \quad (1.32)$$

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

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

with

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

where a was defined in (1.32).

[End of Lecture 4]

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 $A_n \phi \square^n \phi$. 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 ubiquitous⁷ (see the homework) and is the source of many headaches.

• On the term ‘wavefunction renormalization’: who is ϕ ? 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 a mattress), in high energy theory not so much. Classically, we fixed its (multiplicative) normalization by setting the coefficient of $\phi \square \phi$ 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 ϕ^4 theory is then

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

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

$$\text{X} = -ig_P, \quad \text{---} = \frac{i}{k^2 - m_P^2 + i\epsilon} \quad (1.33)$$

⁷At least for most regulators. We’ll see that dim reg is special.

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

$$\begin{array}{c} \diagup \times \diagdown \\ \circ \end{array} = -i\delta_g, \quad \text{---} \begin{array}{c} \times \\ \circ \end{array} \text{---} = -i(\delta Z k^2 + \delta m^2).$$

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 δ_{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 (1.33) 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 ϕ s (such as $3 \leftarrow 3$ scattering of ϕ quanta). How do we know those don't require new counterterms (like a ϕ^6 term, for example)?

1.6.2 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 $D_{\mathcal{A}} \equiv [\mathcal{A}]$, just its engineering dimension. In the limit that $\Lambda \gg$ all other scales, we must then have $\mathcal{A} \sim \Lambda^{D_{\mathcal{A}}}$ (in the absence of cancellations). A log divergent amplitude has $D_{\mathcal{A}} = 0$ (sometimes it's called $D_{\mathcal{A}} = 0^+$).

Let's start simple, and study the ϕ^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 doesn't it (explicitly) depend on any other data of the diagram, such as

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



? 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.

$$\begin{aligned} B_I &= 8 \\ B_E &= 4 \\ V &= 5 \\ L &= 4 \end{aligned}$$

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

graph,

$$L = B_I - (V - 1). \quad (1.34)$$

Math proof⁸: 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: $\prod_{\alpha=1}^{B_I} \int d^D q_\alpha$. 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, (1.34) 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 ϕ^4 theory). In the example, this is $4 + 2 \cdot 8 = 4 \cdot 5$. So we can eliminate

$$B_I = 2V - B_E/2. \quad (1.35)$$

Now we count powers of momenta:

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

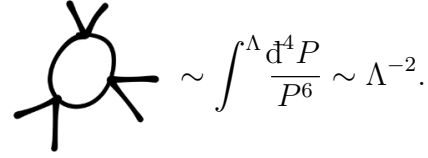
Since we are interested in the UV structure, I've set the mass to zero, as well as all the external momenta. The only scale left in the problem is the cutoff, so the dimensions of \mathcal{A} must be made up by the cutoff:

$$\begin{aligned} D_{\mathcal{A}} = [\mathcal{A}] &= DL - 2B_I \\ &\stackrel{(1.34)}{=} B_I(D - 2) - D(V - 1) \\ &\stackrel{(1.35)}{=} D + \frac{2 - D}{2} B_E + V(D - 4). \end{aligned}$$

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

⁸I learned this one from my class-mate M.B. Schulz.

the cutoff dependence must go away when $\Lambda \rightarrow 0$. This is illustrated by the following diagram with $B_E = 6$:



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.⁹

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

$$S_{\text{bare}}[\phi, \psi] = - \int d^D x \left(\frac{1}{2} \phi (\square + m_\phi^2) \phi + \bar{\psi} (-\not{\partial} + m_\psi) \psi + y \phi \bar{\psi} \psi + \frac{g}{4!} \phi^4 \right).$$

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

$$D_{\mathcal{A}} = D + (D - 4) \left(V_g + \frac{1}{2} V_y \right) + B_E \left(\frac{2 - D}{2} \right) + F_E \left(\frac{1 - D}{2} \right). \quad (1.36)$$

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

To temper your enthusiasm, consider adding a four-fermion interaction: $G(\bar{\psi}\psi)(\bar{\psi}\psi)$ (or maybe $G_V(\bar{\psi}\gamma^\mu\psi)(\bar{\psi}\gamma_\mu\psi)$ or $G_A(\bar{\psi}\gamma^\mu\gamma^5\psi)(\bar{\psi}\gamma_\mu\gamma^5\psi)$ or any other pile of gamma matrices in between, with the indices contracted). When you redo this calculation on

⁹Why isn't it a proof of renormalizability? Consider the following integral:

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

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

the homework, you'll find that in $D = 4$ a diagram (for simplicity, one with no ϕ^4 or Yukawa interactions) has

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

where V_G is the number of insertions of the 4-fermion term. This dependence on the number of four-fermi vertices means that there are worse and worse divergences as we look at higher-order corrections to a given process. Even worse, it means that for any number of external lines F_E no matter how big, there is a large enough order in perturbation theory in G where the cutoff will appear! This means we need $\delta_n(\bar{\psi}\psi)^n$ counterterms for every n , which we'll need to fix with physical input. This is a bit unappetizing, and such an interaction is called “non-renormalizable”. However, when we remember that we only need to make predictions to a given precision (so that we only need to go to a finite order in this process) we will see that such theories are nevertheless quite useful.

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

1.6.3 Naive scale invariance in field theory

[Halpern] Consider a field theory of a scalar field ϕ 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 ϕ :

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

Notice that the $D = 1$ case agrees with our quantum mechanics counting from §1.1. Quantum field theory in $D = 1$ spacetime dimensions *is* quantum mechanics.

Then the self-interaction term has dimensions

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

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

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

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

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

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

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

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

– it's a mass, yay.

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

$$[g] = D \cdot \frac{p_D - p}{p_D} \begin{cases} < 0 \text{ when } p > p_D, & \text{non-renormalizable or irrelevant} \\ = 0 \text{ when } p = p_D, & \text{renormalizable or marginal} \\ > 0 \text{ when } p < p_D, & \text{super-renormalizable or relevant.} \end{cases} \quad (1.37)$$

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

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


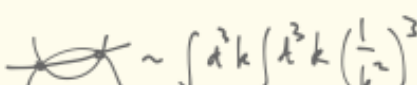
So if $[g] < 0$, c_n must have more and more powers of some mass (inverse length) as n increases. What dimensionful quantity makes up the difference? Sometimes it is


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, Λ_{UV} , the dimensions get made up by the UV cutoff, which has $[\Lambda_{UV}] = 1$. Generically: $c_n = \tilde{c}_n (\Lambda_{UV})^{-n[g]}$, 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 Λ_{IR} is an infrared cutoff or a mass or external momentum, $[\Lambda_{IR}] = 1$. 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:

ϕ^4 in $D = 4$:  $\sim \int d^4k \left(\frac{1}{k^2}\right)^2$ ϕ^6 in $D = 3$:  $\sim \int d^3k \int d^3k \left(\frac{1}{k^2}\right)^3$

ϕ^3 in $D = 6$:  $\sim \int d^4k \left(\frac{1}{k^2}\right)^3$ In $D = 2$, even the propagator for a massless scalar field has logs:


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

The terms involving ‘renormalizable’ in (1.37) are somewhat old-fashioned and come from a high-energy physics point of view where the short-distance physics is unknown, and we want to get as far as we can in that direction with our limited knowledge (in which case the condition ‘renormalizability’ lets us get away with this indefinitely – it lets us imagine we know everything). The latter terms are natural in the opposite situation (like condensed matter physics) where we know some basically correct microscopic description but want to know what happens at low energies. Then an operator like $\frac{1}{M^{2d}}\phi^{2d}$ 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 ϕ 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.

[End of Lecture 5]

1.7 Vertex correction in QED

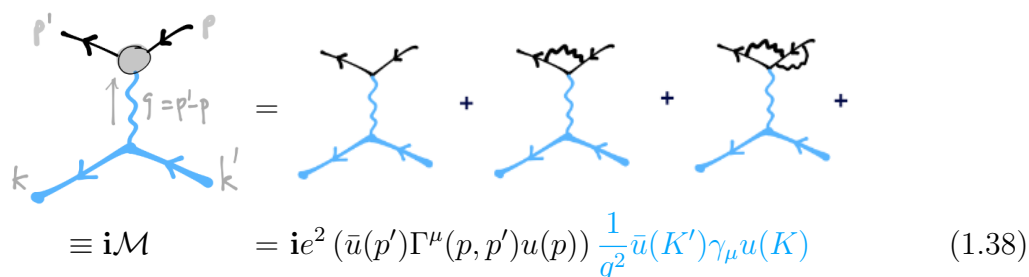
[Peskin chapter 6, Schwartz chapter 17, Zee chapter III.6] Back to work on QED. The

vertex correction  has some great physics payoffs:

- We'll cancel the cutoff dependence we found in the S matrix from δZ .
- We'll compute $g - 2$ (the anomalous magnetic moment) of the electron, the locus of some of the most precise agreement between theory and experiment. (Actually the agreement is so good that it's used as the *definition* of the fine structure constant. But a similar calculation gives the leading anomalous magnetic moment of the muon.)
- We'll see that the exclusive differential cross section $\left(\frac{d\sigma}{d\Omega}\right)_{e\mu \leftarrow e\mu}$ that we've been considering is not really an observable. Actually it is infinity!¹⁰ 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.



$$\equiv i\mathcal{M} = ie^2 (\bar{u}(p')\Gamma^\mu(p, p')u(p)) \frac{1}{q^2} \bar{u}(K')\gamma_\mu u(K) \quad (1.38)$$

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

¹⁰More accurately, the exclusive cross section is zero; the one-loop correction is minus infinity, which is perturbation theory's clumsy attempt to correct the finite tree level answer to make it zero.

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', γ^μ and m, e and numbers. It can't have any γ^5 or $\epsilon^{\mu\nu\rho\sigma}$ by parity symmetry of QED. So on general grounds we can organize it as

$$\Gamma^\mu(p, p') = A\gamma^\mu + B(p + p')^\mu + C(p - p')^\mu \quad (1.39)$$

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:

$$0 \stackrel{\text{Ward}}{=} q_\mu \bar{u}(p') \Gamma^\mu u(p) \stackrel{(1.39)}{=} \bar{u}(p') \left(A \underbrace{\not{q}}_{=\not{p}' - \not{p} \substack{\bar{u}(p') \dots u(p) \\ m-m=0}} + B \underbrace{(p + p') \cdot (p - p')}_{=m^2 - m^2 = 0} + Cq^2 \right) u(p)$$

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 $\sigma^{\mu\nu} \equiv \frac{1}{2}[\gamma^\mu, \gamma^\nu]$) can be used to eliminate the $p + p'$ term¹¹. 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* $F_{1,2}$:

$$\Gamma^\mu(p, p') = \gamma^\mu F_1(q^2) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} F_2(q^2). \quad (1.40)$$

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

¹¹Actually this is why we didn't include a $\sigma^{\mu\nu}$ term. You could ask: what about a term like $\sigma^{\mu\nu}(p + p')^\nu$? Well, there's another Gordon identity that relates that to things we've already included:

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

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

The first term at zero momentum $eF_1(q^2 = 0)$ is the electric charge of the electron (if you don't believe it, use the vertex (1.40) to calculate the Coulomb field of the electron; there are some details on page 186 of Peskin). Since the tree-level bit of F_1 is 1, if by the letter e here we mean the actual charge, then we'd better include counterterms ($\mathcal{L}_{ct} \ni \bar{\psi} \delta_e \gamma^\mu A_\mu \psi$) to make sure it isn't corrected: $F_1(0) = 1$.

The magnetic moment of the electron is the coefficient $\vec{\mu}$ of $\tilde{V}(q) = -\vec{\mu} \cdot \vec{B}(q) + \dots$ in the non-relativistic effective potential. Comparing the non-relativistic limit of $\bar{u}(p') \Gamma^i u(p) A_i(q) = -\vec{\mu} \cdot \vec{B}(q) + \dots$, (see the homework) shows that (see Peskin p. 187)


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

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

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

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

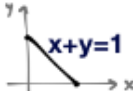
Now that we have some expectation about the form of the answer, and some ideas about what it's for, we sketch the evaluation of the one-loop QED vertex correction:



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

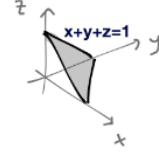
with $k' \equiv k + q$.

Step (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 generalization¹²:

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



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

$$\int \frac{d^4 k N^\mu}{(k^2 + k \cdot (\dots) + \dots)^3}$$

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

$$\Delta = -xyq^2 + (1-z)^2 m^2 + zm_\gamma^2. \quad (1.42)$$

The ℓ -dependence in the numerator is either 1 or ℓ^μ or $\ell^\mu \ell^\nu$. In the integral over ℓ , 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

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

Right now we only need $D = 4$ and $m = 3$, but it turns out to be quite useful to think about them all at once. Like in our discussion of the electron self-energy diagram, we

¹²Peskin outlines a proof by induction of the whole family of such identities on page 190. But here's a simpler proof using *Schwinger parameters*. You'll agree that

$$\frac{1}{A} = \int_0^\infty ds e^{-sA}. \quad (1.41)$$

Applying this identity to each factor gives

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

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

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

Now do the integral over τ , using $\int_0^\infty d\tau \tau^{n-1} e^{-\tau X} = \frac{(n-1)!}{X^n}$ (differentiate (1.41) wrt A), to arrive at

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

can evaluate them by Wick rotating (which changes the denominator to $\ell_E^2 + \Delta$) and going to polar coordinates. This gives:

$$\int \frac{d^D \ell}{(\ell^2 - \Delta)^m} = (-1)^m \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma(m - \frac{D}{2})}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2}}. \quad (1.43)$$

$$\int \frac{d^D \ell \ell^2}{(\ell^2 - \Delta)^m} = (-1)^{m-1} \frac{D}{2} \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma(m - \frac{D}{2} - 1)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2} - 1}. \quad (1.44)$$

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

Step (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_γ replaced by Λ^2 . The dangerous bit comes from the ℓ^2 term we just mentioned, since $m - D/2 - 1 = 3 - 4/2 - 1 = 0$ means logs.

The numerator is

$$\begin{aligned} N^\mu &= \bar{u}(p') \gamma^\nu (\not{k} + \not{q} + m_e) \gamma^\mu (\not{k} + m_e) \gamma_\nu u(p) \\ &= -2(\mathcal{A} \bar{u}(p') \gamma^\mu u(p) + \mathcal{B} \bar{u}(p') \sigma^{\mu\nu} q_\nu u(p) + \mathcal{C} \bar{u}(p') q^\mu u(p)) \end{aligned} \quad (1.45)$$

where

$$\begin{aligned} \mathcal{A} &= -\frac{1}{2} \ell^2 + (1-x)(1-y)q^2 + (1-4z+z^2)m^2 \\ \mathcal{B} &= \mathbf{i} m z (1-z) \\ \mathcal{C} &= m(z-2)(y-x). \end{aligned} \quad (1.46)$$

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

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

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

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

1.7.1 Anomalous magnetic moment

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

$$\begin{aligned}
 F_2(q^2) &= \frac{2m}{e} \cdot (\text{the term with } \mathcal{B}) \\
 &= \frac{2m}{e} 4e^3 (\mathbf{i}m) \int dx dy dz \delta(x+y+z-1) z(1-z) \underbrace{\int \frac{d^4\ell}{(\ell^2 - \Delta)^3}}_{= \frac{-i}{32\pi^2\Delta}} \\
 &= \frac{\alpha}{\pi} m^2 \int dx dy dz \delta(x+y+z-1) \frac{z(1-z)}{(1-z)^2 m^2 - xyq^2}. \tag{1.47}
 \end{aligned}$$

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

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

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

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

1.7.2 IR divergences mean wrong questions.

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

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

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

$$\begin{aligned}
 \int dx dy dz \delta(x+y+z-1) \frac{m^2(1-4z+z^2)}{\Delta} &= m^2 \int_0^1 dz \int_0^{1-z} dy \frac{-2 + 2(1-z) + (1-z)^2}{(1-z)^2 m^2} \\
 &= \int_0^1 dz \frac{-2}{(1-z)} + \text{finite}, \tag{1.48}
 \end{aligned}$$

which diverges at the upper limit of integration. In fact it's divergent even when $q^2 \neq 0$. This is a place where we actually need to include the photon mass, m_γ , for our own safety. The fact that restoring $m_\gamma \neq 0$ in (1.42) regulates this divergence is one way to see that it is indeed an IR divergence. [\[End of Lecture 6\]](#)

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

$$\Gamma^\mu = \gamma^\mu \left(1 - \frac{\alpha}{2\pi} f_{IR}(q^2) \ln \left(\frac{-q^2}{m_\gamma^2} \right) \right) + \text{stuff which is finite as } m_\gamma \rightarrow 0. \quad (1.49)$$

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

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

which blows up when we remove the fake photon mass $m_\gamma \rightarrow 0$. Notice that for t -channel exchange, $-q^2 > 0$, so the argument of the log is positive, the cross-section is real. But notice that the one-loop correction is not only infinite, but *negative* infinity, which simply cannot happen from the definition of the cross section. This is perturbation theory's way of telling us that the answer is $1 - \alpha \cdot \infty + \mathcal{O}(\alpha^2) \simeq 0$ – the putatively small corrections from radiative effects are actually trying to make the answer *zero*.

[Schwartz §20.1] I wanted to just quote the above result for (1.49) 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

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

δ_1 here is a counterterm for the $\Psi \gamma^\mu A_\mu \Psi$ vertex.

We can be more explicit if we consider $-q^2 \gg m_e^2$ so that we can ignore the electron mass everywhere. Then we would choose the counterterm δ_1 so that

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


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


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


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

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

Diversity and inclusion to the rescue. Before you throw up your hands in despair, I would like to bring to your attention another consequence of the masslessness of the photon: It means real (as opposed to virtual) photons can be made with arbitrarily low energy. But a detector has a minimum triggering energy: the detector works by particles doing some physical something to stuff in the detector, and it has a finite energy resolution – it takes a finite amount of energy E_c for those particles to do the stuff. This means that a process with exactly one e and one μ in the final state

 cannot be distinguished from a process ending in $e\mu$ plus a photon of

arbitrarily small energy, such as would result from  (final-state radiation)

or  (initial-state radiation). This ambiguity is present for *any* process with external charged particles.

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

$$\bar{u}(p')\mathcal{M}_0(p', p)u(p) \equiv -i \left(\text{Diagram: A central grey circle with multiple lines radiating from it. Two incoming lines from the left are labeled p' and p. Two outgoing lines to the right are labeled p' and p. A wavy line radiates upwards from the circle.} \right),$$

from more inclusive amplitudes like

$$-i \left(\text{Diagram: Similar to the previous one, but with an additional wavy line radiating from the left side of the circle, labeled k. The incoming line from the left is labeled p' + k.} \right) + \text{Diagram: Similar to the previous one, but with an additional wavy line radiating from the right side of the circle, labeled k. The outgoing line to the right is labeled p - k.} \\ = \bar{u}(p')\gamma^\mu \frac{e}{\not{p}' + \not{k} - m_e} \mathcal{M}_0(p', p)u(p)\epsilon_\mu^*(k) + \bar{u}(p')\mathcal{M}_0(p', p) \frac{e}{\not{p} - \not{k} - m_e} \gamma^\mu u(p)\epsilon_\mu^*(k).$$

Now, by assumption the extra outgoing 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

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

In the denominator we have *e.g.* $(p - k)^2 - m_e^2 = p^2 - m_e^2 - 2p \cdot k + k^2 \sim -2p \cdot k$ since the electron is on shell $p^2 = m_e^2$ and so is the photon $k^2 = 0$. Therefore

$$\mathcal{M}(e\mu + \text{one soft } \gamma \leftarrow e\mu) = e\bar{u}(p')\mathcal{M}_0(p', p)u(p) \left(\frac{p' \cdot \epsilon^*}{p' \cdot k + i\epsilon} - \frac{p \cdot \epsilon^*}{p \cdot k - i\epsilon} \right) \quad (1.50)$$

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

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

before? Notice that the $i\epsilon$ are different because one comes from final state and one from initial. Well, this object is the Fourier transform $\tilde{j}^{\mu}(k) = \int d^4x e^{+ikx} j^{\mu}(x)$ 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 ¹³

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

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 ω^{-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 (1.49)) arises classically as the *number* of soft photons produced by such a process 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 $\frac{1}{2} \int d^3x (E^2 + B^2) = \int d^3k k \hbar \omega_k n_k$. (Note that the integral over k here actually diverges; this is an artifact of the approximation that the momentum change is instantaneous.) See Peskin §6.1 for help.

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

¹³Check it:

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

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

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

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

which means that the lower limit of the k integral gets cut off at m_γ :

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

Being careful about the factors, the actual cross section measured by a detector with energy resolution E_c is¹⁴

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

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

I didn't show explicitly that the coefficient of the log is the same function $f_{IR}(q^2)$. In fact this function is $f_{IR}(q^2) = \frac{1}{2} \log(-q^2/m^2)$, so the product $f_{IR} \ln q^2 \sim \ln^2 q^2$ is the Sudakov double logarithm. A benefit of the calculation which shows that the same f_{IR} appears in both places (Peskin chapter 6.5) is that it also shows that this pattern persists at higher order in α : 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

¹⁴Notice that we add the cross-sections, not the amplitudes, for these processes with different final states. Here's why: even though we don't measure the existence of the photon, *something* does: it gets absorbed by some part of the apparatus or the rest of the world and therefore becomes entangled with 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.

photons. There is a $\frac{1}{2!}$ from Bose statistics of these photons. The result exponentiates, and we get

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

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

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

Sketch of exponentiation of soft photons. [Peskin §6.5] In the following we will just keep track of the bits which diverge when $m_\gamma \rightarrow 0$. Consider a diagram with n soft external photons, summed over ways of distributing them on an initial and final electron line:

$$\sum_{n_f=1}^n \left(\text{diagram with } n \text{ photons on } n_f \text{ lines} \right) = \bar{u}(p') \mathbf{iM}_0 u(p) e^n \prod_{\alpha=1}^n \left(\frac{p'^{\mu_\alpha}}{p' \cdot k_\alpha} - \frac{p^{\mu_\alpha}}{p \cdot k_\alpha} \right) \equiv \mathcal{A}_n.$$

Here the difference in each factor is just as in (1.50), 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 α and a final line β setting $k_\alpha = -k_\beta \equiv k$ and tying them together with a propagator and summing over k :

$$\left(\text{diagram with } n \text{ photons and a loop} \right) = \mathcal{A}_{n-2} \frac{e^2}{2} \int d^4k \frac{-\mathbf{i}\eta_{\rho\sigma}}{k^2 - m_\gamma^2} \left(\frac{p'}{p' \cdot k} - \frac{p}{p \cdot k} \right)^\rho \left(\frac{p'}{-p' \cdot k} - \frac{p}{-p \cdot k} \right)^\sigma \quad (1.51)$$

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

$$\bar{u} \mathbf{iM}_0 u \cdot \mathbf{X} = \left(\text{diagram with } 2 \text{ photons} \right) \cdot \left(\text{diagram with } 2 \text{ photons and a loop} \right)_{\text{soft part}}$$

(where here ‘soft part’ means the part which is singular in m_γ) from which we conclude that

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

(The integral is done in Peskin, page 201.) Taking the most IR-divergent bit with m virtual soft photons (order α^m) for each m gives

$$\mathcal{M}_{\text{virtual soft}} = \sum_{m=0}^{\infty} \left(\text{diagram with } m \text{ wavy lines} \right) = \left(\text{diagram with } m \text{ wavy lines} \right) \underbrace{\sum_m \frac{1}{m!} \mathbf{X}^m}_{e^{\mathbf{X}}}$$

where the $1/m!$ is a symmetry factor from interchanging the virtual soft photons. Notice that this verifies my claim that the $-\infty$ in the one-loop answer is perturbation theory’s way of trying to make the cross-section zero: since $\mathbf{X} \xrightarrow{m_\gamma \rightarrow 0} -\infty$, $d\sigma_{\text{exclusive}} \propto e^{2\mathbf{X}} \xrightarrow{m_\gamma \rightarrow 0} 0$.

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

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

(This was actually exactly the same integral as in the virtual-photon calculation in (1.51).) 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_γ dependence to all orders in α :

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

$$= d\sigma_0 \exp\left(-\frac{\alpha}{\pi} f_{IR}(q^2) \ln \frac{-q^2}{E_c^2}\right).$$

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

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

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

1.7.3 Some magic from gauge invariance of QED

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

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

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

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

Combining these together

$$\begin{aligned} S_{e\mu \leftarrow e\mu} &= \left(\sqrt{Z_2(e)}\right)^2 \left(\text{diagram} + \left(\text{diagram} \cdot \text{diagram} \right) + \dots \right) \\ &= \left(1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \dots\right) e^2 \bar{u}(p') \left(\gamma^\mu \left(1 + \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \dots\right) + \alpha \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \right) u(p) \end{aligned}$$

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 δ_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 \dots multiplying the γ^μ term still contain the vacuum polarization diagram, which is our next subject, and which may be (is) cutoff dependent. Read on.

1.8 Vacuum polarization

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

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

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:

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


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

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

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

But from this point of view it is clear that the magic of the previous subsection is a consequence of gauge invariance, here's why: the demand of gauge invariance relates the coefficients of the $\bar{\psi} \not{\partial} \psi$ and $\bar{\psi} \not{A} \psi$ terms¹⁵. Therefore, any counterterm we need for the $\bar{\psi} \not{\partial} \psi$ term (which comes from the electron self-energy correction and is traditionally called δZ_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 δZ_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 numerous, protons: once we recognize that charge renormalization is a property of the photon itself, it makes clear that quantum corrections cannot mess with the ratio of the charges. A deviation from -1 of the ratio of the charges of electron and proton as a result of interactions might seem plausible given what a mess the proton is, and would be a big deal for atoms. Gauge invariance forbids it.

Just as we defined the electron self-energy (amputated 2-point function) as  = $-\mathbf{i}\Sigma(p)$ (with two spinor indices implied), we define the photon self-energy as

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

(the diagrams on the RHS are amputated). It is a function of q^2 by Lorentz symmetry. (The reason for the difference in sign is that the electron propagator is $\frac{+\mathbf{i}}{p-m}$ while the

¹⁵Notice that the gauge transformation of the rescaled A_μ is $A_\mu \rightarrow A_\mu + \partial_\mu \lambda(x)$, $\psi(x) \rightarrow e^{\mathbf{i}q\lambda(x)}\psi(x)$ so that $D_\mu \psi \equiv (\partial + q\mathbf{i}A)_\mu \psi \rightarrow e^{\mathbf{i}q\lambda} D_\mu \psi$ where q is the charge of the field ($q = -1$ for the electron). This is to be contrasted with the transformation of $\tilde{A}_\mu \rightarrow \tilde{A}_\mu - \partial_\mu \lambda(x)/e$.

photon propagator is $\frac{-i\eta^{\mu\nu}}{q^2}$.) We can parametrize the answer as

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

The Ward identity says

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

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

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

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


$$\Delta_{T\rho}^\mu \Delta_{T\nu}^\rho = \Delta_{T\nu}^\mu \tag{1.52}$$

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

$$\text{wavy line} = \frac{-i\Delta_T}{q^2}$$

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

$$\begin{aligned} \tilde{G}^{(2)}(q) &= \text{wavy line} + \text{wavy line with bubble} + \text{wavy line with two bubbles} + \dots \\ &\stackrel{(1.52)}{=} \frac{-i\Delta_T}{q^2} \left(1 + i\Pi q^2 \Delta_T \left(\frac{-i\Delta_T}{q^2} \right) + i\Pi q^2 \Delta_T \left(\frac{-i\Delta_T}{q^2} \right) i\Pi q^2 \Delta_T \left(\frac{-i\Delta_T}{q^2} \right) + \dots \right) \\ &\stackrel{\Delta_T^2 = \Delta_T}{=} \frac{-i\Delta_T}{q^2} (1 + \Pi \Delta_T + \Pi^2 \Delta_T + \dots) = \frac{-i\Delta_T}{q^2} \frac{1}{1 - \Pi(q^2)}. \end{aligned} \tag{1.53}$$

Does the photon get a mass? If the thing I called A above $q^2 \Pi(q^2) \xrightarrow{q^2 \rightarrow 0} A_0 \neq 0$ (that is, if $\Pi(q^2) \sim \frac{A_0}{q^2}$ or worse), then $\tilde{G} \xrightarrow{q^2 \rightarrow 0} \frac{1}{q^2 - A_0}$ 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 Anderson-Higgs mechanism (not the easiest way to understand it). [\[End of Lecture 7\]](#)

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

$$q_\mu \Pi_2^{\mu\nu}(q^2) = q_\mu \text{wavy line with loop} \propto e^2 \int d^4p \text{tr} \frac{1}{\not{p} + \not{q} - m} \not{q} \frac{1}{\not{p} - m} \gamma^\nu.$$

But here is an identity:

$$\frac{1}{\not{p} + \not{q} - m} \not{q} \frac{1}{\not{p} - m} = \frac{1}{\not{p} - m} - \frac{1}{\not{p} + \not{q} - m}. \quad (1.54)$$

Now, if we shift the integration variable $p \rightarrow p + q$ 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_γ for example for the S -matrix for processes with external photons, like Compton scattering.

Claim: If we do it right¹⁶, the cutoff dependence looks like¹⁷:

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

where Λ 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

¹⁶What 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.

¹⁷The factor in front of the $\ln \Lambda$ can be made to look like it does in other textbooks using $\alpha = \frac{e^2}{4\pi}$, so that

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

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

$$e_0^2 = e^2 \left(1 + \frac{\alpha_0}{4\pi} \frac{2}{3} \ln \Lambda^2 \right) + \mathcal{O}(\alpha^2). \quad (1.55)$$

$$m_0 = m + \mathcal{O}(\alpha_0) = m + \mathcal{O}(\alpha). \quad (1.56)$$

Since the difference between α_0 and α is higher order (in either), our book-keeping is unchanged. Inverting the relationship perturbatively, the renormalized charge is

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

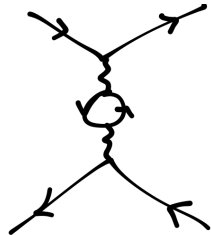
– in QED, the quantum fluctuations *reduce* the charge, as you might expect from the interpretation of this phenomenon as dielectric screening by virtual e^+e^- pairs.

In the example case of $e\mu \leftarrow e\mu$ scattering, the full one-loop UV cutoff dependence then looks like

$$\begin{aligned} S_{e\mu \leftarrow e\mu} &= \sqrt{Z_e^2} \left(1 - \frac{\alpha_0}{4\pi} \ln \Lambda^2 + \frac{\alpha_0}{2\pi} A(m_0) \right) e_0^2 \\ &\quad L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha_0}{4\pi} \ln \Lambda^2 + \frac{\alpha_0}{2\pi} (B + D) + \frac{\alpha_0}{4\pi} \left(-\frac{2}{3} \ln \Lambda^2 \right) \right) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \frac{\alpha_0}{2\pi} C(q^2, m_0) \right] u(p) \\ &= e^2 L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha}{2\pi} (A + B + D) \right) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \frac{\alpha}{2\pi} C \right] u(p) + \mathcal{O}(\alpha^2) \end{aligned} \quad (1.57)$$

where L_μ 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 (1.55) and (1.56) to write everything in terms of the measured e, m . This removes the remaining cutoff dependence.

Claim: this works for all processes to order α^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 is $A+B+D$ we need to specify more carefully a renormalization scheme (other combinations of A, B, D can be changed by gauge transformations and field redefinitions). To do that, I need to give a bit more detail about the integral.

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

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

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

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

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

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

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

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

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

[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 $\int \bar{d}^D p f(p+q) = \int \bar{d}^D p f(p)$ ¹⁹

¹⁹Note that this rule fails for the euclidean momentum cutoff. Also note that this is the property we needed to demonstrate the Ward identity for the vertex correction using (1.54).

2. scaling $\int \bar{d}^D p f(sp) = |s|^{-D} \int \bar{d}^D p f(p)$
3. factorization $\int \bar{d}^D p \int \bar{d}^D q f(p)g(q) = \int \bar{d}^D p f(p) \int \bar{d}^D q g(q)$

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

$$\left(\frac{\pi}{a}\right)^{D/2} = \int d^D x e^{-a\bar{x}^2} = \Omega_{D-1} \int_0^\infty x^{D-1} dx e^{-ax^2} = \frac{1}{2} a^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right) \Omega_{D-1}. \quad (1.58)$$

This defines Ω_{D-1} for general D .

In dim reg, the one-loop vacuum polarization correction *does* satisfy the gauge-invariance Ward identity $\Pi^{\mu\nu} = \Delta_T^{\mu\nu} q^2 \Pi_2(q^2)$. A peek at the tables of dim reg integrals shows that Π_2 is:

$$\begin{aligned} \Pi_2(q^2) &\stackrel{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\stackrel{D \rightarrow 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right) \right) \end{aligned} \quad (1.59)$$

where we have introduced the heralded μ :

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

where γ_E is the Euler-Mascheroni constant, which appears in the Taylor expansion of the Euler gamma function; we define μ in this way so that, like Rosencrantz and Guildenstern in Hamlet, γ_E both appears and disappears from the discussion in this one scene.

In the second line of (4.6), we expanded the Γ -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 Π gets subtracted by the counterterm. One possibility is to demand that the photon propagator is not corrected at $q = 0$, *i.e.* demand $Z_\gamma = 1$. Then the resulting one-loop shift is

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

We'll use this choice below.

Another popular choice, about which more later, is called the $\overline{\text{MS}}$ scheme, in which Π 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 Π_2 . Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ .) The resulting vacuum polarization function is

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

1.8.2 Physics from vacuum polarization

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

$$\tilde{V}(q) = \frac{1}{q^2} \frac{e^2}{1 - \Pi(q^2)} \equiv \frac{e_{\text{eff}}^2(q)}{q^2}. \quad (1.60)$$

This has consequences in both IR and UV.

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

$$\delta\Pi_2(q) = \frac{e^2}{2\pi^2} \int dx x(1-x) \ln\left(1 - \frac{q^2}{m^2} x(1-x)\right) \stackrel{q \ll m}{\simeq} \frac{e^2}{2\pi^2} \int dx x(1-x) \left(-\frac{q^2}{m^2} x(1-x)\right) = -\frac{q^2}{60\pi^2 m^2}$$

which means

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

and hence

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

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 Π_2 varies. The delta function approximation is a good idea for atomic physics, since $\frac{1}{m_e} \ll a_0 = \frac{1}{\alpha m_e}$, the Bohr radius. See Schwartz p. 311 for a bit more on this.

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 get²⁰

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

Therefore, the effective charge in (1.60) at high momentum exchange is

$$e_{\text{eff}}^2(q^2) \stackrel{q^2 \gg m_e^2}{\simeq} \frac{e^2}{1 - \frac{e^2}{12\pi^2} \ln\left(-\frac{q^2}{m^2}\right)}. \quad (1.61)$$

(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. The value $\alpha = \frac{1}{137}$ is the extreme IR value, for $q \ll m_e$.

²⁰The last step is safe since the $x(1-x)$ suppresses the contributions of the endpoints of the x integral, so we can treat $x(1-x)$ as finite.

2 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 wave-function 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 Källén-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’.

[End of Lecture 8]

2.1 Spectral density

[Zee III.8, Appendix 2; Peskin §7.1; 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 ²¹

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

²¹Note 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:

$$-\mathbf{i}\mathcal{D}(x) = \theta(t) \langle 0 | e^{\mathbf{i}\mathbf{P}x} \mathcal{O}(0) e^{-\mathbf{i}\mathbf{P}x} \mathcal{O}^\dagger(0) | 0 \rangle + \theta(-t) \langle 0 | \mathcal{O}^\dagger(0) e^{\mathbf{i}\mathbf{P}x} \mathcal{O}(0) e^{-\mathbf{i}\mathbf{P}x} | 0 \rangle. \quad (2.1)$$

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 (since the components of \mathbf{P}^μ commute with each other):

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

Let's examine the first term in (2.1); 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 :

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

Just like in the 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:

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

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

$$-\mathbf{i}\mathcal{D}(q) = \int d^D x e^{\mathbf{i}q x} \mathbf{i}\mathcal{D}(x) = \mathbf{i}(2\pi)^{D-1} \sum_n \|\mathcal{O}_{0n}\|^2 \left(\frac{\delta^{(D-1)}(\vec{q} - \vec{p}_n)}{q^0 - p_n^0 + \mathbf{i}\epsilon} - \frac{\delta^{(D-1)}(\vec{q} + \vec{p}_n)}{q^0 + p_n^0 - \mathbf{i}\epsilon} \right). \quad (2.2)$$

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

Now suppose that our operator \mathcal{O} is capable of creating a single particle (for example, suppose, if you must, that $\mathcal{O} = \phi$, a perturbative quantum field). Such a state is labelled only by its spatial momentum: $|\vec{k}\rangle$ (here I briefly retreat to non-relativistic normalization of states $\langle \vec{k} | \vec{k}' \rangle = \delta^{D-1}(\vec{k} - \vec{k}')$). The statement that \mathcal{O} can create this state from the vacuum means

$$\langle \vec{k} | \mathcal{O}(0) | 0 \rangle = \frac{Z^{\frac{1}{2}}}{\sqrt{(2\pi)^{D-1} 2\omega_{\vec{k}}}} \quad (2.3)$$

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

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

in terms of m , the *mass* of the particle.²² What is Z ? From (2.3) and the axioms of QM, you can see that it's the probability that \mathcal{O} creates this 1-particle state from the vacuum. In the free field theory it's 1, and it's positive because it's a probability. $1 - Z$ measures the extent to which \mathcal{O} does anything *besides* create this 1-particle state.

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

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

This is a summand in the whole horrible resolution:

$$\mathbb{1} = \mathbb{1}_1 + \dots$$

²²It's been a little while since we spoke explicitly about free fields, so let's remind ourselves about the appearance of $\omega^{-\frac{1}{2}}$ in (2.3), recall the expansion of a free scalar field in creation and annihilation operators:

$$\phi(x) = \int \frac{d^{D-1}\vec{p}}{\sqrt{2\omega_{\vec{p}}}} \left(\mathbf{a}_{\vec{p}} e^{-ipx} + \mathbf{a}_{\vec{p}}^\dagger e^{ipx} \right) .$$

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

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

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

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

I mention this because it lets us define the part of the horrible \sum_n in (2.2) which comes from 1-particle states:

$$\begin{aligned}
\Rightarrow \quad -\mathbf{i}\mathcal{D}(q) &= \dots + \mathbf{i}(2\pi)^{D-1} \int \mathrm{d}^{D-1}\vec{k} \frac{Z}{(2\pi)^{D-1}2\omega_k} \left(\frac{\delta^{D-1}(\vec{q}-\vec{k})}{q^0 - \omega_{\vec{k}} + \mathbf{i}\epsilon} - (\omega_k \rightarrow -\omega_k) \right) \\
&= \dots + \mathbf{i} \frac{Z}{2\omega_q} \left(\frac{1}{q^0 - \omega_q + \mathbf{i}\epsilon} - \frac{1}{q^0 + \omega_q + \mathbf{i}\epsilon} \right) \\
&\stackrel{\text{Lorentz}}{=} \dots + \mathbf{i} \frac{Z}{q^2 - m^2 + \mathbf{i}\epsilon}
\end{aligned}$$

(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 .²³

The imaginary part of \mathcal{D} is called the *spectral density* ρ (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

$$\text{Im} \frac{1}{Q \mp \mathbf{i}\epsilon} = \pm\pi\delta(Q), \quad (\text{for } Q \text{ real}). \quad (2.4)$$

we have

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

More explicitly (for real operators):

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

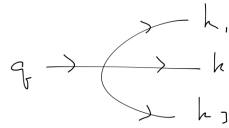
The second term on the RHS vanishes when $q^0 > 0$, since states in \mathcal{H} have energy bigger than the energy of the groundstate. Therefore, the contribution of a 1-particle state to the spectral density is:

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


²³If we hadn't assumed Lorentz invariance, this would be replaced by the statement: if the operator \mathcal{O} can create a state with energy ω from the vacuum with probability Z , then its Green's function has a pole at that frequency, with residue Z .

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

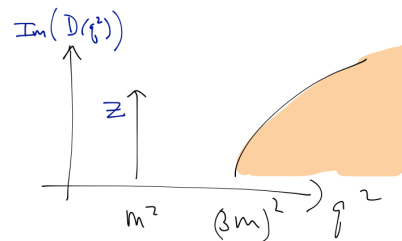
Now what about multiparticle states? The associated sum over such states involves multiple (spatial) momentum integrals, not fixed by the total momentum *e.g.* in ϕ^4

theory, ϕ can make a 3-particle state: $q \rightarrow$  and the three particles must share the momentum q . In this case the sum over all 3-particle states is

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

(Note that I am not saying that a single *real* ϕ particle is decaying to three real ϕ particles; that can't happen if they are massive. Rather, in the diagram  you can think of the particle with momentum q as virtual.)

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 ϕ^3 theory, we would instead find that the particle can decay into two particles, and the sum over two particle states would look like

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

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

Recall that for real x the imaginary part of a function of one variable with a branch cut, (like $\text{Im}(x + i\epsilon)^\nu = \frac{1}{2}((x + i\epsilon)^\nu - (x - i\epsilon)^\nu)$) is equal to (half) the *discontinuity* of the function $((x)^\nu)$ across the branch cut. The discontinuity goes to zero as we approach the branch point. Near the multi-particle continuum, the Green's function has such a branch cut.

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

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

(valid if $\text{Im}G(\omega)$ is analytic in the UHP of ω 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

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

The imaginary part determines the whole function.

Comments:

- The spectral density $\text{Im}\mathcal{D}(q)$ *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

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

still has a pole at $q^2 = m^2$ and a cut starting at the three-particle threshold, $q^2 = (3m)^2$.

- A sometimes useful fact which we’ve basically already shown (for real operators):

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

We can summarize what we’ve learned in the Lorentz-invariant case as follows: In a Lorentz invariant theory, the spectral density ρ for a scalar operator ϕ is a scalar function of p^μ with

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

Claims:

- $\rho(s) = \mathcal{N} \text{Im}\mathcal{D}$ for some number \mathcal{N} (I believe $\mathcal{N} = \pi$ here), 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

$$\mathcal{D}(k^2) = \int ds \rho(s) \frac{1}{k^2 - s + i\epsilon}. \quad (2.5)$$

This is called the Källén-Lehmann spectral representation of the propagator; it represents it as a sum of *free* propagators with different masses, determined by the spectral density.

In particular, this result (2.5) implies that $\mathcal{D}(z = k^2)$ is an analytic function in the complex z -plane away from the support of ρ , *i.e.* away from the momenta where physical states live. Singularities of amplitudes come only from physics.

One consequence (assuming unitarity and Lorentz symmetry) is that at large $|k^2|$, the Green's function is bigger than $\frac{1}{k^2}$, since each term in the integral goes like $\frac{1}{k^2}$ and $\rho(s) \geq 0$ means that there cannot be cancellations between each $\frac{1}{k^2 - s}$ contribution. This means that if the kinetic term for your scalar field has more derivatives, something must break at short distances. Breaking Lorentz symmetry is the easiest way out, for example on a lattice; in a Lorentz-invariant theory, this is an indication that non-renormalizable terms imply more degrees of freedom at high energy. (More on this in subsection §2.2.) For example, consider the theory with Lagrangian $L = (\partial\phi)^2 + \frac{1}{\Lambda^2} (\partial^2\phi)^2$. It's quadratic so we can solve it, and the propagator is

$$\frac{1}{k^2 + k^4/\Lambda^2} = \frac{1}{k^2} - \frac{1}{k^2 - \Lambda^2}$$

which as you can see looks just like a Pauli-Villars regulator. That is, we've added in a ghost field whose pole has a negative residue. As we've seen above, the residue of the pole in the propagator is a *probability*, and hence in a unitary theory had better be positive.

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

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

where ρ_c is just the continuum part. The pole at the particle-mass² survives interactions, with our assumption. (The value of the mass need not be the same as the bare mass!)

- **Sum rule.** Finally, suppose that the field ϕ in question is a canonical field, in the sense that

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

This is a statement both about the normalization of the field, and that its canonical momentum is its time derivative. Then²⁴

$$1 = \int_0^\infty ds \rho(s). \quad (2.7)$$

If we further assume that ϕ 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 ≥ 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.

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

2.2 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 ϕ which has a perturbative cubic interaction:

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

²⁴ Here's how to see this. For free fields (chapter 2) we have

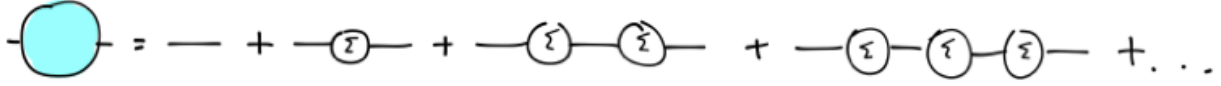
$$\langle 0 | [\phi(x), \phi(y)] | 0 \rangle_{\text{free}} = \Delta_+(x - y, m^2) - \Delta_+(y - x, m^2),$$

where $\Delta_+(x) = \int \frac{d^d p}{2\omega_p} e^{-ip \cdot x} |_{p^0 = \omega_p}$. For an interacting canonical field, we have instead a spectral representation (by exactly the methods above):

$$\langle \Omega | [\phi(x), \phi(y)] | \Omega \rangle = \int d\mu^2 \rho(\mu^2) (\Delta_+(x - y, \mu^2) - \Delta_+(y - x, \mu^2)), \quad (2.6)$$

where ρ is the same spectral density as above. Now take $\partial_{x^0} |_{x^0=y^0}$ of the BHS of (2.6) and use $\partial_t \Delta_+(x - y; \mu^2) |_{x^0=y^0} = -\frac{i}{2} \delta^{(d)}(\vec{x} - \vec{y})$.

Sum the geometric series of 1PI insertions to get

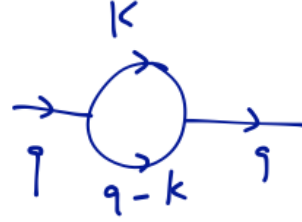


$$i\mathcal{D}_\phi(q) = \frac{i}{q^2 - m^2 - \Sigma(q) + i\epsilon}$$

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

The leading contribution to Σ comes from the one loop diagram at right and is

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



The $\frac{1}{2}$ is a symmetry factor from exchanging the two internal lines of the loop. Consider this function for real q , for which there are actual states of the scalar field – timelike q^μ , with $q^0 > m$. The real part of Σ shifts the mass. But what does it mean if this function has an imaginary part?

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

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

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

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 $\psi(t) \sim e^{-i\mathcal{E}t}$ and has norm

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

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

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

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

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

is a Lorentzian in ω with width Γ . So Γ is sometimes called a *width*.

So: what is $\text{Im}\Sigma_{1\text{ loop}}$ 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

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

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

This next trick, to get rid of the principal part bit, is from Zee’s book (the second edition on p.214; he also does the calculation by brute force in the appendix to that section). We can find a representation for the 1-loop self-energy in terms of real-space propagators: it’s the fourier transform of the amplitude to create two ϕ excitations at the origin at time zero with a single ϕ 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 ϕ field (this is $-ig$ again). Altogether:

$$\begin{aligned} \mathbf{i}\Sigma(q) &= \frac{1}{2} \int d^d x e^{\mathbf{i}q x} (-ig)^2 \mathbf{i}\mathcal{D}(x)\mathbf{i}\mathcal{D}(x) \\ &= \frac{1}{2}g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 + \mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \mathbf{i}\epsilon} \end{aligned} \quad (2.8)$$

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

$$\begin{aligned} 0 &= \frac{1}{2} \int d^d x e^{\mathbf{i}q x} (\mathbf{i}g)^2 \mathbf{i}\mathcal{D}_{\text{adv}}(x)\mathbf{i}\mathcal{D}_{\text{ret}}(x) \\ &= \frac{1}{2}g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 - \sigma_1\mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \sigma_2\mathbf{i}\epsilon} \end{aligned} \quad (2.9)$$

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 σ s). Therefore, we can add the imaginary part of zero

$$\text{Im}(\mathbf{i}0) = \frac{1}{2}g^2 \int d\Phi (\mathcal{P}_1\mathcal{P}_2 + \sigma_1\sigma_2\Delta_1\Delta_2)$$

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

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

The quantity $(1 + \sigma_1\sigma_2)$ is only nonzero (equal to 2) 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:

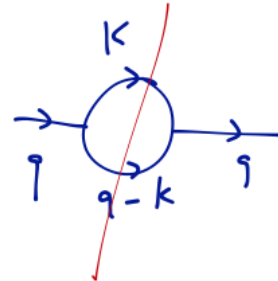
$$\begin{aligned} \text{Im}\Sigma &= \frac{1}{2}g^2 \int d\Phi 2\theta(k_1^0)\theta(k_2^0)\Delta_1\Delta_2 = \frac{1}{2}g^2 2 \int d\Phi \theta(k_1^0)\theta(k_2^0)\pi\delta(k_1^2 - m^2)\pi\delta(k_2^2 - m^2) \\ &= \frac{g^2}{2} \frac{1}{2} \int \frac{d^{D-1}\vec{k}_1}{2\omega_{\vec{k}_1}} \frac{d^{D-1}\vec{k}_2}{2\omega_{\vec{k}_2}} (2\pi)^D \delta^D(k_1 + k_2 - q). \end{aligned}$$

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

$$\text{Im}\Sigma = \frac{1}{2} \sum_{\substack{\text{actual states } n \text{ of 2 particles} \\ \text{into which } \phi \text{ can decay}}} \|\mathcal{A}_{\phi \rightarrow n}\|^2 = m\Gamma. \quad (2.10)$$

In this example the decay amplitude \mathcal{A} is just ig . And the $\frac{1}{2}$ symmetry factor matches the factor that accounts for identical particles in the final state. (The other factor of two is part of the optical theorem, as we'll see next.) In the last step we compared to our expression for the decay rate (p. 94 of [my 215A notes](#)).

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 $\theta(p^0)2\pi\delta(p^2 - m^2) = \theta(p^0)\frac{2\pi\delta(p_0 - \epsilon_p)}{2\epsilon_p}$. 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 *ics* are crucial.

There is a small but important problem with the preceding discussion (pointed out by Brian Campbell-Deem): a single ϕ particle of mass m *cannot* decay into two ϕ particles each of mass m – the kinematics of this example do not allow any final state phase space. But we can make the example viable (without changing the calculation at all) by thinking about a theory of two scalar fields, one light ϕ , one heavy Φ with lagrangian

$$\mathcal{L} = \frac{1}{2} ((\partial\Phi)^2 - M^2\Phi^2 + (\partial\phi)^2 - m^2\phi^2 - g\phi^2\Phi)$$

and thinking about the self-energy for the (unstable) heavy particle.

The general form of (2.10) is a general consequence of unitarity. Recall that the S-matrix is

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

$$\mathbf{H} = \mathbf{H}^\dagger \implies \mathbb{1} = \mathcal{S}\mathcal{S}^\dagger \implies 2\text{Im}\mathcal{T} \equiv i(\mathcal{T}^\dagger - \mathcal{T}) \stackrel{\mathbb{1}=\mathcal{S}\mathcal{S}^\dagger}{=} \mathcal{T}^\dagger\mathcal{T}.$$

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

$$2\text{Im}\mathcal{T}_{fi} = \sum_n \mathcal{T}_{fn}^\dagger \mathcal{T}_{ni}$$

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 ϕ^3 theory here, the intermediate states which can contribute to \sum_n are two-particle states, so that \sum_n will turn into $\int \frac{d\vec{k}_1}{2\omega_{k_1}} \frac{d\vec{k}_2}{2\omega_{k_2}}$, the two-particle density of states.

A bit more explicitly, introducing a basis of scattering states

$$\langle f | \mathcal{T} | i \rangle = \mathcal{T}_{fi} = \delta^4(p_f - p_i) \mathcal{M}_{fi}, \quad \mathcal{T}_{fi}^\dagger = \delta^4(p_f - p_i) \mathcal{M}_{if}^*, \quad (\text{recall that } \delta^d \equiv (2\pi)^d \delta^d)$$

we have (denoting N is the number of particles)

$$\begin{aligned} \langle F | \mathcal{T}^\dagger \mathbb{1} \mathcal{T} | I \rangle &= \sum_N \langle F | \mathcal{T}^\dagger \prod_{f=1}^N \int \frac{d^3 q_f}{2E_f} | \{q_f\} \rangle \langle \{q_f\} | \mathcal{T} | I \rangle \\ &= \sum_N \prod_{f=1}^N \int \frac{d^3 q_f}{2E_f} \delta^4(p_F - \sum_f q_f) \mathcal{M}_{\{q_f\}F}^* \delta^4(p_I - \sum_f q_f) \mathcal{M}_{\{q_f\}I} \end{aligned}$$

Now notice that we have a $\delta^4(p_F - p_I)$ on both sides, and

$$\prod_{f=1}^N \int \frac{d^3 q_f}{2E_f} \delta^4(p_F - \sum_f q_f) = \int d\Pi_N$$

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

$$\mathbf{i}(\mathcal{M}_{IF}^* - \mathcal{M}_{FI}) = \sum_N \int d\Pi_N \mathcal{M}_{\{q_f\}F}^* \mathcal{M}_{\{q_f\}I}.$$


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



$$2\text{Im}\mathcal{M}_{II} = \sum_N \int d\Pi_N |\mathcal{M}_{\{q_f\}I}|^2.$$

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

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

In more complicated examples (such as a box diagram contributing to 2-2 scattering), there can be more than one way to cut the diagram. Different ways of cutting the diagram correspond to discontinuities in different kinematical variables. To get the whole imaginary part, we have to add these up. A physical cut is a way of separating all initial-state particles from all final-state particles by cutting only internal lines. So

for example, a t -channel tree-level diagram (like ) never has any imaginary part; this makes sense because the momentum of the exchanged particle is spacelike.

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

instead we must find them as resonances in scattering amplitudes of other particles: $\text{Im} \left(\text{diagram with wavy internal line} \right)$.

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

$$\mathcal{M} = - \left(\sqrt{Z} \right)^2 \Sigma = -Z\Sigma \tag{2.11}$$

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

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{p^2 - m_0^2 - \Sigma(p)} \simeq \frac{\mathbf{i}}{(p^2 - m^2) \underbrace{(1 - \partial_{p^2} A|_{m^2})}_{=Z^{-1}} - \mathbf{i}B} = \frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}BZ}. \quad (2.12)$$

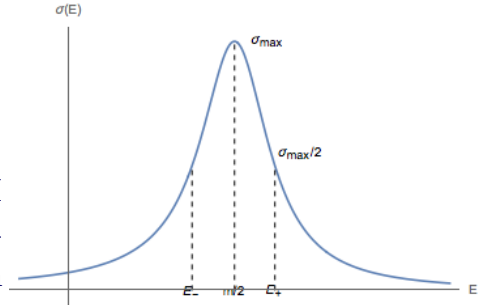
In terms of the *particle width* $\Gamma_w \equiv -ZB(m^2)/m$, this is

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

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

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

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



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

$$\Gamma_w = -\frac{BZ}{m} \stackrel{(2.11)}{=} -\frac{1}{m} (-\text{Im} \mathcal{M}_{1 \rightarrow 1}) \stackrel{\text{optical}}{=} \frac{1}{m} \frac{1}{2} \sum_n \int_f d\Pi_n |\mathcal{M}_{\{q_f\}1}|^2 = \Gamma$$

the last equation of which is exactly our formula for the decay rate. If it is not the case that $\Gamma \ll m$, *i.e.* if the resonance is too broad, the Taylor expansion of the inverse propagator we did in (2.12) 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 :

$$\sigma(E \gg \dots) \sim G^2 E^{-2-2k}. \quad (2.13)$$

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.

2.3 How to study hadrons with perturbative QCD

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

$$\sigma^{\text{anything} \leftarrow e^+e^-} \stackrel{\text{optical thm}}{=} \frac{1}{2s} \text{Im} \mathcal{M}(e^+e^- \leftarrow e^+e^-) \quad (2.14)$$

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 α (small), but to all orders in the strong coupling α_s (big at low energies), the contributions of QCD look like

$$\mathbf{i} \mathcal{M}_h = \begin{array}{c} \text{Diagram: } e^+ \text{ and } e^- \text{ lines meeting at a vertex, with a hadronic blob in the middle. Momenta } k, k_+, q, k_+ \text{ are labeled.} \end{array} = (-ie)^2 \bar{u}(k) \gamma_\mu v(k_+) \frac{-\mathbf{i}}{s} \mathbf{i} \Pi_h^{\mu\nu}(q) \frac{-\mathbf{i}}{s} \bar{v}(k_+) \gamma_\nu u(k)$$

with

$$\begin{array}{c} \text{Diagram: } \text{Hadronic vacuum polarization blob with external photon lines.} \end{array} = \mathbf{i} \Pi_h^{\mu\nu}(q) \stackrel{\text{Ward}}{=} \mathbf{i} (q^2 \eta^{\mu\nu} - q^\mu q^\nu) \Pi_h(q^2)$$

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 (2.14):

$$\sigma^{\text{hadrons} \leftarrow e^+e^-} = \frac{1}{4} \sum_{\text{spins}} \frac{\text{Im} \mathcal{M}_h}{2s} = -\frac{4\pi\alpha}{s} \text{Im} \Pi_h(s). \quad (2.15)$$

(The initial and final spins are equal and we average over initial spins. We can ignore the longitudinal term $q^\mu q^\nu$ by the Ward identity. The spinor trace is $\sum_{\text{spins}} \bar{u}(k) \gamma_\mu v(k_+) \bar{v}(k_+) \gamma^\mu u(k) =$

$-2k \cdot k_+ = -s$.) As a reality check, consider the contribution from one loop of a heavy lepton of mass $M^2 \gg m_e^2$:

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

and

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

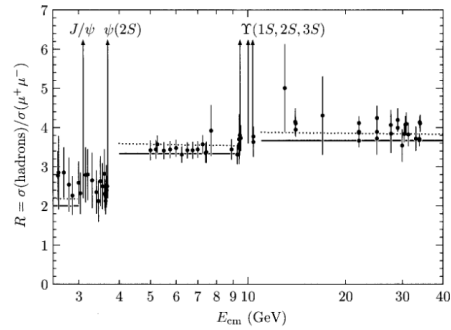
with

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

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

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

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



[End of Lecture 10]

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 Π_h is a two-point function of the quark contributions to the EM current:

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

(Here, the quark fields q_f are Dirac spinors, with Lagrangian $L_q = \sum_f \bar{q}_f (\mathbf{i}\not{D} - m_f) q_f$, $D_\mu = \partial_\mu - \mathbf{i}Q_f A_\mu + \dots$, where the \dots is the coupling to the gluon field which we’ll discuss soon enough. They have a color index which runs from 1 to 3 which I’ve suppressed.) Maybe it looks like we are taking $x \rightarrow 0$ and therefore studying short distances. But

if we are interested in large *timelike* q^μ here, that means that dominant contributions to the x integral are when the two points are timelike separated, and in the resolution of the identity in between the two J s includes physical states of QCD with lots of *real* hadrons. In contrast, the limit where we can do (maybe later we will learn how) perturbative QCD is when $q^2 = -Q_0^2 < 0$ is spacelike. (Preview: We can use the operator product expansion of the two currents.)

How can we use this knowledge to find the answer in the physical regime of $q^2 > 0$? The fact that Π_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:

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

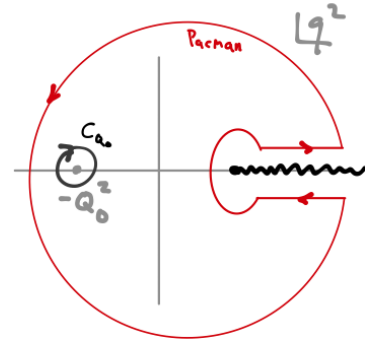
The idea here is that the RHS can be computed by perturbative QCD.

But non-analyticities in $\Pi(q^2)$ only come from physical reasons – intermediate states going on-shell. Since the physical states all have $q^2 > 0$, we can deform the contour freely away from the branch cut.

On the other hand, we know from the (appropriate generalization to currents of the) spectral representation sum rule (2.7) that $\Pi_h(q^2) \stackrel{|q|\gg\dots}{\lesssim} \log(q^2)$, so for $n \geq 1$, the contour at infinity can be ignored.

Therefore

$$\begin{aligned} I_n &= -4\pi\alpha \oint \frac{dq^2}{2\pi i} \frac{\Pi_h(q^2)}{(q^2 + Q_0^2)^{n+1}} \\ &= -4\pi\alpha \int \frac{dq^2}{4\pi i} \frac{\text{Disc}\Pi_h}{(q^2 + Q_0^2)^{n+1}} \\ &\stackrel{(2.15)}{=} \frac{1}{\pi} \int_{s_{\text{threshold}}}^{\infty} ds \frac{s}{(s + Q_0^2)^{n+1}} \sigma^{\text{hadrons} \leftarrow e^+ e^-}(s). \end{aligned}$$



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

3 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, two preparatory sections: we give another parable from QM, and spend some time addressing the basic and instructive question of where do field theories come from.

3.1 A parable on integrating out degrees of freedom

Here's another parable from QM which gives some useful perspective on renormalization in QFT and on the notion of effective field theory.

[Banks p. 138] Consider a system of two coupled harmonic oscillators. We will assume one of the springs is much stiffer than the other: let's call their natural frequencies ω_0, Ω , with $\omega_0 \ll \Omega$. The euclidean-time action is

$$S[Q, q] = \int dt \left[\frac{1}{2} (\dot{q}^2 + \omega_0^2 q^2) + \frac{1}{2} (\dot{Q}^2 + \Omega^2 Q^2) + gQq^2 \right] \equiv S_{\omega_0}[q] + S_{\Omega}[Q] + S_{\text{int}}[Q, q].$$

(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 :

$$\begin{aligned} e^{-S_{\text{eff}}[q]} &:= \int [dQ] e^{-S[q,Q]} \\ &= e^{-S_{\omega_0}[q]} \langle e^{-S_{\text{int}}[Q,q]} \rangle_Q \end{aligned}$$

Here $\langle \dots \rangle_Q$ indicates the expectation value of ... in the (free) theory of Q , with the action $S_\Omega[Q]$. It is a gaussian integral (because of our choice of S_{int}):

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

This last equality is an application of the ‘fundamental theorem of path integrals,’ *i.e.* the gaussian integral. Here $J(s) \equiv gq(s)^2$. The normalization factor \mathcal{N} is independent of J and hence of q . And $G(s,t)$ is the inverse of the linear operator appearing in S_Ω , the euclidean Green’s function:

$$S_\Omega[Q] = \int ds dt Q(s)G^{-1}(s,t)Q(t).$$

More usefully, G satisfies

$$(-\partial_s^2 + \Omega^2) G(s,t) = \delta(s-t) .$$

The fact that our system is time-translation invariant means $G(s,t) = G(s-t)$. We can solve this equation in fourier space: $G(s) = \int d\omega e^{-i\omega s} G_\omega$ makes it algebraic:

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

and we have

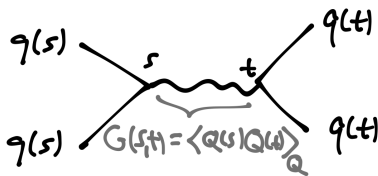
$$G(s) = \int d\omega e^{-i\omega s} \frac{1}{\omega^2 + \Omega^2} . \quad (3.1)$$

So we have:

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

or taking logs

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt ds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2 . \quad (3.2)$$



Q mediates an interaction of four qs , 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 qs what can it do besides turn back into two qs ? 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 (3.1):

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

Plug this back into (3.2):

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

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 Ω .

(3.3)

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 qs , which involve increasingly many derivatives. The first term is a quartic potential term for q : $\Delta V = \frac{g}{\Omega^2} q^4$; the leading effect of the fluctuations of Q is to *shift* the quartic self-coupling of q by a finite amount (note that we could have included a bare $\lambda_0 q^4$ potential term).

Notice that if we keep going in this expansion, we get terms with *more than two derivatives* of q . This is OK. We've just derived the right way to think about such terms: we treat them as a perturbation, and they are part of a never-ending series of terms which become less and less important for low-energy questions. If we want to ask questions about q at energies of order ω , 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 Δ at energy ω , 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

$$\sum_n c_n (\partial_t^n q)^2 q^2, \quad \text{with } c_n \sim \frac{1}{\Omega^{2n}},$$

exactly the situation where the S -matrix grows too fast at high energies that we discussed above in (2.13). 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.

3.1.1 Attempt to consolidate understanding

We've just done some coarse graining: focusing on the dofs we care about (q), and actively ignoring the dofs we don't care about (Q), except to the extent that they affect those we do (e.g. the self-interactions of q).

Above, we did a calculation in a QM model with two SHOs. This is a paradigm of QFT in many ways. For one thing, free quantum fields are bunches of harmonic oscillators with natural frequency depending on k , $\Omega = \sqrt{k^2 + m^2}$. Here we kept just two of these modes (one with large k , one with small k) for clarity. Perhaps more importantly, QM is just QFT in 0+1d. The more general QFT path integral just involves more integration variables. The idea of the Wilsonian RG (for continuum field theory) is essentially to do the integrals over the modes in descending order of wavenumber, and at each stage make the expansion described above to get a local action. And notice that basically all possible terms are generated, consistent with the symmetries (here for example, there is a \mathbb{Z}_2 symmetry under which $q \rightarrow -q$, so there are no odd powers of q).

The result of that calculation was that fluctuations of Q mediate various q^4 interactions. It adds to the action for q the following: $\Delta S_{\text{eff}}[q] \sim \int dt ds q^2(t) G(t-s) q^2(s)$, as in Fig. 3.3.

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 ω_0 , e.g.

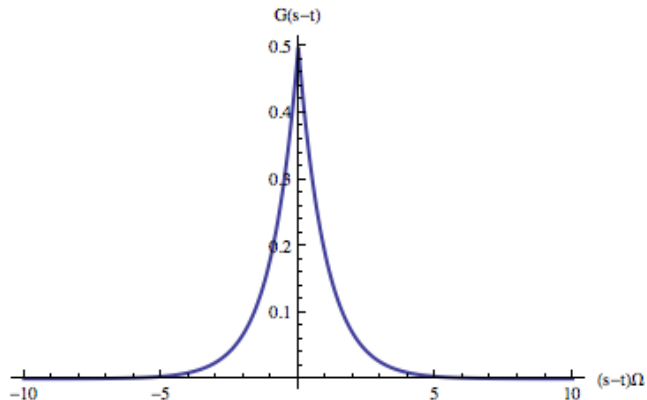
$$\langle q(t)q(0) \rangle \quad \text{with } t \sim \frac{1}{\omega_0} \gg \Omega$$

Then we can Taylor expand the function $G(t - s)$, and we find a series of corrections in powers of $\frac{1}{t\Omega}$ (or more accurately, powers of $\frac{\partial_t}{\Omega}$).

(Notice that it's not so useful to integrate out light degrees of freedom to get an action for the heavy degrees of freedom; that would necessarily be nonlocal and stay nonlocal and we wouldn't be able to treat it using ordinary techniques.)

The crucial point is that the scary non-locality of the effective action that we saw only extends a distance of order $\frac{1}{\Omega}$; the kernel $G(s - t)$ looks like this:

The mechanism we've just discussed is an essential ingredient in getting *any* physics done at all. *Why* can we do physics despite the fact that we do not understand the theory of quantum gravity which governs Planckian distances? We happily do lots of physics without worrying about this! This is because the effect of those Planckian quantum gravity fluctuations – whatever they are, call them Q – on the degrees of freedom we do care about (*e.g.* the Standard Model, or an atom, or the sandwich you made this morning, call them collectively q) are encoded in terms in the effective action of q which are suppressed by powers of the high energy scale M_{Planck} , whose role in the toy model is played by Ω . And the natural energy scale of your sandwich is much less than M_{Planck} .



I picked the Planck scale as the scale to ignore here for rhetorical drama, and because we really are ignorant of what physics goes on there. But this idea is equally relevant for *e.g.* being able to describe water waves by hydrodynamics (a classical field theory) without worrying about atomic physics, or to understand the physics of atoms without needing to understand nuclear physics, or to understand the nuclear interactions without knowing about the Higgs boson, and so on deeper into the onion of physics.

This wonderful situation, which makes physics possible, has a price: since physics at low energies is so insensitive to high energy physics, it makes it hard to learn about high energy physics! People have been very clever and have learned a lot in spite of this vexing property of the RG called decoupling. We can hope that will continue. (Cosmological inflation plays a similar role in hiding the physics of the early universe. It's like whoever designed this game is *trying* to hide this stuff from us.)

The explicit functional form of $G(s)$ (the inverse of the (euclidean) kinetic operator

for Q) is:

$$G(s) = \int \tilde{d}\omega \frac{e^{-i\omega s}}{\omega^2 + \Omega^2} = e^{-\Omega|s|} \frac{1}{2\Omega}. \quad (3.4)$$

Do it by residues: the integrand has poles at $\omega = \pm i\Omega$. The absolute value of $|s|$ is crucial, and comes from the fact that the contour at infinity converges in the upper (lower) half plane for $s < 0$ ($s > 0$).

Next, some comments about ingredients in this discussion, which provide a useful opportunity to review/introduce some important QFT technology:

- Please don't be confused by the formal similarity of the above manipulations with the construction of the generating functional of correlation functions of Q :

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

It's true that what we did above amounts precisely to constructing $Z[J]$, and plugging in $J = g_0 q^2$. But the motivation is different: in the above q is also a dynamical variable, so we don't get to pick q and differentiate with respect to it; we are merely postponing doing the path integral over q until later.

- Having said that, what is this quantity $G(s)$ above? It is the (euclidean) two-point function of Q :

$$G(s, t) = \langle Q(s)Q(t) \rangle_Q = \frac{\delta}{\delta J(t)} \frac{\delta}{\delta J(s)} \log Z[J].$$

The middle expression makes it clearer that $G(s, t) = G(s - t)$ since nobody has chosen the origin of the time axis in this problem. This euclidean Green's function, the inverse of $-\partial_\tau^2 + \Omega^2$, is unique, once we demand that it falls off at large separation (unlike the real-time Green's function).

- **Adding more labels.** Quantum mechanics is quantum field theory in 0+1 dimensions. Except for our ability to do all the integrals, everything we are doing here generalizes to quantum field theory in more dimensions: quantum field theory *is* quantum mechanics (with infinitely many degrees of freedom). With more spatial dimensions, we'll want to use the variable x for the spatial coordinates (which are just labels on the fields!) and it was in anticipation of this step that I used q instead of x for my oscillator position variables.

3.2 Where do field theories come from?

3.2.1 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

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

Let’s choose the euclidean action to be

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

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

$$Z = \int [ds] e^{-S[s]} = \sum_{\{s_i\}} e^{+\beta J \sum_{\langle i, j \rangle} s_i s_j} . \quad (3.5)$$

(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*²⁵, *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 $s_i \rightarrow -s_i$ by a nonzero *magnetization*, $\langle s_i \rangle \neq 0$.

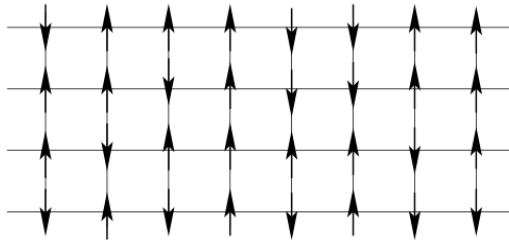


Figure 1: A configuration of classical Ising spins on the 2d square lattice. [from Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040]

Kramers-Wannier duality. To see that there is a special value of βJ , 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 βJ favors contributions from spins which agree). It is possible to rewrite the partition sum in terms of these

²⁵This 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.

disagreements. (For more on this, see [the lecture notes here](#).) The answer is identical to the original model, except with βJ 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.

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

Comment on analyticity in βJ versus the critical point. [Zee §V.3] The Ising model defined by (3.5) is a model of a magnet (more specifically, when $\beta J > 0$ which makes neighboring spins want to align, a *ferromagnet*). 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 β is a pure number (it depends on the number of spatial dimensions)²⁶. In terms of the Ising model, the magnetization is²⁷

$$\langle M \rangle = \frac{1}{Z} \sum_{\{s_i\}} e^{-H(s)/T} \frac{\sum_i s_i}{\mathcal{V}} . \quad (3.6)$$

(\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.

3.2.2 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.

²⁶The name is conventional; don't confuse it with the inverse temperature.

²⁷In many real magnets, the magnetization can point in any direction in three-space – it's a vector \vec{M} . We are simplifying our lives.

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

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

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 you may encounter called the euclidean effective action $\Gamma[\phi_c]$ (up to factors of β), where the analog of M is called the 'classical field' ϕ_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} ²⁸ then rotation invariance (or even just $M \rightarrow -M$ symmetry) demands that

$$G = V \left(r \vec{M}^2 + u \left(\vec{M}^2 \right)^2 + \dots \right)$$

where r, u 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:

$$r(T) = r_1(T - T_c) + \dots$$

with $r_1 > 0$ 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 $|\vec{M}| = \sqrt{-\frac{r}{2u}} \sim (T_c - T)^{\frac{1}{2}}$. This is the *mean field theory* description of a second-order phase transition. It's not the right value of β (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

²⁸In (3.6), 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'.

answer what is the effective *action*, $G[M(\vec{x})]$. The Landau-Ginzburg effective action is

$$G[M] = \int d^d \vec{x} \left(r \vec{M}^2 + u (\vec{M}^2)^2 + Z \partial_i \vec{M} \cdot \partial_i \vec{M} + \dots \right) \quad (3.7)$$

– 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{r} (they are all the same as a consequence of the spin rotation symmetry). So $\frac{1}{\sqrt{r}}$ is a length scale, to which we turn next.

[End of Lecture 11]

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 $M = 0$):

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

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):

$$\begin{aligned} M(x) &= \int d^3 y G_2(x, y) H(y) = \int d^3 y \left(\int d^3 k \frac{e^{i\vec{k} \cdot (\vec{x} - \vec{y})}}{k^2 + r} \right) H(y) \\ &= \int d^3 y \frac{1}{4\pi |\vec{x} - \vec{y}|} e^{-\sqrt{r} |\vec{x} - \vec{y}|} H(y). \end{aligned} \quad (3.8)$$

The Green's function

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

is diagonal in the vector index I, J so I've suppressed it in (3.8). 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* ξ , which will depend on the parameters. In the LG mean field theory, we find $\xi = \frac{1}{\sqrt{r}}$. The LG theory predicts the behavior of ξ 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 (3.5). Away from the special value of βJ , 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, ξ 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

$$M(x) \equiv \frac{\sum_{i \in R_x} \langle s_i \rangle}{\text{Vol}(R_x)} \quad (3.9)$$

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

3.2.3 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 (3.9) on the way.

The blocking (or 'decimation') transformation can be implemented in more detail for Ising spins on the 2d square lattice as follows (Fig. 2). 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:

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

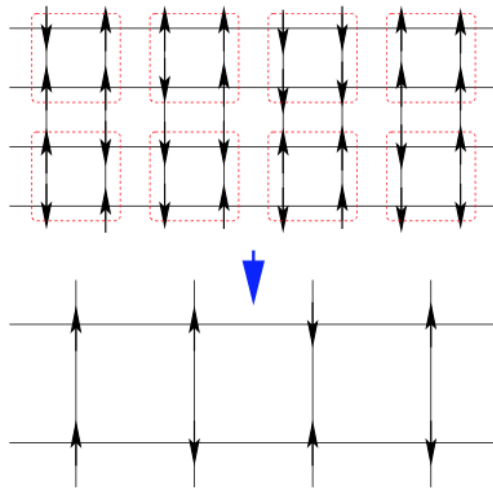


Figure 2: A blocking transformation.

[from Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040]

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

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

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:

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

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:

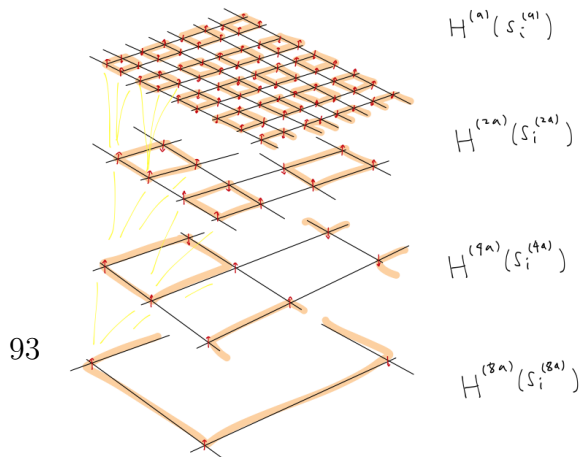
$$H^{(2a)}[s^{(2a)}] = -J^{(2a)} \sum_{\langle i, j \rangle} s_i^{(2a)} s_j^{(2a)} + -K^{(2a)} \sum_{\langle\langle i, j \rangle\rangle} s_i^{(2a)} s_j^{(2a)} + \dots$$

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*:

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



The couplings $J, K\dots$ 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.

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

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

The fixed point hamiltonian H_\star , which is not changed by the rescaling operation, is scale invariant. What can its correlation length be if it is invariant under $\xi \rightarrow \xi/2$? 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 [3.3](#).

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_* , and move away from it slightly by changing one of the couplings a little bit:

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

What does the RG do to this to leading order in δg ? The possibilities are:

- If the flow takes it back to the original fixed point, \mathcal{O} (and its associated coupling δg) is called *irrelevant*.
- If the flow takes it away from the original fixed point, \mathcal{O} is called a *relevant* perturbation of H_* .
- The new H might also be a fixed point, at least to this order in δg . 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 δg , then \mathcal{O} is called *exactly marginal*. Usually it goes one way or the other and is called *marginally relevant* or *marginally irrelevant*.

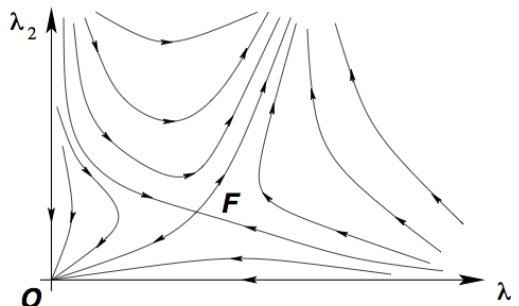


Figure 3: A possible set of RG flows for a system with two couplings $\lambda_{1,2}$. [from [Álvarez-Gaumé and Vázquez-Mozo, hep-th/0510040](#)]

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).

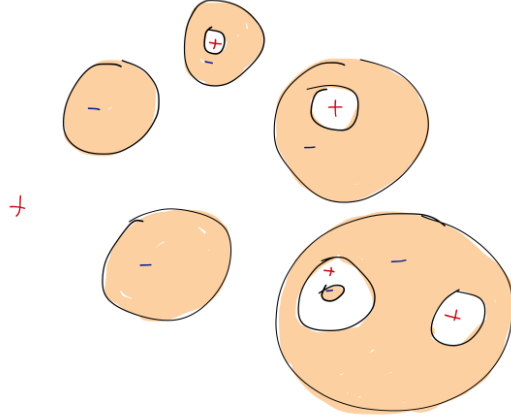
The basic point is that there is a scale-invariant field theory (often a conformal field theory) which describes the intrinsic properties of the critical point; the critical exponents are dimensions of operators in this field theory.

3.3 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 ϕ^4 theory in Euclidean space, with negative m^2 (and no ϕ^k terms with odd k). This potential has two minima and a \mathbb{Z}_2 symmetry that interchanges them, $\phi \rightarrow -\phi$. If we squint at a configuration of ϕ , we can label regions of space by the sign of ϕ (as in the figure at right). The kinetic term for ϕ 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 above²⁹.



We will study the integral

$$Z_\Lambda \equiv \int_\Lambda [D\phi] e^{-\int d^D x \mathcal{L}(\phi)}. \quad (3.10)$$

Here the specification \int_Λ says that we integrate over field configurations $\phi(x) = \int d^D k e^{ikx} \phi_k$ 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 spacing³⁰ – there just aren't modes of shorter wavelength. We are using (again) a cutoff on the euclidean momenta $k_E^2 \leq \Lambda^2$.

We want to understand (3.10) 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

²⁹ For a more sophisticated argument for this equivalence, see page 7-9 of Polyakov, *Gauge Fields and Strings*.

³⁰ This cutoff is not precisely the same as have a lattice; with a lattice, the momentum space is periodic: $e^{ikx_n} = e^{ik(na)} = e^{i(k + \frac{2\pi}{a})(na)}$ for $n \in \mathbb{Z}$. Morally it is the same.

configurations into pieces:

$$\phi(x) = \int \bar{d}k e^{ikx} \phi_k \equiv \phi^< + \phi^> .$$

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:

$$e^{-\int d^D x \delta\mathcal{L}(\phi^<)} \equiv \int [D\phi^>] e^{-\int d^D x \mathcal{L}_1(\phi^<, \phi^>)} \quad (3.11)$$

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

$$Z_\Lambda = \int_{\Lambda-\delta\Lambda} [D\phi^<] e^{-\int d^D x (\mathcal{L}(\phi^<) + \delta\mathcal{L}(\phi^<))} . \quad (3.12)$$

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

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 + \sum_n g_n \phi^n + \dots \quad (3.13)$$

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

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

(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 (3.13) again, with different coefficients? The result is to shift the couplings $g_n \rightarrow g_n + \delta g_n$. (This includes the coefficient of the kinetic term and also of the higher-derivative terms

which are hidden in the ... in (3.13). 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_{Λ} with different couplings; we accomplish this by defining

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

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 $e^{ikx} = e^{ik'x'}$. Plug this into the action

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

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

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

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

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

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

Ignore the interaction corrections, δg_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 δL term! This term can make an important difference, even in perturbation theory, for the fate of marginal operators (such as ϕ^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 (3.10) 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.

3.4 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

$$S[\phi, \phi^*] = \frac{1}{2} \sum_{n,i} |\phi(n) - \phi(n+i)|^2 + \sum_n u_0 |\phi(n)|^4 . \quad (3.14)$$

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

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

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

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

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 $a^2 \partial_\mu \phi \partial^\mu \phi^*$. The next term $\int \mathrm{d}^D k a^4 k^4 |\phi_k|^2 = \int d^D x a^4 \phi^* \sum_\mu \partial_\mu^4 \phi$, 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.

31

³¹Confession: 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.

The path integral is defined by

$$\begin{aligned}
Z &\equiv \int \underbrace{[d\phi^* d\phi]_{|k| < \Lambda_0}}_{\substack{\text{dRe } \phi(k) \text{dIm } \phi(k) \\ \pi}} e^{-S[\phi, \phi^*]} \\
&\equiv \prod_{|k| < \Lambda_0} \frac{\text{dRe } \phi(k) \text{dIm } \phi(k)}{\pi} \\
&= \prod_{|k| < \Lambda_0} \frac{d\phi^*(k) d\phi(k)}{2\pi i}
\end{aligned} \tag{3.15}$$

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

$$\phi(k) \rightarrow e^{i\theta} \phi(k), \phi^*(k) \rightarrow e^{-i\theta} \phi^*(k). \tag{3.16}$$

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

³².

[End of Lecture 12]

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

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

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 $\Lambda - \delta\Lambda \equiv \Lambda/s$, and³³ $s > 1$, we will regard s as close to 1: $s - 1 \ll 1$.

$$\begin{aligned}
Z &= \int \prod_{0 \leq |k| \leq \Lambda/s} d\phi_{<}(k) \left(\int \prod_{\Lambda/s \leq |k| \leq \Lambda} d\phi_{>}(k) e^{-\left(\underbrace{S_0[\phi^<]}_{\text{quadratic}} + S_0[\phi^>] + \underbrace{S_{\text{int}}[\phi^<, \phi^>]}_{\text{mixes fast and slow}} \right)} \right) \\
&= \int [d\phi^<] e^{-S_0[\phi^<]} \underbrace{\left\langle e^{-S_{\text{int}}[\phi^<, \phi^>]} \right\rangle_{0, >}}_{\text{average over } \phi^>, \text{ with gaussian measure}} Z_{0, >}
\end{aligned} \tag{3.17}$$

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 $|\tilde{k}| \in (0, \Lambda)$.

³²Actually, the symmetry of (3.15) is $O(2)$, since $(\phi_1, \phi_2) \rightarrow (-\phi_1, \phi_2)$ is also a symmetry and has determinant minus one.

³³I note that $s = b$ from the previous subsection; sorry for the proliferation of redundant letters.

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 rescale the field variable at each step:

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

We will choose the ‘wavefunction renormalization’ factor ζ so that the kinetic terms are fixed.

RG for free field

If $S_{\text{int}} = 0$, then (3.17) gives

$$\tilde{S}[\phi_{<}] = \int_{|k|<\Lambda/s} \mathrm{d}^D k \phi_{<}^*(k) k^2 \phi_{>}(k) \stackrel{\text{steps 2 and 3}}{=} s^{-D-2} \zeta^2 \int_{|\tilde{k}|<\Lambda} \tilde{\phi}^*(\tilde{k}) \tilde{k}^2 \tilde{\phi}(\tilde{k}) \mathrm{d}^D \tilde{k}.$$

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

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

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: $\phi(x) = \int \mathrm{d}^D k \phi(k) e^{ikx}$. 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 (3.16). We can order them by their degree in ϕ . The first nontrivial case preserving the symmetry is

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

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

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

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

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

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

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

Quartic perturbation

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

This is some shorthand notation for

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

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

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

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

$$\langle e^{-\Omega} \rangle = e^{-\langle \Omega \rangle + \frac{1}{2}(\langle \Omega^2 \rangle - \langle \Omega \rangle^2) + \dots} \quad (3.18)$$

So

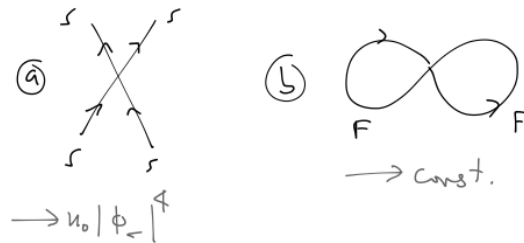
$$\tilde{\delta S} = \underbrace{\langle \delta S \rangle_{>, 0}}_{\sim u_0} - \frac{1}{2} \underbrace{\left(\langle \delta S^2 \rangle_{>, 0} - \langle \delta S \rangle_{>, 0}^2 \right)}_{\sim u_0^2} + \dots$$

So this expansion is a perturbative expansion in u_0 .

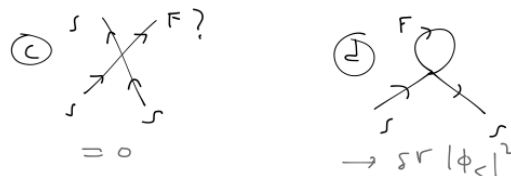
First the first term ($\sim u_0$):

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

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



- (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

$$\begin{aligned}
\widetilde{\delta S}_2(\phi_<) &= \frac{u_0}{(2!)^2} \int_{|k|<\Lambda} \langle(\phi_>^*(4)\phi_<^*(3) + \phi_>^*(3)\phi_<^*(4))(\phi_>(2)\phi_<(1) + \phi_>(1)\phi_<(2))\rangle_{0,>} \\
&= u_0 \int_{|k|<\Lambda/s} d^D k \phi_<^*(k) \phi_<(k) \cdot \underbrace{\int_{\Lambda/s}^{\Lambda} d^D p \frac{1}{p^2}}_{= \frac{\Omega_{D-1}}{(2\pi)^D} \int_{\Lambda/s}^{\Lambda} k^{D-3} dk} \\
&\stackrel{D=4}{=} \frac{2\pi^2}{(2\pi)^4} \frac{\Lambda^2}{2} (1 - s^{-2}).
\end{aligned} \tag{3.19}$$

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

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

The correction to the mass is of order the cutoff.

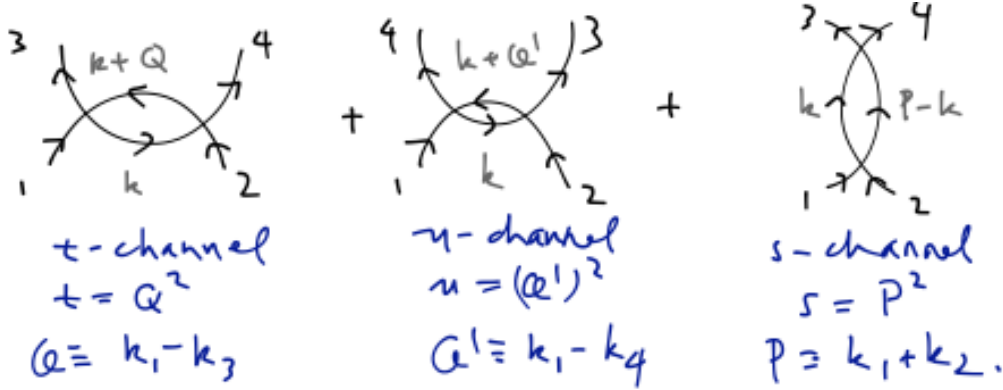
In D dimensions, we get instead

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

The next term in the cumulant expansion

Now for the $\mathcal{O}(u_0^2)$ term in $\widetilde{\delta S}$. The diagrammatic representation of $\frac{1}{2} (\langle\delta S^2\rangle - \langle\delta S\rangle^2)$ 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 $u = u_0 + u_1 k^2 + \dots$. We care about the sign of δu_0 , because in $D = 4$ it is marginal. Even small corrections will make a big difference.

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

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

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:

$$\tilde{u}(k=0) = \tilde{u}_0 = u_0 - u_0^2 \underbrace{\frac{5}{2} \int_{d\Lambda} \frac{k^3 dk}{k^4}}_{=\log s} \cdot \underbrace{\frac{\Omega_3}{(2\pi)^4}}_{=\frac{1}{8\pi^2}}.$$

The bit which depends on the external momenta k can be expanded in a series in k^2 ; these terms produce things like $\phi^* \nabla^2 \phi \phi^* \phi$, which are irrelevant in $D > 2$.

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 ℓ is the IR.

$$\begin{cases} \frac{du_0}{d\ell} = -\frac{5}{16\pi^2} u_0^2 \equiv -b u_0^2 \\ \frac{d\hat{r}_0}{d\ell} = 2\hat{r}_0 + \frac{u_0}{16\pi^2} = 2r_0 + a u_0 \end{cases} \quad (3.20)$$

Here $a, b > 0$ are constants, and $\hat{r}_0 \equiv r_0 \Lambda^2$ is the mass² 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:

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

u_0 is a marginally irrelevant perturbation of the gaussian fixed point. This theory is not asymptotically free³⁴ 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:

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

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 $r_0(0) = -\frac{a}{2}u_0(0)$.

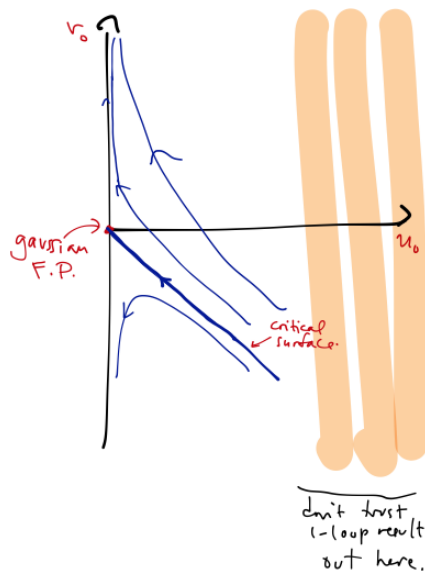
Following Wilson and Fisher, it is an extremely good idea to consider dimensions other than 4, $D \equiv 4 - \epsilon$. We've already been willing to do this as a regulator of short-distance physics; it turns out that it also resolves some short-distance physics in the phase diagram. If $D \neq 4$, the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

$$\frac{du_0}{d\ell} = \epsilon u_0 - bu_0^2. \quad (3.21)$$

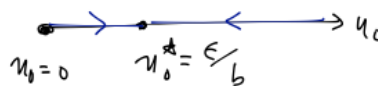
For $\epsilon > 0$ ($D < 4$) there is another fixed point at $u_0^* = \epsilon/b > 0$. And in fact the Gaussian FP is unstable, and this *Wilson-Fisher fixed point* is the stable one in the IR (see fig at right, which is drawn along the critical surface leading to $r_0(\infty) = 0$). This situation allows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

As $\epsilon \rightarrow 0$, 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



$\epsilon = 4 - D > 0$:



³⁴This 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.

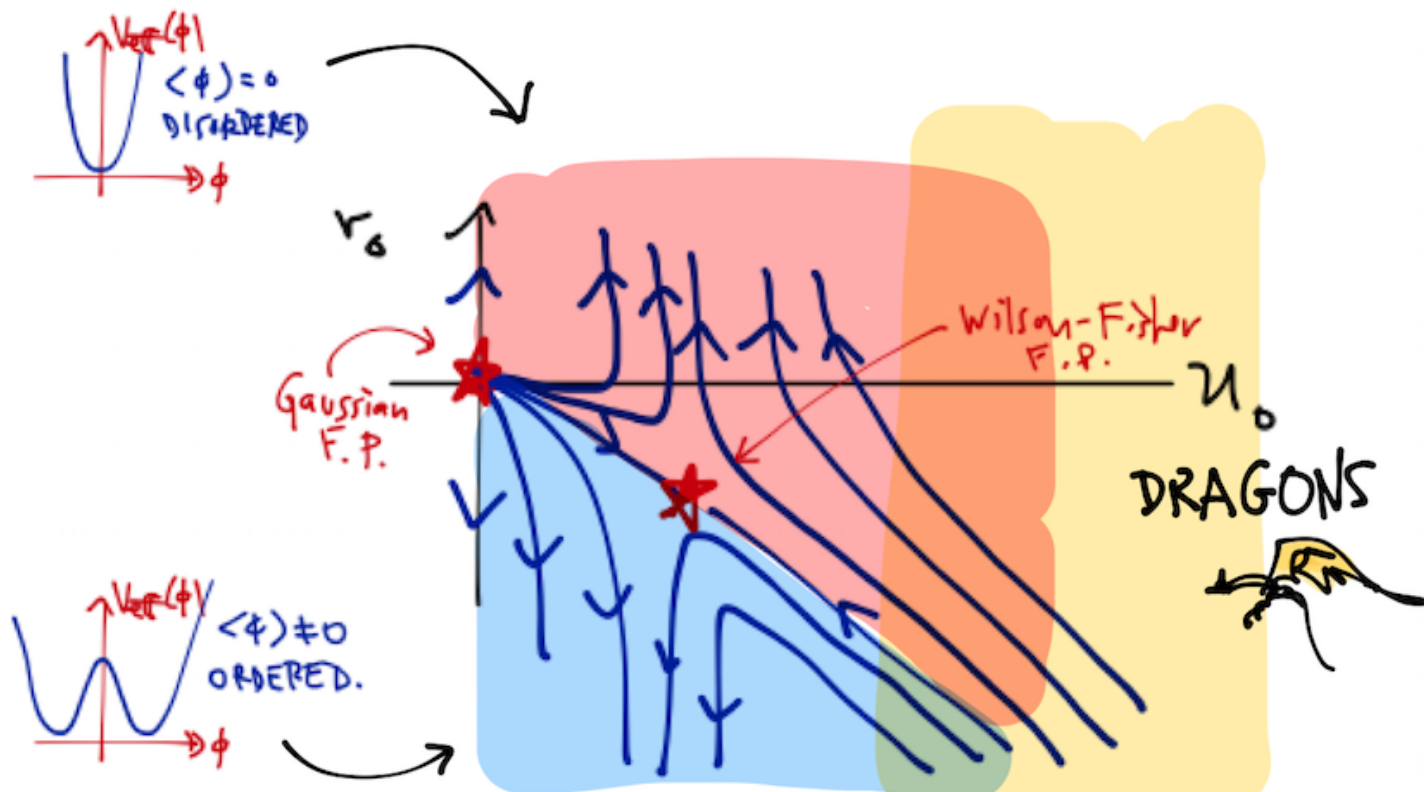


Figure 4: The ϕ^4 phase diagram. If $r_0(\ell = \infty) > 0$, the effective potential for the uniform ‘magnetization’ has a minimum at the origin; this is the disordered phase, where there is no magnetization. If $r_0(\ell = \infty) = V''_{\text{eff}} < 0$, the effective potential has minima away from the origin, and the groundstate breaks the symmetry (here $\phi \rightarrow e^{i\theta}\phi$); 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 $O(2)$ symmetry (called the XY model). WLOG, we can choose normalize our fields so that the coefficient β 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 ϵ 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 , (1.58). An also-correct answer that some people (*e.g.* me) find more satisfying is the following. Suppose we can define our QFT by a discrete model, defined on a discretized space (like in (3.14)). 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.

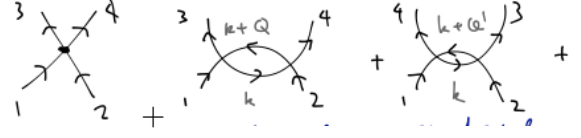
- Elimination of modes does not introduce new singularities into the couplings. At each step of the RG, we integrate out a finite-width shell in momentum space – we are doing integrals which are convergent in the infrared and ultraviolet.
- The RG plays nicely with symmetries. In particular any symmetry *of the regulated model* is a symmetry of the long-wavelength effective action.³⁵
- Some people conclude from the field theory calculation of the ϕ^4 beta function that ϕ^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’.

[End of Lecture 13]

³⁵The 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.

3.4.1 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 $\Gamma_4(k_4\dots k_1) \equiv \Gamma(4321)$, the 1PI 4-point vertex, is cutoff inde-



pendent. 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 $O(2)$ 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

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

And in particular, the bit that matters for the running of the coupling is

$$\Gamma(0000) = u_0 - u_0^2 \frac{5}{32\pi^2} \log \frac{\Lambda^2}{r_0} + \mathcal{O}(u_0^3).$$

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

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

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 $\langle \phi(k)\phi^*(k) \rangle$ 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.

3.4.2 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:

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

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 ± 1) or the liquid-gas critical point) because it simply ignored the effects of fluctuations of the modes of nonzero wavelength, *i.e.* the δL bit in (3.12). 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, η . The simplest critical exponent to understand from what we've done so far is η , the exponent associated with the anomalous dimension of the field ϕ 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 ϕ (that is, $\Gamma_2(p) = 1/\tilde{G}(p)$) as

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

where ξ is the correlation length and Λ is the UV cutoff. This looks a bit crazy – at nonzero η , 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 $e^{i\alpha X}$ the massless scalar field X in 1+1 dimensions (see the homework).

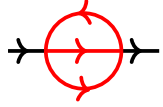
$\Gamma_2(p)$ is the 1PI momentum space 2-point vertex, *i.e.* the kinetic operator. We can interpret a nonzero η as saying that the dimension of ϕ , which in the free theory was $\Delta_0 = \frac{2-D}{2}$, has been modified by the interactions to $\Delta = \frac{2-D}{2} - \eta/2$. $\eta/2$ is the *anomalous dimension* of ϕ . 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.

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

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

It comes from the wavefunction renormalization.

In the ϕ^4 theory, $\eta = 0$ at one loop. The leading correction to η comes from the ‘sunrise’ (or ‘eyeball’) diagram at right, at two loops. (I draw the $\phi^>$ lines in red and the $\phi^<$ lines in black.) So in this model, $\eta \sim g_*^2 \sim \epsilon^2$.



[Ma, *Modern Theory of Critical Phenomena*, p. 209] Tarun Grover gave me a hard time for not emphasizing enough the fact that at the Wilson-Fisher fixed point, the anomalous dimension of the order-parameter field is nonzero – it is not a free field. He called it “the central result.” So here is another perspective on this calculation, which allows us to get the actual value. It is another example where it is easier to study field theory in real space, rather than momentum space.

Return to our expression for the correction to the effective action for the slow modes from integrating out the fast modes from the cumulant expansion, (3.18). But now write δS in position space:

$$\delta S[\phi^<] = \frac{u_0}{4} \left\langle \int d^D x |\phi^< + \phi^>|^4(x) \right\rangle_{>,0} - \frac{(u_0/4)^2}{2} \left\langle \int d^D x |\phi^< + \phi^>|^4(x) \int d^D y |\phi^< + \phi^>|^4(y) \right\rangle_{>,0,c} + \dots$$

where the subscript c indicates connected. We will look for terms in this expansion which look like $\int d^D x \partial \phi^2$. The bit from the $\mathcal{O}(u_0)$ term is of the form

$$\text{---} \text{---} \text{---} \text{---} \text{---} \text{---} = u_0 \int dx \phi(x)^2 G_>(x-x)$$

and so doesn’t give a correction to the kinetic term, only to the mass, as expected.

The terms with two slow modes involve six fast modes, and have exactly the form

of the eyeball diagram above (but now interpreted as a position-space diagram)³⁶:

$$\delta S[\phi^<] \ni \text{diagram} = -\frac{n+2}{2} u_0^2 \int d^D x \int d^D y \phi^<(x) G_>(x-y)^3 \phi^<(y). \quad (3.22)$$

Since the fast modes involve only small wavelengths, their propagator must be short-ranged; therefore we can Taylor expand

$$\phi^<(y) = \phi^<(x) + \vec{r} \cdot \vec{\nabla} \phi^<(x) + \frac{1}{2} (\vec{r} \cdot \vec{\nabla})^2 \phi^<(x) + \dots, \quad (3.23)$$

where $r \equiv y - x$. This gives

$$\delta S \ni -\frac{1}{2} u_0^2 \int d^D x \phi^<(x) \int d^D y G_>(r)^3 \left(\underbrace{\phi^<(x)}_{\text{mass correction}} + \underbrace{\vec{r} \cdot \vec{\nabla} \phi^<(x)}_{\text{vanishes by rotation symmetry}} + \frac{1}{2} (\vec{r} \cdot \vec{\nabla})^2 \phi^<(x) + \dots \right). \quad (3.24)$$

So this last term is what we are looking for, and it takes the form (after an IBP)

$$+ \int d^D x r_2 (\partial \phi^<(x))^2 \delta Z$$

with

$$\delta Z = r_2^{-1} \frac{n+2}{4} u_0^2 \int d^D r \frac{r^2}{D} G_>(r)^3 \quad (3.25)$$

where we used rotation invariance: $\int d^D r r^i r^j f(|r|) = \int d^D r \frac{r^2 \delta^{ij}}{D} f(|r|)$.

Here we can evaluate $G_>$ directly in $D = 4$ (since the differences will be in the $\mathcal{O}(\epsilon^3)$ slush):

$$r_2 G_>(x-y) \equiv \int_{\Lambda/b}^{\Lambda} d^D k \frac{e^{i\mathbf{k}(x-y)}}{k^2} \quad (3.26)$$

$$= \int_{\Lambda/b}^{\Lambda} k^{4-1-2} dk \frac{\Omega_2}{(2\pi)^4} \underbrace{\int_{-1}^1 d\theta \sin^2 \theta e^{ikr \cos \theta}}_{= \frac{\pi J_1(kr)}{kr}} \quad (3.27)$$

$$= \frac{\Omega_2}{(2\pi)^4} \frac{\pi}{r} \int_{\Lambda/b}^{\Lambda} dk J_1(kr) \quad (3.28)$$

$$= \frac{4\pi}{(2\pi)^4} \frac{\pi}{r^2} (J_0(r\Lambda/b) - J_0(r\Lambda)) = \frac{1}{4\pi^2 r^2} (J_0(r\Lambda/b) - J_0(r\Lambda)) \quad (3.29)$$

³⁶Although I drew the diagrams appropriate to the XY model, in this calculation, I have not been careful about the numerical prefactor, which depends on the number of components n of the order parameter field. This prefactor directly determines the numerical factor in η at the WF fixed point, which is a universal constant of nature, like π or e , and therefore worth determining. It's my factors of two that you should watch out for.

where J_n are Bessel functions.

So we have

$$\delta Z = r_2^{-1} \frac{(n+2)u_0^2}{4D} \left(\frac{1}{4\pi}\right)^3 \Omega_3 \int_0^\infty dr r^{D-1} r^2 G_>(r)^3 \quad (3.30)$$

$$\stackrel{D \rightarrow 4}{=} r_2^{-4} \frac{(n+2)u_0^2}{16} \left(\frac{1}{4\pi^2}\right)^3 2\pi^2 \int_0^\infty \frac{dr}{r} (J_0(r\Lambda/b) - J_0(r\Lambda))^3 \quad (3.31)$$

$$= r_2^{-4} \frac{(n+2)u_0^2}{2^{10}\pi^4} \int_0^\infty \frac{dr}{r} (J_0(r/b) - J_0(r))^3 \quad (3.32)$$

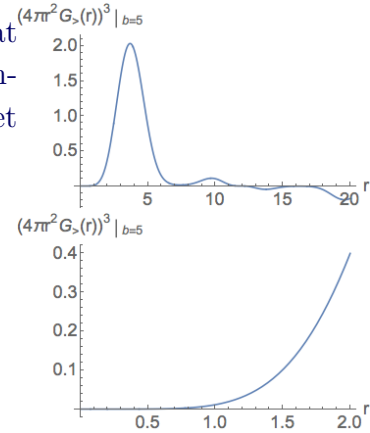
where $r \equiv r\Lambda$ and the cutoff dependence drops out.

To find the dependence on b , again the crucial idea is that the $G_>(r)$ is short-ranged; the BesselJ oscillates, but the envelope decays when its argument is of order a few. To get the idea, treat $J_0(r) \sim \theta(1-r)$, $J_0(r/b) \sim \theta(b-r)$, so

$$(4\pi r^2 G_>(r))^3 = (J_0(r/b) - J_0(r))^3 \sim \theta(r-1)\theta(b-r)$$

and

$$\int_0^\infty \frac{dr}{r} (J_0(r/b) - J_0(r))^3 \sim \int_1^b \frac{dr}{r} = \ln b.$$



The oscillations give additive corrections to this answer, which are independent of b ³⁷. Therefore, we find

$$\delta Z = r_2^{-4} c u_0^2 \ln b,$$

where c is a numerical number made of 2s and π s. The anomalous dimension of ϕ is then $\eta = \partial_{\ln b} \delta Z = c r_2^{-4} u_0^2$.

³⁷Actually, Mathematica can do the integrals

$$\int_0^\infty \frac{dr}{r} (J_0(r/b) - J_0(r)) = \ln b \quad (3.33)$$

$$\int_0^\infty \frac{dr}{r} (J_0(r/b) - J_0(r))^2 = \ln b, \quad (3.34)$$

and they both give *exactly* $\ln b$, but it doesn't like higher powers. The latter integral is the position space expression for the diagrams which correct u_0 at one loop, such as:



Correlation length exponent, ν . We now turn to the correlation length exponent, ν . Recall that 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: ξ diverges at a rate determined by the exponent ν .

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 the fixed point. The latter we will evaluate subsequently in our example. (There is actually an easier way to do this, discussed in §3.4.3, but this will be instructive.)

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

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

Under a step of the RG, $\xi_1 \rightarrow s^{-1}\xi_1$, $\delta_1 \rightarrow s^\Delta \delta_1$, where I have defined Δ 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

$$\xi = \xi_1 \left(\frac{\delta}{\delta_1} \right)^{-\frac{1}{\Delta}} \quad (3.35)$$

which is the definition of the correlation length exponent ν , 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 η . Away from the critical point, there is another scale, namely the size of the perturbation – the deviation of the knob δ from its critical value, such as $T - T_c$. Therefore, dimensional analysis says that $G(x)$ takes the form

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

where the argument of the *scaling function* Φ 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 $\xi \sim \delta^{-\frac{1}{\Delta}}$ which says $\nu = \frac{1}{\Delta}$.

In the case of ϕ^4 theory, r_0 is the parameter that an experimentalist must carefully tune to access the critical point (what I just called δ) – 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 ϕ , and we get $\nu^{-1} = \Delta_{|\phi|^2}^{(0)} = 2\frac{D-2}{2} = D-2$. 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 ϕ ! This requires a bit of a digression. [\[End of Lecture 14\]](#)

Renormalization of composite operators.

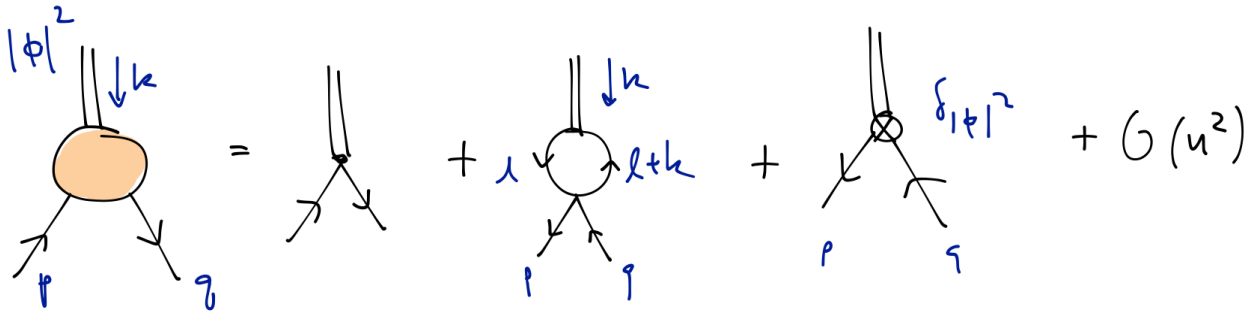
[Peskin §12.4] Perturbing the Wilson-Fisher fixed point by this seemingly-innocuous quadratic operator, is then no longer quite so innocent. In particular, we must define what we mean by the operator $|\phi|^2$! One way to define it (from the counterterms point of view, now, following Peskin and Zinn-Justin) is by adding an extra renormalization condition³⁸. 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

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

The subscript on $\mathcal{O}_\Lambda(k)$ is to emphasize that its (multiplicative) normalization is defined by a renormalization condition at scale (spacelike momentum) Λ . 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 $\mathcal{O}_\infty \equiv \phi^* \phi$ is the bare product of fields.



We can represent the implementation of this prescription diagrammatically. In the diagram above, the double line is a new kind of thing – it represents the insertion of \mathcal{O}_Λ . The vertex where it meets the two ϕ lines is *not* the 4-point vertex associated with the interaction – two ϕ s can turn into two ϕ s even in the free theory. The one-loop,

³⁸ Note that various factors differ from Peskin’s discussion in §12.4 because I am discussing a complex field $\phi \neq \phi^*$; this changes the symmetry factors.

1PI correction to this correlator is (the second diagram on the RHS of the figure)³⁹

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

where c is a number (I think it is $c = \frac{\Gamma(2-\frac{D}{2})}{(4\pi)^2}$) 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) $|\phi|_\Lambda^2$ by writing a renormalization group equation for our 3-point function

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

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

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

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

which for our example with $n = 2$ gives the anomalous dimension of $|\phi|^2$ to be (just the first term to this order since δ_Z is the wavefunction renormalization of ϕ , 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 $u_0^* = \epsilon/b$,

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

³⁹At higher order in u_0 , the wavefunction renormalization of ϕ will also contribute to the renormalization of $|\phi|^2$.

(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 lattice⁴⁰ 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 ϵ 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

$$\mathcal{O}_\Lambda^i = (Z^{-1}(\Lambda))_{ij} \mathcal{O}_\infty^j$$

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

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

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

⁴⁰To be more precise, the current results for the 3d XY model are

$$\nu^{\text{experiment}} = 0.6709(1), \quad \nu^{\text{Monte Carlo}} = 0.67169(7)$$

which disagree by 8σ ! The value which is both most reliable and most precise comes from the conformal bootstrap and is

$$\nu^{\text{conformal bootstrap}} = 0.67175(2),$$

in agreement with the MC result. For a discussion of the current state see [this paper](#).

3.4.3 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 ϕ^4 theory again, with a number of improvements:

- Instead of two components, we'll do N component fields, with $\mathcal{U} = \int d^D x u_0 (\phi^a \phi^a)^2$ (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, $\phi^a \rightarrow R_b^a \phi^b$, with $R^t R = \mathbb{1}_{N \times N}$, and perturb about the gaussian fixed point with (euclidean) action

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

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:

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

Again the $\langle \dots \rangle_{0,>}$ 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

$$\langle \phi_>^a(q_1) \phi_>^b(q_2) \rangle_{0,>} = \frac{\delta^{ab} \not\phi(q_1 + q_2)}{r_0 + q_1^2 r_2}.$$

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

$$\log \langle e^{-\mathcal{U}} \rangle_{0,>} = - \underbrace{\langle \mathcal{U} \rangle_{0,>}}_1 + \frac{1}{2} \underbrace{\left(\langle \mathcal{U}^2 \rangle_{0,>} - \langle \mathcal{U} \rangle_{0,>}^2 \right)}_2$$

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

Diagrammatically, these 16 terms decompose as in Fig. 5.

Figure 5: 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 $\mathbf{1}_3$ and $\mathbf{1}_4$ can be understood as arising from the fact that $\mathbf{1}_3$ has a symmetry which exchanges the fast lines but not the slow lines, while $\mathbf{1}_4$ 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

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

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

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

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

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. 6. The interesting part of the

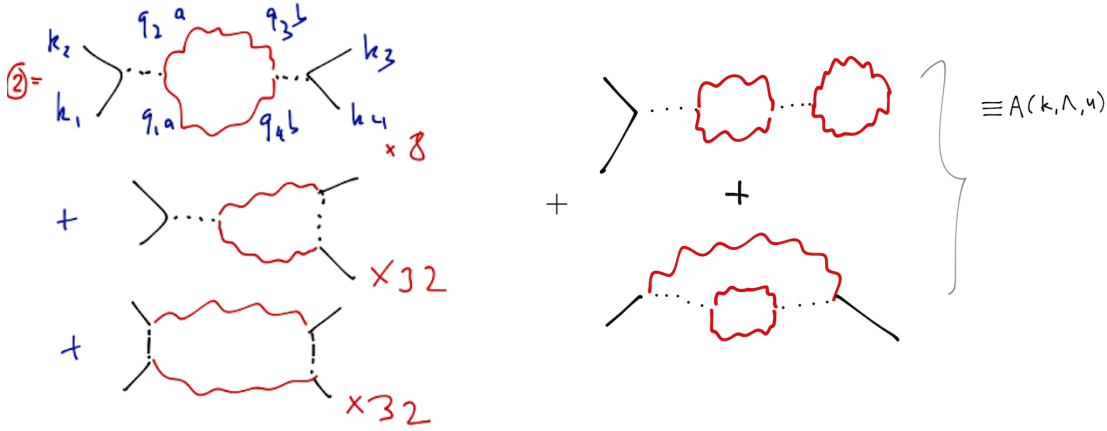


Figure 6: 2nd order corrections from the quartic perturbation of the Gaussian fixed point of the $O(N)$ model. The left column of diagrams are corrections to the quartic interaction, and the right column correct quadratic terms. In fact the top right diagram is independent of the external momentum and hence only corrects r_0 ; the bottom right diagram (that looks like a sheep) also corrects the kinetic term (along with one more I didn't draw which differs in how the flavor indices are contracted). 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 $N \rightarrow \infty$, $u_0 \rightarrow 0$, holding $u_0 N$ fixed. The quantity $u_0 N$ is often called the 't Hooft coupling.

second order bit

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

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

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

with

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

– 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

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

The shifts are:

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

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

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

Now let's make the RG step infinitesimal:

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

$$\begin{cases} \frac{dr_0}{d\ell} = 2r_0 + \frac{4(N+2)K_D \Lambda^D}{r_0 + r_2 \Lambda^2} u_0 - A u_0^2 + \mathcal{O}(u_0^3) \\ \frac{du_0}{d\ell} = (4 - D)u_0 - \frac{4(N+8)K_D \Lambda^D}{(r_0 + r_2 \Lambda^2)^2} u_0^2 + \mathcal{O}(u_0^3) \end{cases} \quad (3.36)$$

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:

$$\begin{aligned}
\tilde{r}_0 &= r_0 + \ell \frac{dr_0}{d\ell} = s^2 \left(r_0 + 4u(N+2) \int_{\Lambda/s}^{\Lambda} \frac{d^D q}{r_0 + r_2 q^2} + \mathcal{O}(u_0^2) \right) \\
&= (1+2\ell) \left(r_0 + 4u_0(N+2) \frac{\Omega_{D-1}}{(2\pi)^D} \Lambda^D \frac{1}{r_0 + r_2 \Lambda^2} \ell + \mathcal{O}(u_0^2) \right) \\
&= \left(2r_0 + \frac{4u_0(N+2)}{r_0 + r_2 \Lambda^2} K_D \Lambda^D \right) \ell + \mathcal{O}(u_0^2). \tag{3.37}
\end{aligned}$$

Now we are home. (3.36) has two fixed points. One is the free fixed point at the origin where nothing happens. The other (Wilson-Fisher) fixed point is at

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

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

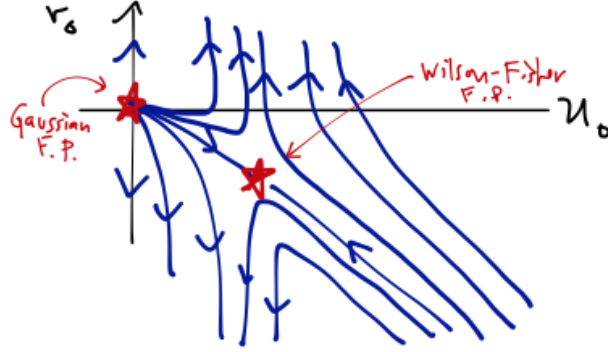


Figure 7: The ϕ^4 phase diagram, for $\epsilon > 0$.

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

$$\frac{d}{d\ell} \begin{pmatrix} \delta r_0 \\ \delta u_0 \end{pmatrix} = M_\star \begin{pmatrix} \delta r_0 \\ \delta u_0 \end{pmatrix} \tag{3.38}$$

The matrix M_\star is a 2x2 matrix whose eigensystem describes the flows near the fixed point. For the Wilson-Fisher fixed point, it looks like

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

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

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

which is a stable direction. An implicit claim I am making here is that if we included any of the other possible operators (like ϕ^6 or $\vec{\nabla}^2\phi\vec{\nabla}^2\phi$) in our action, and therefore had a bigger $K \times K$ matrix M_\star associated to K possible couplings, *all the other eigenvalues would be negative* – *i.e.* all the other operators are irrelevant at the fixed point.

So y_r determines the correlation length exponent. Its eigenvector is δr_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. (3.35)). ξ 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.

3.5 The operator product expansion and conformal perturbation theory

[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. Next we discuss a point of view which makes clear some of the data in the beta functions is universal. It also gives a more general perspective on the epsilon expansion and why it works. And it leads to the modern viewpoint on conformal field theory.

Operator product expansion (OPE). Suppose we want to understand a 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 locations $\{x_l\}$; suppose that the two operators we've picked out are closer to each other than to any of the others:

$$|x_1 - x_2| \ll |x_{1,2} - x_l|, \quad \forall l.$$

Then from the point of view of the collection Φ , $\phi_i \phi_j$ 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, *e.g.*:

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

A shorthand for this collection of statements (for any Φ) is the OPE

$$\phi_i(x_1) \phi_j(x_2) \sim \sum_k C_{ijk}(x_1 - x_2) \phi_k(x_1) \tag{3.39}$$

which is to be understood 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 special basis of operators $\{\mathcal{O}_i\}$ where

$$\langle \phi_i(x)\phi_j(0) \rangle_\star = \frac{\delta_{ij}}{r^{2\Delta_i}} \quad (3.40)$$

(here, for the simple case of scalar operators) where Δ_i is the *scaling dimension* of ϕ_i . The \star indicates that this correlator is evaluated at the fixed point. (3.40) defines the multiplicative normalizations of the ϕ_k . This basis is the same as the operators multiplying eigenvectors of the scaling matrix M_\star in (3.38), and the Δ_k are related to the eigenvalues (by $y_k = d - \Delta_k$).

Given (3.40), we can order the contributions to \sum_k in the OPE (3.39) by increasing Δ_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,

$$\mathcal{O}_i(x_1)\mathcal{O}_j(x_2) \sim \sum_k \frac{c_{ijk}}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}} \mathcal{O}_k(x_1) \quad (3.41)$$

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 (3.41) by $\mathcal{O}_k(x_3)$ and take the expectation value at the fixed point:

$$\begin{aligned} \langle \mathcal{O}_i(x_1)\mathcal{O}_j(x_2)\mathcal{O}_k(x_3) \rangle_\star &\stackrel{(3.41)}{=} \sum_{k'} \frac{c_{ijk'}}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_{k'}}} \langle \mathcal{O}_{k'}(x_1)\mathcal{O}_k(x_3) \rangle_\star \\ &\stackrel{(3.40)}{=} \frac{c_{ijk}}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}} \frac{1}{|x_1 - x_3|^{2\Delta_k}} \end{aligned} \quad (3.42)$$

(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 later but for now it is a distraction.

Conformal perturbation theory. 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.) For definiteness and simplicity let's think about the equilibrium partition function

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

– we set the temperature equal to 1 and include it in the couplings, so H is dimensionless. We can parametrize it as

$$H = H_\star + \sum_x \sum_i g_i a^{\Delta_i} \mathcal{O}_i(x) \quad (3.43)$$

where a is the short distance cutoff (*e.g.* the lattice spacing), and \mathcal{O}_i has dimensions of length $^{-\Delta_i}$ as you can check from (3.40). So g_i are de-dimensionalized couplings which we will treat as small and expand in⁴¹.

Then

$$\begin{aligned} Z &= \underbrace{Z_\star}_{\equiv \text{tr} e^{-H_\star}} \left\langle e^{-\sum_x \sum_i g_i a^{\Delta_i} \mathcal{O}_i(x)} \right\rangle_\star \\ &\stackrel{\sum_x \simeq \frac{1}{a^d} \int d^d x}{\simeq} Z_\star \left(1 - \sum_i g_i \int \langle \mathcal{O}_i(x) \rangle_\star \frac{d^d x}{a^{d-\Delta_i}} \right. \\ &\quad + \frac{1}{2} \sum_{ij} g_i g_j \int \frac{d^d x_1 d^d x_2}{a^{2d-\Delta_i-\Delta_j}} \langle \mathcal{O}_i(x_1) \mathcal{O}_j(x_2) \rangle_\star \\ &\quad \left. - \frac{1}{3!} \sum_{ijk} g_i g_j g_k \int \int \int \frac{\prod_{a=1}^3 d^d x_a}{a^{3d-\Delta_i-\Delta_j-\Delta_k}} \langle \mathcal{O}_i(x_1) \mathcal{O}_j(x_2) \mathcal{O}_k(x_3) \rangle_\star + \dots \right). \end{aligned}$$

Comments:

- We used the fact that near the fixed point, the correlation length is much larger than the lattice spacing to replace $\sum_x \simeq \frac{1}{a^d} \int d^d x$.
- There is still a UV cutoff on all the integrals – the operators can't get within a lattice spacing of each other: $|x_i - x_j| > a$.
- 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.

⁴¹Don't be put off by the word 'conformal' in the name 'conformal perturbation theory' – it just means doing perturbation theory about a general fixed point, not necessarily the gaussian one.

Now let's do the RG dance. We'll take the high-energy point of view here: while preserving Z , we make an infinitesimal change of the cutoff,

$$a \rightarrow ba = (1 + \ell)a, \quad 0 < \delta l \ll 1 .$$

The price for preserving Z is letting the couplings run $g_i = g_i(b)$. 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 $-L$ is fixed during the RG transformation, independent of b .
- The leading-in- ℓ effects of (1) and (2) are additive and so may be considered separately:

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

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

$$(2) \quad \sum_{i,j} g_i g_j \int_{|x_1-x_2| \in (a, a(1+\ell))} \frac{d^d x_1 d^d x_2}{a^{2d-\Delta_i-\Delta_j}} \underbrace{\langle \mathcal{O}_i(x_1) \mathcal{O}_j(x_2) \rangle}_* \\ = \sum_k c_{ijk} |x_1-x_2|^{\Delta_k-\Delta_i-\Delta_j} \langle \mathcal{O}_k \rangle_* \\ = \ell \sum_{ijk} g_i g_j c_{ijk} \Omega_{d-1} a^{-2d+\Delta_k} \int \langle \mathcal{O}_k \rangle_*$$

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

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

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: $\frac{dg}{d\ell} = (\delta_1 g + \delta_2 g) / \ell$

$$\frac{dg_k}{d\ell} = (d - \Delta_k)g_k - \frac{1}{2} \Omega_{d-1} \sum_{ij} c_{ijk} g_i g_j + \mathcal{O}(g^3) . \quad (3.44)$$

⁴² 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 translates our understanding of relevant and irrelevant at the initial fixed point in terms of the scaling dimensions Δ_k : g_k is relevant if $\Delta_k < d$.

⁴² 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 $\langle \rangle_*$ to obtain the result (3.44). 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 $\langle \mathcal{O}_i(x) \rangle_* = 0$. 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.

(3.44) says that to find the interaction bit of the beta function for g_k , we look at all the OPEs between operators in the perturbed hamiltonian (3.43) which produce g_k on the RHS. [End of Lecture 15]

Let's reconsider the Ising model from this point of view:

$$\begin{aligned}
H &= -\frac{1}{2} \sum_{x,x'} J(x-x') S(x) S(x') - h \sum_x S(x) \\
&\simeq -\frac{1}{2} \sum_{x,x'} J(x-x') S(x) S(x') - h \sum_x S(x) + \lambda \sum_x (S(x)^2 - 1)^2 \\
&\simeq \int d^d x \left(\frac{1}{2} (\vec{\nabla} \phi)^2 + r_0 a^{-2} \phi^2 + u_0 a^{d-4} \phi^4 + h a^{-1-d/2} \phi \right) \tag{3.45}
\end{aligned}$$

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 (3.45) I've temporarily included a Zeeman-field term hS which breaks the $\phi \rightarrow -\phi$ 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

$$H_{*,0} = \int d^d x \frac{1}{2} (\vec{\nabla} \phi)^2 .$$

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

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

It is convenient to rescale the couplings of the perturbing operators by $g_i \rightarrow \frac{2}{\Omega_{d-1}} g_i$ to remove the annoying $\Omega_{d-1}/2$ factor from the beta function equation. Then the RG equations (3.44) say

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

So we just need to know a few numbers, which we can compute by doing Wick contractions with free fields.

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

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

so that $\langle : \phi^n : \rangle = 0$, and specifically⁴³

$$\mathcal{O}_2 = \phi^2 - \langle \phi^2 \rangle, \quad \mathcal{O}_4 = \phi^4 - 6 \langle \phi^2 \rangle \phi^2 + \langle \phi^4 \rangle. \quad (3.48)$$

This amounts to a shift in couplings $r_0 \rightarrow r_0 + 3u \langle \phi^2 \rangle_*$. 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, consider a correlator of the form above:

$$\langle \mathcal{O}_n(x_1) \mathcal{O}_m(x_2) \Phi \rangle \quad \begin{array}{c} \phi_2(x_1) \text{ (circle with two dots)} \\ \phi_2(x_2) \text{ (circle with two dots)} \end{array} \quad \mathbb{Z}$$

⁴³The coefficients in (3.48) disagree with Cardy's book. Here's where these numbers come from. The self-contractions are annoying both because they are more terms, and also because they are infinite. We want to define the \mathcal{O}_n so that they are both orthonormal and finite. When I write $\langle \phi^2 \rangle$, you can imagine that I am separating the locations of the two operators by some cutoff ϵ , so $\langle \phi^2 \rangle = \langle \phi(\epsilon) \phi(0) \rangle = \epsilon^{2-d}$; the goal is to subtract off all the bits which are singular as $\epsilon \rightarrow 0$, and then take the limit.

We can do this inductively. In particular, \mathcal{O}_n is orthogonal to the identity operator $\mathcal{O}_0 = \mathbb{1}$ says $\langle \mathcal{O}_n \rangle = 0$. This fixes $\mathcal{O}_2 = \phi^2 - \langle \phi^2 \rangle$. To save writing let $G_0 \equiv \langle \phi^2(x) \rangle$. Now let

$$\mathcal{O}_4 = \phi^4 + a \phi^2 \langle \phi^2 \rangle + b \langle \phi^4 \rangle = \phi^4 + a \phi^2 G_0 + 3b G_0^2.$$

First we demand

$$0 \stackrel{!}{=} \langle \mathcal{O}_4 \rangle = 3G_0^2 + aG_0^2 + 3bG_0^2$$

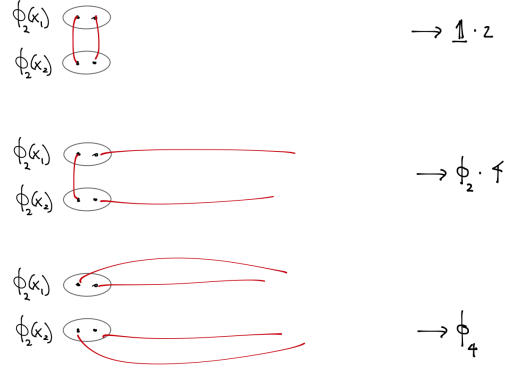
which requires $0 = 3 + a + 3b$. The next demand is that

$$0 \stackrel{!}{=} \langle \mathcal{O}_4(x) \mathcal{O}_2(0) \rangle = \langle \phi^4(x) \phi^2(0) \rangle - \langle \phi^4 \rangle G_0 + a (\langle \phi^2(x) G_0 \phi^2(0) \rangle - \langle \phi^2 \rangle G_0^2) + 3b G_0^2 \langle \mathcal{O}_2(0) \rangle \quad (3.46)$$

$$= 3G_0^3 + 12G_x^2 G_0 - 3G_0^3 + a (2G_x G_0 + G_0^3 - G_0^3) = G_x G_0 (12 + 2a) + G_0^3. \quad (3.47)$$

which requires $a = -6$, and hence $b = +1$. Notice, however, that this changes nothing about the operational definition (omit self-contractions). Thanks to Aria Yom for questioning the expression in (3.48).

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 \mathcal{O}_n with those of \mathcal{O}_m , then what's left looks like the identity operator to Φ ; that's the leading term, if it's there, since the identity has dimension 0, the lowest possible. More generally, some number of ϕ s will be left over and will need to be contracted with bits of Φ to get a nonzero correlation function. For example, the contributions to $\mathcal{O}_2 \cdot \mathcal{O}_2$ are depicted at right. In determining the combinatoric factors, note that permuting the legs on the right does not change anything, they are identical.



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):

$$\begin{cases} \mathcal{O}_2 \mathcal{O}_2 & \sim 2\mathbb{1} + 4\mathcal{O}_2 + \mathcal{O}_4 + \dots \\ \mathcal{O}_2 \mathcal{O}_4 & \sim 12\mathcal{O}_2 + 8\mathcal{O}_4 + \dots \\ \mathcal{O}_4 \mathcal{O}_4 & \sim 24\mathbb{1} + 96\mathcal{O}_2 + 72\mathcal{O}_4 + \dots \end{cases}$$

Notice that the *symmetric* operators (the ones we might add to the action preserving the symmetry) form a closed subalgebra of the operator algebra.

At $h = 0$, the result is (the $N = 1$ case of the result in §3.4.3)

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

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

The difference in numerical numbers in the values of the fixed point couplings relative to our previous calculation comes from our different parametrization (recall that we shifted the definition of r when we switched to a basis of normal-ordered operators in (3.48)) – that is not universal information. We can extract something universal and independent of our choices as follows. Linearizing the RG flow about the *new* fixed point,

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

gives

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

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

4 Effective field theory

4.1 Introduction to effective field theory

[Some nice lecture notes on effective field theory can be found here: [J. Polchinski](#), [A. Manohar](#), [I. Rothstein](#), [D. B. Kaplan](#), [H. Georgi](#), [A. Manohar](#)]

Having internalized Wilson’s perspective on renormalization – namely that we should include all possible operators consistent with symmetries and let the dynamics decide which are important at low energies – we are led immediately to the idea of an *effective field theory* (EFT), or, *how to do physics without a theory of everything*. (You may notice that all the physics that has been done has been done without a theory of everything.) It is a weaponized version of selective inattention.

The basic idea is that the Hamiltonian (or the action) should contain all terms consistent with symmetries, organized according to an expansion in decreasing relevance to low energy physics. This is an implementation of the *totalitarian principle of physics*, that anything that can happen must happen.

Diatribes about ‘renormalizability’. There is no reason to demand that a field theory that we have found to describe physics in some regime should be a valid description of the world to arbitrarily short (or long!) distances. This is a happy statement: there can always be new physics that has been so far hidden from us. Rather, an EFT comes with a regime of validity, and with necessary cutoffs. As we will discuss, in a useful implementation of an EFT, the cutoff implies a small parameter in which we can expand (and hence compute). (In the example of $S_{\text{eff}}[q]$ of the §3.1, the small parameter is ω/Ω .)

Caring about renormalizability is pretending to know about physics at arbitrarily short distances. Which you don’t.

Even when theories are renormalizable, this apparent victory is often false. For example, QED requires only two independent counterterms (for the mass and for the fine structure constant), and is therefore by the old-fashioned definition renormalizable, but it is superseded by the electroweak theory above 80GeV. Also: the coupling in QED actually increases logarithmically at shorter distances, and ultimately reaches a *Landau pole* at **SOME RIDICULOUSLY HIGH ENERGY** (of order $e^{+\frac{c}{\alpha}}$ where $\alpha \sim \frac{1}{137}$ is the fine structure constant (*e.g.* at the scale of atomic physics) and c is some numerical number. Plugging in numbers gives something like 10^{330} GeV, which is quite a bit larger than the Planck scale). This is of course completely irrelevant for physics and even in principle because of the previous remark about electroweak unification. And if not because of that, because of the Planck scale. A heartbreaking historical fact is

that Landau and many other smart people gave up on QFT as a whole because of this silly fantasy about QED in an unphysical regime.

We will see below that even in QFTs which are non-renormalizable in the strict sense, there is a more useful notion of renormalizability: effective field theories come with a small parameter (often some ratio of mass scales), in which we may expand the action. A useful EFT requires a finite number of counterterms *at each order in the expansion*.

Furthermore, I claim that this is *always* the definition of renormalizability that we are using, even if we are using a theory which is renormalizable in the traditional sense, which allows us to pretend that there is no cutoff. That is, there could always be corrections of order $\left(\frac{E}{E_{\text{new}}}\right)^n$ where E is some energy scale of physics that we are doing and E_{new} is some UV scale where new physics might come in; for large enough n , this is too small for us to have seen. The property of renormalizability that actually matters is that we need a finite number of counterterms at each order in the expansion in $\frac{E}{E_{\text{new}}}$.

Renormalizable QFTs are in some sense *less* powerful than non-renormalizable ones – the latter have the decency to tell us when they are giving the wrong answer! That is, they tell us at what energy new physics *must* come in; with a renormalizable theory we may blithely pretend that it is valid in some ridiculously inappropriate regime like 10^{330} GeV.

Notions of EFT. There is a dichotomy in the way EFTs are used. Sometimes one knows a lot about the UV theory (*e.g.*

- electroweak gauge theory,
- QCD,
- electrons in a solid,
- water molecules

...) but it is complicated and unwieldy for the questions one wants to answer, so instead one develops an effective field theory involving just the appropriate and important dofs (*e.g.*, respectively,

- Fermi theory of weak interactions (or QED or ...),
- chiral lagrangian (or HQET or SCET or hydrodynamics of quark-gluon plasma or ...),

- Landau Fermi liquid theory (or the Hubbard model or a topological field theory or ...),
- hydrodynamics (or some theory of phonons in ice or ...)

...). As you can see from the preceding lists of examples, even a single UV theory can have many different IR EFTs depending on what phase it is in, and depending on what question one wants to ask. The relationship between the pairs of theories above is always coarse-graining from the UV to the IR, though exactly what plays the role of the RG parameter can vary wildly. For example, in the case of the Fermi liquid theory, the scaling is $\omega \rightarrow 0$, and momenta scale towards the Fermi surface, not $\vec{k} = 0$.

A second situation is when one knows a description of some low-energy physics up to some UV scale, and wants to try to infer what the UV theory might be. This is a common situation in physics! Prominent examples include: the Standard Model, and quantized Einstein gravity. Occasionally we (humans) actually learn some physics and an example of an EFT from the second category moves to the first category.

Instructions for EFT. Answer the following questions:

1. what are the dofs?
2. what are the symmetries?
3. where is the cutoff, Λ , on its validity?

Then write down all interactions between the dofs which preserve the symmetries, in an expansion in derivatives, with higher-dimension operators suppressed by more powers of the UV scale, Λ .

I must also emphasize two distinct usages of the term ‘effective field theory’ which are common, and which the discussion above is guilty of conflating (this (often slippery) distinction is emphasized in the review article by Georgi linked at the beginning of this subsection). The Wilsonian perspective advocated above produces a low-energy description of the physics which is really just a way of solving (if you can) the original model; very reductively, it’s just a physically well-motivated order for doing the integrals. If you really integrate out the high energy modes exactly, you will get a non-local action for the low energy modes. This is to be contrasted with the local actions one uses in practice, by truncating the derivative expansion. It is the latter which is really the action of the effective field theory, as opposed to the *full theory*, with some of the

integrals done already. The latter will give correct answers for physics below the cutoff scale, and it will give them much more easily.

Some more comments:

- Sometimes (in condensed matter circles) this approach of just writing all terms consistent with symmetries is called Landau theory.

- Do not underestimate the difficulty of step 1 of the EFT instructions. As we'll see in some examples below, the correct low-energy dofs can look nothing at all like the microscopic dofs.

- The Wilson RG justifies this procedure: coarse graining by integrating out short-wavelength modes produces all terms consistent with the symmetries.

- When we say “what are the symmetries?” we mean the symmetries \mathbf{G} of the (regulated) microscopic theory. \mathbf{G} must be a symmetry of the low-energy EFT. Sometimes new symmetries can emerge at low energies. This procedure explains how this happens: if there are *no* relevant or marginal operators invariant under \mathbf{G} which violate a symmetry \mathbf{K} , then physics at lower and lower energies will be more and more \mathbf{K} -symmetric.

Here are some interesting and/or important examples where EFT has been useful (some of which we will discuss in more detail below) and where you can learn about them:

- Hydrodynamics [[Kovtun](#)]
- Fermi liquid theory [[J. Polchinski](#), R. Shankar, *Rev. Mod. Phys.* **66** (1994) 129]
- chiral perturbation theory [[D. B. Kaplan](#), §4]
- heavy quark effective theory [[D. B. Kaplan](#), §1.3, Manohar and Wise, *Heavy Quark Physics*]
- random surface growth (KPZ) [Zee, chapter VI]
- color superconductors [[D. B. Kaplan](#), §5]
- gravitational radiation from binary mergers [[Goldberger](#), [Rothstein](#), [Porto](#)]
- soft collinear effective theory [[Becher](#), [Stewart](#)]
- magnets [Zee, chapter VI.5, [hep-ph/9311264v1](#)]

- effective field theory of cosmological inflation [[Senatore et al](#), [Cheung et al](#), [Porto](#)]
- effective field theory of dark matter direct detection [[Fitzpatrick et al](#)]

There are many others, the length of this list was limited by how long I was willing to spend digging up references.

[End of Lecture 16]

4.2 The color of the sky

[from [hep-ph/9606222](#) and [nucl-th/0510023](#)] Why is the sky blue? Basically, it's because the blue light from the sun scatters in the atmosphere more than the red light, and you (I hope) only look at the scattered light.

Here is an understanding of this fact using the EFT logic. Consider the scattering of photons off atoms (in a gas) at low energies. Low energy means that the photon does not have enough energy to probe the substructure of the atom – it can't excite the electrons or the nuclei. This means that the atom is just a particle, with some mass M .

The dofs are just the photon field and the field that creates an atom.

The symmetries are Lorentz invariance and charge conjugation invariance and parity. We'll use the usual redundant description of the photon which has also gauge invariance.

The cutoff is the energy ΔE that it takes to excite atomic energy levels we've left out of the discussion. We allow no inelastic scattering. This means we require

$$E_\gamma \ll \Delta E \sim \frac{\alpha}{a_0} \ll a_0^{-1} \ll M_{\text{atom}} \quad (4.1)$$

Because of this separation of scales, we can also ignore the recoil of the atom, and treat it as infinitely heavy.

Since there are no charged objects in sight – atoms are neutral – gauge invariance means the Lagrangian can depend only on the field strength $F_{\mu\nu}$. Let's call the field which destroys an atom with velocity v ϕ_v . $v^\mu v_\mu = 1$ and $v_\mu = (1, 0, 0, 0)_\mu$ in the atom's rest frame. The (Lorentz-singlet) Lagrangian can depend on v^μ . We can write a Lagrangian for the free atoms as

$$L_{\text{atom}} = \phi_v^\dagger \mathbf{i} v^\mu \partial_\mu \phi_v \quad .$$

This action is related by a boost to the statement that the atom at rest has zero energy – in the rest frame of the atom, the eom is just $\partial_t \phi_{v=(1, \vec{0})} = 0$. (If we didn't define the zero of energy to be at the rest mass, there would be an additional term $\gamma_v M_{\text{atom}} \phi_v^\dagger \phi_v$, $\gamma_v \equiv \frac{1}{\sqrt{1-v^2}}$.) Notice that the kinetic term $\phi_v^\dagger \frac{\vec{\nabla}^2}{2M_{\text{atom}}} \phi_v$ is a very small correction given our hierarchy of scales (4.1).

So the Lagrangian density is

$$L_{\text{Maxwell}}[A] + L_{\text{atom}}[\phi_v] + L_{\text{int}}[A, \phi_v]$$

and we must determine L_{int} . It is made from local, Hermitian, gauge-invariant, Lorentz invariant operators we can construct out of $\phi_v, F_{\mu\nu}, v_\mu, \partial_\mu$ (it can only depend on $F_{\mu\nu} =$

$\partial_\mu A_\nu - \partial_\nu A_\mu$, and not A_μ directly, by gauge invariance, because the atom, and hence ϕ_v , is neutral.). It should actually only depend on the combination $\phi_v^\dagger \phi_v$ since we will not create and destroy atoms. (Notice that we didn't have to specify the statistics of the atoms or ϕ_v .) Therefore

$$L_{\text{int}} = c_1 \phi_v^\dagger \phi_v F_{\mu\nu} F^{\mu\nu} + c_2 \phi_v^\dagger \phi_v v^\sigma F_{\sigma\mu} v_\lambda F^{\lambda\mu} + c_3 \phi_v^\dagger \phi_v (v^\lambda \partial_\lambda) F_{\mu\nu} F^{\mu\nu} + \dots$$

... indicates terms with more derivatives and more powers of velocity (*i.e.* an expansion in $\partial \cdot v$). Which are the most important terms at low energies? Demanding that the Maxwell term dominate, we get the power counting rules (so time and space should scale the same way):

$$[\partial_\mu] = 1, \quad [F_{\mu\nu}] = 2$$

This then implies $[\phi_v] = 3/2, [v] = 0$ and therefore

$$[c_1] = [c_2] = -3, [c_3] = -4.$$

Terms with more partials are more irrelevant.

What makes up these dimensions? They must come from the length scales that we have integrated out to get this description – the size of the atom $a_0 \sim (\alpha m_e)^{-1}$ and the energy gap between the ground state and the electronic excited states $\Delta E \sim \alpha^2 m_e$. For $E_\gamma \ll \Delta E, a_0^{-1}$, we can just keep the two leading terms.

In the rest frame of the atom, these two leading terms $c_{1,2}$ represent just the scattering of E and B respectively. To determine their coefficients one would have to do a matching calculation to a more complete theory (compute transition rates in a theory that does include extra energy levels of the atom). But a reasonable guess is just that the scale of new physics (in this case atomic physics) makes up the dimensions: $c_1 \simeq c_2 \simeq a_0^3$. (In fact the magnetic term c_2 comes with extra factor of v/c which suppresses it.) The scattering cross section then goes like $\sigma \sim c_i^2 \sim a_0^6$; dimensional analysis ($[\sigma] = -2$ is an area, $[a_0^6] = -6$) then tells us that we have to make up four powers with the only other scale around:

$$\sigma \propto E_\gamma^4 a_0^6.$$

(The factor of E_γ^2 in the amplitude arises from $\vec{E} \propto \partial_t \vec{A}$.) Blue light, which has about twice the energy of red light, is therefore scattered 16 times as much.

The leading term that we left out is the one with coefficient c_3 . The size of this coefficient determines when our approximations break down. We might expect this to come from the next smallest of our neglected scales, namely ΔE . That is, we expect

$$\sigma \propto E_\gamma^4 a_0^6 \left(1 + \mathcal{O} \left(\frac{E_\gamma}{\Delta E} \right) \right).$$

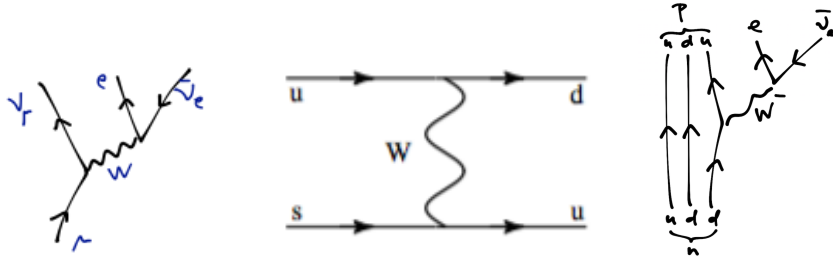
The ratio in the correction terms is appreciable for UV light.

4.3 Fermi theory of Weak Interactions

[from §5 of A. Manohar's EFT lectures] As another example of EFT, let's think about part of the Standard Model.

$$L_{EW} \ni -\frac{1}{2} (\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+) (\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}) + M_W W_\mu^+ W^{-\mu} \quad (4.2)$$

$$- \frac{ig}{\sqrt{2}} \bar{\psi}_i \gamma^\mu P_L \psi_j W_\mu^+ V_{ij} \quad + \text{terms involving } Z \text{ bosons}$$



Some things intermediate, off-shell W bosons can do: μ decay, $\Delta S = 1$ processes, neutron decay

If we are asking questions with external momenta less than M_W , we can integrate out W and make our lives simpler:

$$\delta S_{eff} \sim \left(\frac{ig}{\sqrt{2}} \right)^2 V_{ij} V_{kl}^* \int d^D p \frac{-ig_{\mu\nu}}{p^2 - M_W^2} (\bar{\psi}_i \gamma^\mu P_L \psi_j)(p) (\bar{\psi}_k \gamma^\nu P_L \psi_\ell)(-p)$$

(I am lying a little bit about the W propagator in that I am not explicitly projecting out the fourth polarization with the negative residue. Also, the W carries electric charge, so the charges of $\bar{\psi}_i$ and ψ_j in (4.2) must differ by one.) This is non-local at scales $p \gtrsim M_W$ (recall the discussion of the subsection §3.1). But for $p^2 \ll M_W^2$,

$$\frac{1}{p^2 - M_W^2} \stackrel{p^2 \ll M_W^2}{\simeq} -\frac{1}{M_W^2} \left(1 + \underbrace{\frac{p^2}{M_W^2} + \frac{p^4}{M_W^4} + \dots}_{\text{derivative couplings}} \right) \quad (4.3)$$

$$S_F = -\frac{4G_F}{\sqrt{2}} V_{ij} V_{kl}^* \int d^4 x (\bar{\psi}_i \gamma^\mu P_L \psi_j)(x) (\bar{\psi}_k \gamma_\mu P_L \psi_\ell)(x) + \mathcal{O}\left(\frac{1}{M_W^2}\right) + \text{kinetic terms for fermions} \quad (4.4)$$

where $G_F/\sqrt{2} \equiv \frac{g^2}{8M_W^2}$ is the Fermi coupling. We can use this (Fermi's) theory to compute the amplitudes above, and it is much simpler than the full electroweak theory (for example I don't have to lie about the form of the propagator of the W -boson like I


did above). It was discovered first and used quite effectively long before the existence of W s was suspected.

On the other hand, this theory is *not the same* as the electroweak theory; for example it is not renormalizable, while the EW theory is. Its point in life is to help facilitate the expansion in $1/M_W$. There is something about the expression (4.4) that should make you nervous, namely the big red **1** in the $1/M_W^2$ corrections: what makes up the dimensions? This becomes an issue when we ask about ...

4.4 Loops in EFT

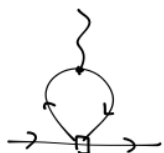
Suppose we try to define the Fermi theory S_F with a euclidean momentum cutoff $|k_E| < \Lambda$. We expect that we'll have to set $\Lambda \sim M_W$. A simple example which shows that this is problematic arises by asking about radiative corrections in the 4-Fermi theory to the coupling between the fermions and the photon (or the Z boson).

We are just trying to estimate the magnitude of this correction, so don't worry about the factors and the gamma matrices:



$$\sim I \equiv \underbrace{\frac{1}{M_W^2}}_{\propto G_F} \underbrace{\int^\Lambda d^4k \frac{1}{k} \frac{1}{k} \text{tr}(\gamma \dots)}_{\sim \int^\Lambda k dk \sim \Lambda^2 \sim M_W^2} \sim \mathcal{O}(1).$$

Even worse, consider what happens if we use the vertex coming from the $\left(\frac{p^2}{M_W^2}\right)^\ell$ correction in (4.3)



$$\sim I_\ell \equiv \frac{1}{M_W^2} \int^\Lambda d^4k \frac{1}{k^2} \left(\frac{k^2}{M_W^2}\right)^\ell \sim \mathcal{O}(1)$$

– it's also unsuppressed by powers of ... well, anything. This is a problem.

Fix: A way to fix this is to use a “mass-independent subtraction scheme”, such as dimensional regularization and minimal subtraction ($\overline{\text{MS}}$). The crucial feature is that the dimensionful cutoff parameter appears only inside logarithms ($\log \mu$), and not as free-standing powers (μ^2).

With such a scheme, we'd get instead

$$I \sim \frac{m^2}{M_W^2} \log \mu \quad I_\ell \sim \left(\frac{m^2}{M_W^2}\right)^{\ell+1} \log \mu$$

where m is some mass scale *other* than the RG scale μ (like a fermion mass parameter, or an external momentum, or a dynamical scale like Λ_{QCD}).

We will give a more detailed example next. The point is that in a mass-independent scheme, the regulator doesn't produce new dimensionful things that can cancel out the factors of M_W in the denominator. It respects the 'power counting': if you see 2ℓ powers of $1/M_W$ in the coefficient of some term in the action, that's how many powers will suppress its contributions to amplitudes. This means that the EFT is like a renormalizable theory *at each order in the expansion* (here in $1/M_W$), in that there is only a finite number of allowed vertices that contribute at each order (counterterms for which need to be fixed by a renormalization condition). The insatiable appetite for counterterms is still insatiable, but it eats only a finite number at each order in the expansion. Eventually you'll get to an order in the expansion that's too small to care about, at which point the EFT will have eaten only a finite number of counterterms.

There is a price for these wonderful features of mass-independent schemes, which has two aspects:

- Heavy particles (of mass m) don't decouple when $\mu < m$. For example, in a mass-independent scheme for a gauge theory, heavy charged particles contribute to the beta function for the gauge coupling even at $\mu \ll m$.
- Perturbation theory will break down at *low* energies, when $\mu < m$; in the example just mentioned this happens because the coupling keeps running.

We will show both these properties very explicitly in the next subsection. The solution of both these problems is to integrate out the heavy particles by hand at $\mu = m$, and make a new EFT for $\mu < m$ which simply omits that field. Processes for which we should set $\mu < m$ don't have enough energy to make the heavy particles in external states anyway. (For some situations where you should still worry about them, see Aneesh Manohar's notes linked above.)

4.4.1 Comparison of schemes, case study

The case study we will make is the contribution of a charged fermion of mass m to the running of the QED gauge coupling.

Recall that the QED Lagrangian is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \bar{\psi}(\mathbf{i}\not{D} - m)\psi$$

with $D_\mu = \partial_\mu - ieA_\mu$. By redefining the field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ by a constant factor we can move around where the e appears, *i.e.* by writing $\tilde{A} = eA$, we can make the

gauge kinetic term look like $\frac{1}{4e^2} \tilde{F}_{\mu\nu} \tilde{F}^{\mu\nu}$. This means that the charge renormalization can be seen *either* in the vacuum polarization, the correction to the photon propagator:



. I will call this diagram $i\Pi_{\mu\nu}$.

So the information about the running of the coupling is encoded in the gauge field two-point function:

$$\Pi_{\mu\nu} \equiv \langle A_\mu(p) A_\nu(q) \rangle = (p_\mu p_\nu - p^2 g_{\mu\nu}) \delta(p+q) \Pi(p^2) .$$

The factor $P_{\mu\nu} \equiv p_\mu p_\nu - p^2 g_{\mu\nu}$ is guaranteed to be the polarization structure by the gauge invariance Ward identity: $p^\mu \langle A_\mu(p) A_\nu(q) \rangle = 0$. That is: $p^\mu P_{\mu\nu} = 0$, and there is no other symmetric tensor made from p^μ which satisfies this. This determines the correlator up to a function of p^2 , which we have called $\Pi(p^2)$.

The choice of scheme shows up in our choice of renormalization condition to impose on $\Pi(p^2)$:

Mass-dependent scheme: subtract the value of the graph at $p^2 = -M^2$ (a very off-shell, euclidean, momentum). That is, we impose a renormalization condition which says

$$\Pi(p^2 = -M^2) \stackrel{!}{=} 1 \tag{4.5}$$

(which is the tree-level answer with the normalization above).

The contribution of a fermion of mass m and charge e is (factoring out the momentum-conserving delta function):

$$p, \mu \quad \text{---} \circlearrowleft \text{---} \quad p, \nu = - \int \bar{d}^D k \text{tr} \left((-ie\gamma^\mu) \frac{-i(\not{k} + m)}{k^2 - m^2} (-ie\gamma^\nu) \frac{-i(\not{p} + \not{k} + m)}{(p+k)^2 - m^2} \right)$$

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

$$i\Pi^{\mu\nu} = -e^2 \int \bar{d}^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{(\ell^2 - \Delta)^2}$$

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

$$N^{\mu\nu} = 2\ell^\mu \ell^\nu - g^{\mu\nu} \ell^2 - 2x(1-x)p^\mu p^\nu + g^{\mu\nu} (m^2 + x(1-x)p^2) + \text{terms linear in } \ell^\mu .$$

In dim reg, the one-loop vacuum polarization correction satisfies the gauge invariance Ward identity $\Pi^{\mu\nu} = P^{\mu\nu} \delta\Pi_2$ (unlike the euclidean momentum cutoff which is not gauge invariant). A peek at the tables of dim reg integrals shows that $\delta\Pi_2$ is:

$$\begin{aligned} \delta\Pi_2(p^2) &\stackrel{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\stackrel{D \rightarrow 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right) \right) \end{aligned} \quad (4.6)$$

where we have introduced the heralded μ :

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

where γ_E is the Euler-Mascheroni constant. In the second line of (4.6), we expanded the Γ -function about $D = 4$; there are other singularities at other integer dimensions.

Mass-dependent scheme: Now back to our discussion of schemes. I remind you that in a mass-independent scheme, we demand that the counterterm cancels $\delta\Pi_2$ when we set the external momentum to $p^2 = -M^2$, so that the whole contribution at order e^2 is :

$$\begin{aligned} 0 &\stackrel{(4.5)!}{=} \Pi_2^{(M)}(p^2 = -M^2) = \underbrace{\delta_{F^2}^{(M)}}_{\text{counterterm coefficient for } \frac{1}{4}F_{\mu\nu}F^{\mu\nu}} + \delta\Pi_2 \\ \implies \Pi_2^{(M)}(p^2) &= \frac{e^2}{2\pi^2} \int dx x(1-x) \log\left(\frac{m^2 - x(1-x)p^2}{m^2 + x(1-x)M^2}\right). \end{aligned}$$

Notice that the μ s go away in this scheme.

Mass-Independent scheme: This is to be contrasted with what we get in a mass-independent scheme, such as $\overline{\text{MS}}$, in which Π 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 Π_2 . Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ .) The resulting vacuum polarization function is

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

Next we will talk about beta functions, and verify the claim above about the failure of decoupling. First let me say some words about what is failing. What is failing – the

price we are paying for our power counting – is the basic principle of the RG, namely that physics at low energies shouldn't care about physics at high energies, except for small corrections to couplings. An informal version of this statement is: you don't need to know about nuclear physics to make toast. A more formal version is the *Appelquist-Carazzone Decoupling Theorem*, which I will not state (Phys. Rev. D11, 28565 (1975)). So it's something we must and will fix.

Beta functions. \boxed{M} : First in the mass-dependent scheme. The fermion contribution to the beta function for the EM coupling is

$$\beta_e^{(M)} = \frac{e}{2} M \partial_M \Pi_2(p^2) = -\frac{1}{2} \left(\frac{e^3}{2\pi} \right) \int_0^1 dx x(1-x) \left(\frac{-2M^2 x(1-x)}{m^2 + M^2 x(1-x)} \right) + \mathcal{O}(e^5)$$

$$\begin{cases} \stackrel{m \ll M}{\simeq} \frac{e^3}{2\pi^2} \int_0^1 dx x(1-x) = \frac{e^3}{12\pi^2} \\ \stackrel{m \gg M}{\simeq} \frac{e^3}{2\pi^2} \int_0^1 dx x(1-x) \frac{M^2 x(1-x)}{m^2} = \frac{e^3}{60\pi^2} \frac{M^2}{m^2} \end{cases} \quad (4.7)$$

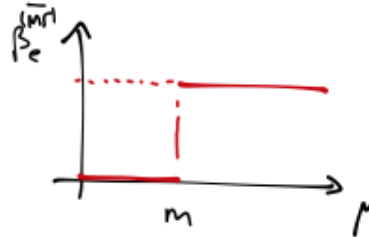
$$\boxed{\overline{\text{MS}}}: \beta_e^{(\overline{\text{MS}})} = \frac{e}{2} \mu \partial_\mu \Pi_2(p^2) = -\frac{1}{2} \frac{e^3}{2\pi^2} \underbrace{\int_0^1 dx x(1-x)}_{=1/6} \underbrace{\mu \partial_\mu \log \frac{m^2 - p^2 x(1-x)}{\mu^2}}_{=-2}$$

$$= \frac{e^3}{12\pi^2}. \quad (4.8)$$

Also, the $\overline{\text{MS}}$ vacuum polarization behaves for small external momenta like

$$\Pi_2(p^2 \ll m^2) \simeq -\frac{e^3}{2\pi^2} \int_0^1 dx x(1-x) \underbrace{\log \frac{m^2}{\mu^2}}_{\gg 1, \text{ for } \mu \ll m! \text{ bad!}}$$

As I mentioned, the resolution of both these problems is simply to define a new EFT for $\mu < m$ which omits the heavy field. Then the strong coupling problem goes away and the heavy fields do decouple. The price is that we have to do this by hand, and the beta function jumps at $\mu = m$; the coupling is continuous, though.



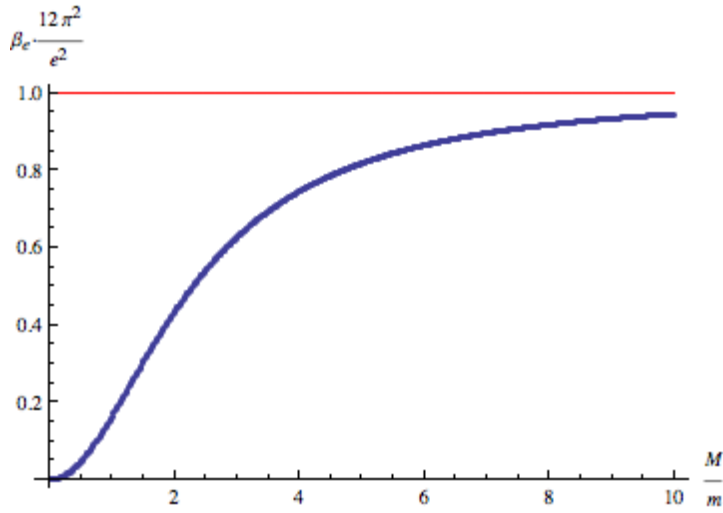
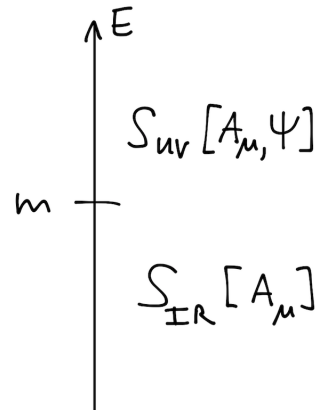


Figure 8: The blue curve is the mass-dependent-scheme beta function; at scales $M \ll m$, the mass of the heavy fermion, the fermion sensibly stops screening the charge. The red line is the $\overline{\text{MS}}$ beta function, which is just a constant, pinned at the UV value.

The couplings in the low energy EFT (here, a theory of just the photon) are determined by *matching*: this means compute a bunch of physical quantities in both descriptions, and solve for the couplings in the IR theory in terms of those of the UV theory.



4.5 The Anderson-Higgs Mechanism and Superconductors

Landau-Ginzburg EFT of superconductors: Massive vector fields as gauge fields. Consider a massive vector field B_μ with Lagrangian density

$$\mathcal{L}_B = -\frac{1}{4e^2}(dB)_{\mu\nu}(dB)^{\mu\nu} + \frac{1}{2}m^2 B_\mu B^\mu \quad (4.9)$$

where $(dB)_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$. (Note the funny-looking sign of the mass term which comes from $B^\mu B_\mu = B_0^2 - B_i^2$.) The mass term is not invariant under $B_\mu \rightarrow B_\mu + \partial_\mu \lambda$, 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 θ is a new degree of freedom. Since B is invariant under the transformation

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

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

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

is the field strength of A . The mass term becomes

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

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

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

it is the same as coupling to A_μ .

Who is θ ? Our previous point of view was that it is fake and we can just choose the gauge parameter $\lambda(x)$ to get rid of it, and set $\theta(x) = 0$ (in which case, $B = A$). This is called *unitary gauge*, and gives us back the Proca theory of $B = A$.

Consider, as an aside the following model of a single complex scalar:

$$\mathcal{L}_{\text{global}} \equiv +\frac{1}{2}|\partial_\mu \Phi|^2 - V(|\Phi|)$$

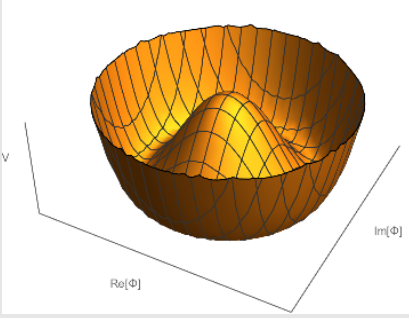
and let's take

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

for some couplings κ, v . This potential has a $U(1)$ symmetry $\Phi \rightarrow e^{i\lambda}\Phi$, and a circle of minima at $|\Phi|^2 = v^2$ (if $v^2 > 0$, which we'll assume). In polar coordinates in field space, $\Phi \equiv \rho e^{i\theta}$, the Lagrangian is

$$\mathcal{L}_{\text{global}} = +\frac{1}{2}\rho^2(\partial_\mu\theta)^2 + \frac{1}{2}(\partial\rho)^2 - V(\rho) \stackrel{E \ll V''(v)}{\simeq} \frac{1}{2}v^2(\partial\theta)^2.$$

In the last step, we observed that the excitations of the ρ field are have mass-squared $V''(v)$ about the minimum; below that energy scale, we can integrate it out and ignore it. The θ field is the massless Goldstone boson, which parametrizes the circle of minima.



Now consider the following theory, related to the previous by *gauging* the $U(1)$ symmetry:

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

where Φ is a complex, charged scalar field whose covariant derivative is $D_\mu\Phi = (\partial_\mu - \mathbf{i}A_\mu)\Phi$, with the same V as above. This is called an *Abelian Higgs model*. The $U(1)$ symmetry is *gauged*, in the sense that $A_\mu \rightarrow A_\mu + \partial_\mu\lambda(x)$, $\Phi(x) \rightarrow e^{i\lambda(x)}\Phi(x)$ is an invariance of the action, and we've learned to regard such a *local* invariance as a redundancy of the description.

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

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

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

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

That is: in the limit of large κ , the excitations of ρ are hard to make, and we get back \mathcal{L}_B . For any value of κ , we can say that the gauge field eats the would-be Goldstone boson θ and becomes heavy, in a manner consistent with gauge invariance⁴⁴. This is the Anderson-Higgs mechanism.

⁴⁴You can check that the mixing with θ is exactly what's required to make $\Pi(q)$ singular enough at $q = 0$ to give A a mass consistent with the Ward identity, as in our discussion at (1.53).

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

It is also a description of what happens to the EM field in a superconductor: the photon gets a mass; the resulting expulsion of magnetic flux is called the Meissner effect. For example, if we immerse a region $x > 0$ with $\Phi = v$ in an external constant magnetic field B_0 , $0 = \partial_\mu F^{\mu\nu} - m^2 A^\nu \implies B(x) = B e^{-mx}$. Another consequence of the mass is that if we do manage to sneak some magnetic flux into a superconductor, the flux lines will bunch up into a localized string. This can be shown by solving the equations of motion of the model above (this is a bonus problem on the homework). This is called a vortex (or vortex string in $3d$) because of what Φ does in this configuration: its phase winds around the defect. In a superconductor, the role of Φ is played by the Cooper pair field (which has electric charge two). The fact that Φ has charge two is visible in the flux quantization of the vortices (this is part of the homework bonus problem). I hope to say more about its origins in terms of electrons later in §4.5.1.

I mention here the Meissner effect and the associated collimation of flux lines partly because it is helpful for developing a picture of confinement. In particular: think about the energetics of a magnetic monopole (suppose we had one available⁴⁵) in a superconductor. If we try to insert it into a superconductor, it will trail behind it a vortex string along which all of its exiting magnetic flux is localized. This string has a finite tension (energy per unit length), as you'll study on the homework. If we make the superconducting region larger and larger, the energy of the monopole configuration grows linearly in the size – it is not a finite energy object in the thermodynamic limit. If monopoles were dynamical excitations of rest mass M_m , it would eventually become energetically favorable to pop an antimonopole out of the vacuum, so that the flux string connects the monopole to the antimonopole – this object can have finite energy inside the superconductor. But notice that in a region where electric charge is condensed, a single monopole is *confined* by the magnetic flux string it must carry around. A confining state of a gauge theory is like this for the *electric* charges, because *magnetic* charge is condensed. [End of Lecture 17]

Who is Φ ? In the last bit, we developed an effective (Landau-Ginzburg) description of superconductors which reproduces the Meissner effect (that magnetic flux is expelled

⁴⁵ [Here](#) is the paper about the only one that's been detected by humans so far.

or collimated into flux tubes); it is called the Abelian Higgs model:

$$\mathcal{F} = \frac{1}{4} F_{ij} F_{ij} + |D_i \Phi|^2 + a |\Phi|^2 + \frac{1}{2} b |\Phi|^4 + \dots \quad (4.10)$$

with $D_i \Phi \equiv (\partial_i - 2eiA_i) \Phi$. Here A is the photon field. This is a slight modification of the previous expression to indicate that the Higgs field Φ has electric charge two. We could have guessed this description by playing the EFT game, knowing that the dofs involved are the photon and a charge-two scalar field. But who is this charge-two scalar field? (Relatedly: what is the cutoff on the validity of this description?)

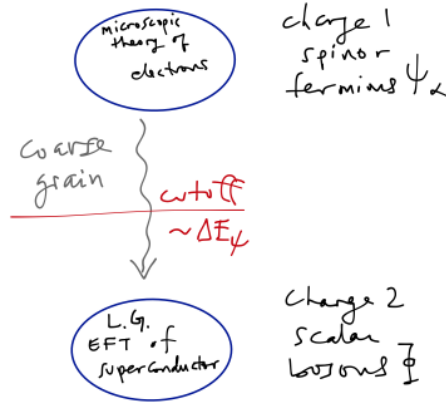
New IR dofs. A feature of this example that I want you to notice: the microscopic description of real superconductor involves electrons – charge $1e$ spinor fermions, created by some fermionic operator ψ_α , $\alpha = \uparrow, \downarrow$.

We are describing the low-energy physics of a system of electrons in terms of a bosonic field, which (in simple ‘s-wave’ superconductors) is roughly related to the electron field by

$$\Phi \sim \psi_\alpha \psi_\beta \epsilon^{\alpha\beta}; \quad (4.11)$$

Φ is called a Cooper pair field. At least, the charges and the spins and the statistics work out. The details of this relationship are not the important point I wanted to emphasize. Rather I wanted

to emphasize the dramatic difference in the correct choice of variables between the UV description (spinor fermions) and the IR description (scalar bosons). One reason that this is possible is that it costs a large energy to make a fermionic excitation of the superconductor. This can be understood roughly as follows: The microscopic theory of the electrons looks something like (ignoring the coupling to electromagnetism for now, except for a screened (and therefore short-ranged) repulsion which ultimately is the Coulomb interaction)



$$S[\psi] = S_2[\psi] + \int dt d^d x u \psi^\dagger \psi \psi^\dagger \psi + h.c. \quad (4.12)$$

where

$$S_2 = \int dt \int d^d k \psi_k^\dagger (\mathbf{i}\partial_t - \epsilon(k)) \psi_k.$$

Spin is important here so that $\psi_\uparrow^\dagger \psi_\uparrow \psi_\downarrow^\dagger \psi_\downarrow$ is nonzero. A mean field theory description of the condensation of Cooper pairs (4.11) is obtained by replacing the quartic term in (4.12) by expectation values:

$$S_{MFT}[\psi] = S_2[\psi] - \int dt d^d x u \langle \psi \psi \rangle \psi^\dagger \psi^\dagger + h.c.$$

$$= S_2[\psi] - \int dt d^d x u \Phi \psi^\dagger \psi^\dagger + h.c. \quad (4.13)$$

So an expectation value for Φ is a mass for the fermions. It is a funny kind of symmetry-breaking mass, but if you diagonalize the quadratic operator in (4.13) (actually it is done below) you will find that it costs an energy of order $\Delta E_\psi = u \langle \Phi \rangle$ to excite a fermion. That's the cutoff on the LG EFT.

A general lesson from this example is: the useful degrees of freedom at low energies can be very different from the microscopic dofs.

4.5.1 Lightning discussion of BCS.

I am sure that some of you are nervous about the step from $S[\psi]$ to $S_{MFT}[\psi]$ above. To make ourselves feel better about it, I will say a few more words about the steps from the microscopic model of electrons (4.12) to the LG theory of Cooper pairs (these steps were taken by Bardeen, Cooper and Schrieffer (BCS)).

First recall the *Hubbard-Stratonovich transformation* aka *completing the square*. In 0+0 dimensional field theory:

$$e^{-iux^4} = \frac{1}{\sqrt{i\pi u}} \int_{-\infty}^{\infty} d\sigma e^{-\frac{1}{iu}\sigma^2 - 2ix^2\sigma}. \quad (4.14)$$

At the cost of introducing an extra field σ , we turn a quartic term in x into a quadratic term in x . The RHS of (4.14) is gaussian in x and we know how to integrate it over x . (The version with \mathbf{i} is relevant for the real-time integral.) Notice the weird extra factor of \mathbf{i} lurking in (4.14). This can be understood as arising because we are trying to use a scalar field σ , to mediate a repulsive interaction (which it is, for positive u) (see Zee p. 193, 2nd Ed).

Actually, we'll need a complex H-S field:

$$e^{-iux^2\bar{x}^2} = \frac{1}{i\pi u} \int_{\mathbb{C}} d^2\sigma e^{-\frac{1}{iu}|\sigma|^2 - ix^2\bar{\sigma} + i\bar{x}^2\sigma}, \quad (4.15)$$

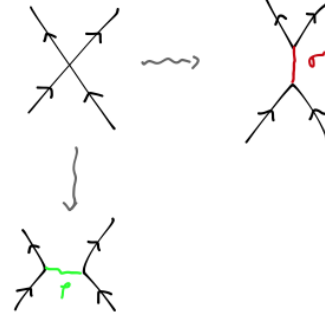
where $\int_{\mathbb{C}} d^2\sigma \dots \equiv \int_{-\infty}^{\infty} d\text{Re}\sigma \int_{-\infty}^{\infty} d\text{Im}\sigma \dots$ (The field-independent prefactor is, as usual, not important for path integrals.)

We can use a field theory generalization of (4.15) to 'decouple' the 4-fermion interaction in (4.12):

$$Z = \int [D\psi D\psi^\dagger] e^{iS[\psi]} = \int [D\psi D\psi^\dagger D\sigma D\sigma^\dagger] e^{iS_2[\psi] + i \int d^D x (\bar{\sigma} \psi_\uparrow \psi_\downarrow + h.c.) - \int d^D x \frac{|\sigma|^2(x)}{iu}}. \quad (4.16)$$

The point of this is that now the fermion integral is gaussian. At the saddle point of the σ integral (which is exact because it is gaussian), σ is the Cooper pair field, $\sigma_{\text{saddle}} = u\psi_{\uparrow}\psi_{\downarrow}$.

Notice that we made a choice here about in which ‘channel’ to make the decoupling – we could have instead introduces a different auxiliary field ρ and written $S[\rho, \psi] = \int \rho\psi^{\dagger}\psi + \int \frac{\rho^2}{2u}$, which would break up the 4-fermion interaction in the t -channel (as an interaction of the fermion density $\psi^{\dagger}\psi$) instead of the s (BCS) channel (as an interaction of Cooper pairs ψ^2). At this stage both are correct, but they lead to different mean-field approximations below. That the BCS mean field theory wins is a consequence of the RG.



How can you resist doing the fermion integral in (4.16)? Let’s study the case where the single-fermion dispersion is $\epsilon(k) = \frac{\vec{k}^2}{2m} - \mu$.

$$I_{\psi}[\sigma] \equiv \int [D\psi D\psi^{\dagger}] e^{\mathbf{i} \int dt d^d x (\psi^{\dagger} (\mathbf{i}\partial_t - \frac{\nabla^2}{2m} - \mu) \psi + \bar{\sigma} \psi \psi + \bar{\psi} \bar{\psi} \sigma)}$$

The action here can be written as the integral of

$$L = (\bar{\psi} \ \psi) \begin{pmatrix} \mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla) & \sigma \\ \bar{\sigma} & -(\mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla)) \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} \equiv (\bar{\psi} \ \psi) M \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix}$$

so the functional integral is

$$I_{\psi}[\sigma] = \det M = e^{\text{tr} \log M(\sigma)}.$$

If σ is constant (which will lower the energy), the matrix M is diagonal in momentum space, and the integral remaining to be done is

$$\int [D\sigma D\sigma^{\dagger}] e^{-\int d^D x \frac{|\sigma(x)|^2}{2iu} + \int d^D k \log(\omega^2 - \epsilon_k^2 - |\sigma|^2)}.$$

It is often possible to do this integral by saddle point. This can be justified, for example, by the largeness of the volume of the Fermi surface, $\{k | \epsilon(k) = \mu\}$, or by a large number N of species of fermions. The result is an equation which determines σ , which as we saw earlier determines the fermion gap.

$$0 = \frac{\delta \text{exponent}}{\delta \bar{\sigma}} = \mathbf{i} \frac{\sigma}{2u} + \int d\omega d^d k \frac{2\sigma}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon}.$$

We can do the frequency integral by residues:

$$\int d\omega \frac{1}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} = \frac{1}{2\pi} 2\pi \mathbf{i} \frac{1}{2\sqrt{\epsilon_k^2 + |\sigma|^2}}.$$

The resulting equation is naturally called the *gap equation*:

$$1 = -2u \int \mathrm{d}^d p' \frac{1}{\sqrt{\epsilon(p')^2 + |\sigma|^2}} \quad (4.17)$$

which you can imagine solving self-consistently for σ ⁴⁶. Plugging back into the action (4.16) says that σ determines the energy cost to have electrons around; more precisely, σ is the energy required to break a Cooper pair.

Comments:

- Notice that a solution of (4.17) requires $u < 0$, an *attractive* interaction. Superconductivity happens because the u that appears here is not the bare interaction between electrons, which is certainly repulsive (and long-ranged). This is where the phonons come in in the BCS discussion.
- If we hadn't restricted to a delta-function 4-fermion interaction $u(p, p') = u_0$ at the outset, we would have found a more general equation like

$$\sigma(\vec{p}) = -\frac{1}{2} \int \mathrm{d}^d p' \frac{u(p, p') \sigma(\vec{p}')}{\sqrt{\epsilon(p')^2 + |\sigma(p')|^2}}.$$

- A conservative perspective on the preceding calculation is that we have made a variational ansatz for the groundstate wavefunction, and the equation we solve for σ is minimizing the variational energy – finding the best wavefunction within the ansatz.
- I haven't included here effects of the fluctuations of the sigma field about its saddle point. In fact, they make the four-fermion interaction which leads to Cooper pairing marginally *relevant*. This breaks the degeneracy in deciding how to split up the $\psi\psi\psi^\dagger\psi^\dagger$ into *e.g.* $\psi\psi\sigma$ or $\psi^\dagger\psi\rho$. BCS wins. This is explained beautifully in Polchinski, lecture 2, and R. Shankar. I will summarize the EFT framework for understanding this in §4.6.

⁴⁶I should have said: and in fact one can solve it. As we will learn in the next section, the integral is dominated by the behavior near the Fermi surface, near which $\epsilon(p') \simeq v_F \ell \equiv \varepsilon$; this approximation is valid for $|\varepsilon| < E_D$, some UV cutoff on this description. The result is

$$1 = -2u \int \frac{\mathrm{d}^d p'}{\sqrt{\epsilon(p')^2 + |\sigma|^2}} \simeq -2u \int_{FS} \frac{\mathrm{d}^{d-1} k}{v_F} \int_{-E_D}^{E_D} \frac{d\varepsilon}{\sqrt{\varepsilon^2 + |\sigma|^2}} = Nu \log \left(\frac{E_D + \sqrt{E_D^2 + |\sigma|^2}}{|\sigma|} \right)$$

where $N \equiv \int_{FS} \frac{\mathrm{d}^{d-1} k}{2\pi v_F}$ is the density of states at the Fermi surface. The solution for σ is

$$|\sigma| = \frac{2E_D e^{\frac{1}{2Nu}}}{e^{\frac{1}{Nu}} - 1} \stackrel{Nu \ll 1}{\simeq} 2E_D e^{-\frac{1}{Nu}}.$$

Notice that this is non-perturbative in the coupling strength u .

- I've tried to give the most efficient introduction I could here. I left out any possibility of k -dependence or spin dependence of the interactions or the pair field, and I've conflated the pair field with the gap. In particular, I've been sloppy about the dependence on k of σ above.
- You can study a very closely related manipulation on the problem set, in examples (the $O(N)$ model and the Gross-Neveu model) where the saddle point is justified by large N .

4.5.2 Non-relativistic scalar fields

[Zee §III.5, V.1, Kaplan nucl-th/0510023 §1.2.1] In the previous discussion of the EFT for a superconductor, I just wrote the free energy, and so we didn't have to think about whether the complex scalar in question was relativistic or not.

It is not. In real superconductors, at least. How should we think about a non-relativistic field? A simple answer comes from realizing that a relativistic field which can make a boson of mass m can certainly make a boson of mass m which is moving slowly, with $v \ll c$. By taking a limit of the relativistic model, then, we can make a description which is useful for describing the interactions of an indefinite number of bosons moving slowly in some Lorentz frame. A situation that calls for such a description, for example, is a large collection of ^4He atoms.

Non-relativistic limit of a relativistic scalar field. A non-relativistic particle in a relativistic theory (like the ϕ^4 theory that we've been spending time with) has energy

$$E = \sqrt{p^2 + m^2} \stackrel{\text{if } v \ll c}{\approx} m + \frac{p^2}{2m} + \dots$$

This means that the field that creates and annihilates it looks like

$$\phi(\vec{x}, t) = \sum_{\vec{k}} \frac{1}{\sqrt{2E_{\vec{k}}}} \left(a_{\vec{k}} e^{iE_{\vec{k}}t - i\vec{k}\cdot\vec{x}} + h.c. \right)$$

In particular, we have

$$\dot{\phi}^2 \simeq m^2 \phi^2$$

and the BHS of this equation is large. To remove this large number let's change variables:

$$\phi(x, t) \equiv \frac{1}{\sqrt{2m}} \left(e^{-imt} \underbrace{\Phi(x, t)}_{\text{complex, } \Phi \ll m\Phi} + h.c. \right).$$

Notice that Φ is complex, even if ϕ is real.

Let's think about the action governing this NR sector of the theory. We can drop terms with unequal numbers of Φ and Φ^* since such terms would come with a factor of e^{imt} which gives zero when integrated over time. Starting from $(\partial\phi)^2 - m^2\phi^2 - \lambda\phi^4$ we get:

$$L_{\text{real time}} = \Phi^* \left(\mathbf{i}\partial_t + \frac{\vec{\nabla}^2}{2m} \right) \Phi - g^2 (\Phi^*\Phi)^2 + \dots \quad (4.18)$$

with $g^2 = \frac{\lambda}{4m^2}$.

Notice that Φ is a complex field and its action has a U(1) symmetry, $\Phi \rightarrow e^{i\alpha}\Phi$, even though the full theory did not. The associated conserved charge is the number of particles:

$$j_0 = \Phi^*\Phi, j_i = \frac{\mathbf{i}}{2m} (\Phi^*\partial_i\Phi - \partial_i\Phi^*\Phi), \quad \partial_t j_0 - \nabla \cdot \vec{j} = 0.$$

Notice that the 'mass term' $\Phi^*\Phi$ is then actually the chemical potential term, which encourages a nonzero density of particles to be present.

This is another example of an *emergent* symmetry (like baryon number in the SM): a symmetry of an EFT that is not a symmetry of the microscopic theory. The ... in (4.18) include terms which break this symmetry, but they are irrelevant.

To see more precisely what we mean by irrelevant, let's think about scaling. To keep this kinetic term fixed we must scale time and space differently:

$$x \rightarrow \tilde{x} = sx, \quad t \rightarrow \tilde{t} = s^2t, \quad \Phi \rightarrow \tilde{\Phi}(\tilde{x}, \tilde{t}) = \zeta\Phi(sx, s^2t).$$

A fixed point with this scaling rule has dynamical exponent $z = 2$. The scaling of the bare action (with no mode elimination step) is

$$\begin{aligned} S_E^{(0)} &= \int \underbrace{dt d^d \vec{x}}_{=s^{d+z} d\tilde{t} d^d \tilde{x}} \left(\underbrace{\Phi^*(sx, s^2t) \left(\partial_t - \frac{\vec{\nabla}^2}{2m} \right) \Phi(sx, s^2t)}_{=s^{-2} \left(\tilde{\partial}_t - \frac{\tilde{\nabla}^2}{2m} \right)} - g^2 (\Phi^*\Phi(sx, s^2t))^2 + \dots \right) \\ &= \underbrace{s^{d+z-2} \zeta^{-2}}_{\stackrel{!}{=} 1 \Rightarrow \zeta = s^{-d/2}} \int d\tilde{t} d^d \tilde{x} \left(\tilde{\Phi}^* \left(\tilde{\partial}_t - \frac{\tilde{\nabla}^2}{2m} \right) \tilde{\Phi} - \zeta^{-2} g^2 (\tilde{\Phi}^*\tilde{\Phi}(\tilde{x}, \tilde{t}))^2 + \dots \right) \quad (4.19) \end{aligned}$$

From this we learn that $\tilde{g} = s^{2-d}g \rightarrow 0$ in the IR – the quartic term is irrelevant in $D = d+1 = 3+1$ with nonrelativistic scaling! Where does it become marginal? (Hint: look back at the first lecture.)

Number and phase angle. In the NR theory, the canonical momentum for Φ is just $\frac{\partial L}{\partial \dot{\Phi}} \sim \Phi^*$, with no derivatives. This statement becomes more shocking if we change variables to $\Phi = \sqrt{\rho}e^{i\theta}$ (which would be useful *e.g.* if we knew ρ didn't want to be zero); the action density is

$$L = \frac{\mathbf{i}}{2}\partial_t\rho - \rho\partial_t\theta - \frac{1}{2m}\left(\rho(\nabla\theta)^2 + \frac{1}{4\rho}(\nabla\rho)^2\right) - g^2\rho^2. \quad (4.20)$$

The first term is a total derivative. The second term says that the canonical momentum for the phase variable θ is $\rho = \Phi^*\Phi = j_0$, the particle number density. Quantumly, then:

$$[\hat{\rho}(\vec{x}, t), \hat{\theta}(\vec{x}', t)] = \mathbf{i}\delta^d(\vec{x} - \vec{x}').$$

Number and phase are canonically conjugate variables. If we fix the phase, the amplitude is maximally uncertain.

If we integrate over space, $N \equiv \int d^d x \rho(\vec{x}, t)$ gives the total number of particles, which is time independent, and satisfies $[N, \theta] = \mathbf{i}$.

This relation explains why there's no Higgs boson in most non-relativistic superconductors and superfluids (in the absence of some extra assumption of particle-hole symmetry). In the NR theory with first order time derivative, the would-be amplitude mode which oscillates about the minimum of $V(\rho)$ is actually just the conjugate momentum for the goldstone boson!

4.5.3 Superfluids.

[Zee §V.1] Let me amplify the previous remark. A superconductor is just a superfluid coupled to an external U(1) gauge field, so we've already understood something about superfluids.

The effective field theory has the basic lagrangian (4.20), with $\langle\rho\rangle = \bar{\rho} \neq 0$. This nonzero density can be accomplished by adding an appropriate chemical potential to (4.20); up to an uninteresting constant, this is

$$L = \frac{\mathbf{i}}{2}\partial_t\rho - \rho\partial_t\theta - \frac{1}{2m}\left(\rho(\nabla\theta)^2 + \frac{1}{4\rho}(\nabla\rho)^2\right) - g^2(\rho - \bar{\rho})^2.$$

Expand around such a condensed state in small fluctuations $\sqrt{\rho} = \sqrt{\bar{\rho}} + h$, $h \ll \sqrt{\bar{\rho}}$:

$$L = -2\sqrt{\bar{\rho}}h\partial_t\theta - \frac{\bar{\rho}}{2m}\left(\vec{\nabla}\theta\right)^2 - \frac{1}{2m}\left(\vec{\nabla}h\right)^2 - 4g^2\bar{\rho}h^2 + \dots$$

Notice that h , the fluctuation of the amplitude mode, is playing the role of the canonical momentum of the goldstone mode θ . The effects of the fluctuations can be incorporated

by doing the gaussian integral over h (What suppresses self-interactions of h ?), and the result is

$$\begin{aligned} L &= \bar{\rho} \partial_t \theta \frac{1}{4g^2 \bar{\rho} - \frac{\nabla^2}{2m}} \bar{\rho} \partial_t \theta - \frac{\bar{\rho}}{2m} (\vec{\nabla} \theta)^2 \\ &= \frac{1}{4g^2} (\partial_t \theta)^2 - \frac{\bar{\rho}}{2m} (\nabla \theta)^2 + \dots \end{aligned} \quad (4.21)$$

where in the second line we are expanding in the small wavenumber k of the modes, that is, we are constructing an action for Goldstone modes whose wavenumber is $k \ll \sqrt{9g^2 \bar{\rho} m}$ so we can ignore higher gradient terms.

The linearly dispersing mode in this superfluid that we have found, sometimes called the phonon, has dispersion relation

$$\omega^2 = \frac{2g^2 \bar{\rho}}{m} k^2.$$

This mode has an emergent Lorentz symmetry with a lightcone with velocity $v_c = g\sqrt{2\bar{\rho}/m}$. The fact that the sound velocity involves g – which determined the steepness of the walls of the wine-bottle potential – is a consequence of the non-relativistic dispersion of the bosons. In the relativistic theory, in contrast, we have $L = \partial_\mu \Phi^* \partial^\mu \Phi - \kappa (\Phi^* \Phi - v^2)^2$ and we can take $\kappa \rightarrow \infty$ fixing v and still get a linearly dispersing mode by plugging in $\Phi = e^{i\theta} v$.

What does this have to do with the phenomenology of superfluids, like dissipationless flow? The importance of the linearly dispersing phonon mode of the superfluid is that there is *no other* low energy excitation of the fluid. With a classical pile of (*e.g.* non interacting) bosons, a chunk of moving fluid can donate some small momentum \vec{k} to a single boson at energy cost $\frac{(\hbar \vec{k})^2}{2m}$. A quadratic dispersion means more modes at small k than a linear one (the density of states is $N(E) \propto k^{D-1} \frac{dk}{dE}$). With only a linearly dispersing mode at low energies, there is a critical velocity below which a non-relativistic chunk of fluid cannot give up any momentum [Landau]: conserving momentum $M\vec{v} = M\vec{v}' + \hbar\vec{k}$ says the change in energy (which must be negative for this to happen on its own) is (eliminate $v' = v - \hbar k/M$):

$$\frac{1}{2} M (v')^2 + \hbar \omega(k) - \frac{1}{2} M v^2 = -\hbar k v + \frac{(\hbar k)^2}{2M} + \hbar \omega(k) = (-v + v_c) k + \frac{(\hbar k)^2}{2M}.$$

For small k , this is only negative when $v > v_c$.

You can ask: an ordinary liquid also has a linearly dispersing sound mode; why doesn't Landau's argument mean that it has superfluid flow? The answer is that it has *other* modes with softer dispersion (so more contribution at low energies), in particular diffusion modes, with $\omega \propto k^2$ (there is an important factor of i in there).

The Goldstone boson has a compact target space, $\theta(x) \equiv \theta(x) + 2\pi$, since, after all, it is the phase of the boson field. This is significant because it means that as the phase wanders around in space, it can come back to its initial value after going around the circle – such a loop encloses a *vortex*. Somewhere inside, we must have $\Phi = 0$. And actually, our discussion of the vortices of the Abelian Higgs model did not depend on the form of the time-derivative terms. There is much more to say about this.

[End of Lecture 18]

4.6 Effective field theory of Fermi surfaces

In previous subsections, we gave various descriptions of superconductors, appropriate at increasing energies. At the lowest energies, there was just a massive photon, (4.9). At higher energies, there was a Cooper-pair field, (4.10). At even higher energies, where we can break apart Cooper pairs, there are electrons (4.12). In this subsection, we peel away one more layer of the onion: at even higher energies, those electrons are no longer paired up and constitute a metal.

[Polchinski, lecture 2 (I recommend these notes very strongly), and R. Shankar] Electrically conducting solids are a remarkable phenomenon. An arbitrarily small electric field \vec{E} leads to a nonzero current $\vec{j} = \sigma \vec{E}$. This means that there must be gapless modes with energies much less than the natural cutoff scale in the problem.

Scales involved: The Planck scale of solid state physics (made by the logic by which Planck made his quantum gravity energy scale, namely by making a quantity with dimensions of energy out of the available constants) is

$$E_0 = \frac{1}{2} \frac{e^4 m}{\hbar^2} = \frac{1}{2} \frac{e^2}{a_0} \sim 13\text{eV}$$

(where $m \equiv m_e$ is the electron mass and the factor of 2 is an abuse of outside information) which is the energy scale of *chemistry*. Chemistry is to solids as the melting of spacetime is to particle physics. As with particle physics, however, there are other scales involved. In particular a solid involves a lattice of nuclei, each with $M \gg m$ (approximately the proton mass). So m/M is a useful small parameter which controls the coupling between the electrons and the lattice vibrations. Also, the actual speed of light $c \gg v_F$ can generally also be treated as ∞ to first approximation. v_F/c suppresses spin orbit couplings (though large atomic numbers enhance them: $\lambda_{\text{SO}} \propto Z v_F/c$).

Let us attempt to construct a Wilsonian-natural effective field theory of this phenomenon. The answer is called Landau Fermi Liquid Theory. What are the right low-energy degrees of freedom? Let's make a guess that they are like electrons – fermions

with spin and electric charge. They will not have exactly the properties of free electrons, since they must incorporate the effects of interactions with all their friends. The ‘dressed’ electrons are called quasielectrons, or more generally quasiparticles.

Given the strong interactions between so many particles, why should the dofs have anything at all to do with electrons? Landau’s motivation for this description (which is not always correct) is that we can imagine starting from the free theory and adiabatically turning up the interactions. If we don’t encounter any phase transition along the way, we can follow each state of the free theory, and use the same labels in the interacting theory.

We will show that there is a nearly-RG-stable fixed point describing gapless quasielectrons. Notice that we are not trying to match this description directly to some microscopic lattice model of a solid; rather we will do bottom-up effective field theory.

Having guessed the necessary dofs, let’s try to write an action for them consistent with the symmetries. A good starting point is the free theory:

$$S_{\text{free}}[\psi] = \int dt \, d^d p \, (\mathbf{i}\psi_\sigma^\dagger(p) \partial_t \psi_\sigma(p) - (\epsilon(p) - \epsilon_F) \psi_\sigma^\dagger(p) \psi_\sigma(p))$$

where σ is a spin index, ϵ_F is the Fermi energy (zero-temperature chemical potential), and $\epsilon(p)$ is the single-particle dispersion relation. For non-interacting non-relativistic electrons in free space, we have $\epsilon(p) = \frac{p^2}{2m}$. It will be useful to leave this as a general function of p .^{47 48}

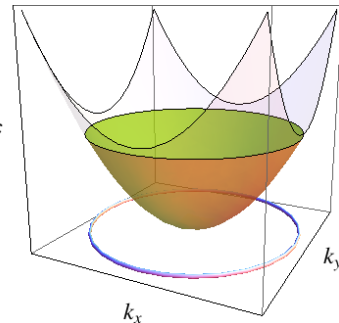
The groundstate of the free theory is the filled Fermi sea:

$$|\text{gs}\rangle = \prod_{p|\epsilon(p) < \epsilon_F} \psi_p^\dagger |0\rangle, \quad \psi_p |0\rangle = 0, \quad \forall p.$$

(If you don’t like continuous products, put the system in a box so that p is a discrete label.) The Fermi surface is the set of points in momentum space at the boundary of the filled states:

$$\text{FS} \equiv \{p | \epsilon(p) = \epsilon_F\}.$$

The low-lying excitations are made by adding an electron just above the FS or removing an electron (creating a hole) just below.



⁴⁷Notice that we are assuming translation invariance. I am not saying anything at the moment about whether translation invariance is discrete (the ions make a periodic potential) or continuous.

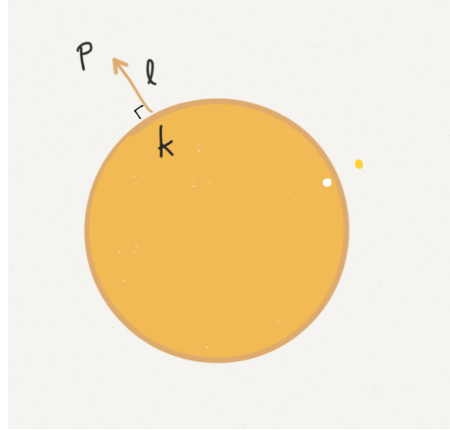
⁴⁸We have chosen the normalization of ψ to fix the coefficient of the ∂_t term (this rescaling may depend on p).

In order to define the power-counting rules for our EFT, we would like to define a scaling transformation which focuses on the low-energy excitations. We scale energies by a factor $E \rightarrow bE, b < 1$. In relativistic QFT, \vec{p} scales like E , toward zero, $\vec{p} \rightarrow b\vec{p}$, since all the low-energy stuff is near the single special point $\vec{p} = 0$. Here the situation is much more interesting because there is a whole surface of low-energy stuff on the FS. This will lead to what's called *hyperscaling violation* – we can't just count powers of momentum.

One way to implement this is to introduce a hierarchical labeling of points in momentum space, by breaking the momentum space into *patches* around the FS. (An analogous strategy of labeling is also used in heavy quark EFT and in SCET.)

We'll use a slightly different strategy, following Polchinski. To specify a point \vec{p} , we pick the nearest point \vec{k} on the FS, $\epsilon(\vec{k}) = \epsilon_F$ (draw a line perpendicular to the FS from \vec{p}), and let

$$\vec{p} = \vec{k} + \vec{\ell}.$$



So $d - 1$ of the components are determined by \vec{k} and one is determined by ℓ . (There are some exceptional cases if the FS gets too wiggly. Ignore these for now.)

$$\epsilon(p) - \epsilon_F = \ell v_F(\vec{k}) + \mathcal{O}(\ell^2), \quad v_F \equiv \partial_p \epsilon|_{p=k}.$$

So a scaling rule which accomplishes our goal of focusing on the FS is

$$E \rightarrow bE, \quad \vec{k} \rightarrow \vec{k}, \quad \vec{\ell} \rightarrow b\vec{\ell}.$$

This implies

$$dt \rightarrow b^{-1}dt, \quad d^{d-1}\vec{k} \rightarrow d^{d-1}\vec{k}, \quad d\vec{\ell} \rightarrow bd\vec{\ell}, \quad \partial_t \rightarrow b\partial_t$$

$$S_{\text{free}} = \int \underbrace{dt d^{d-1}\vec{k} d\vec{\ell}}_{\sim b^0} \left(\mathbf{i}\psi^\dagger(p) \underbrace{\partial_t}_{\sim b^1} \psi(p) - \underbrace{\ell v_F(k)}_{\sim b^1} \psi^\dagger(p)\psi(p) \right)$$

In order to make this go like b^0 we require $\psi \rightarrow b^{-\frac{1}{2}}\psi$ near the free fixed point.

Next we will play the EFT game. To do so we must enumerate the symmetries we demand of our EFT:

1. Particle number, $\psi \rightarrow e^{i\theta}\psi$

2. Spatial symmetries: time-translation invariance, and either (a) continuous translation invariance and rotation invariance (as for *e.g.* liquid ${}^3\text{He}$) or (b) lattice symmetries. This means that momentum space is periodically identified, roughly $p \simeq p + 2\pi/a$ where a is the lattice spacing (the set of independent momenta is called the Brillouin zone (BZ)) and p is only conserved modulo an inverse lattice vector $2\pi/a$. There can also be some remnant of rotation invariance preserved by the lattice. Case (b) reduces to case (a) if the Fermi surface does not go near the edges of the BZ.
3. Spin rotation symmetry, $\text{SU}(n)$ if $\sigma = 1..n$. In the limit with $c \rightarrow \infty$, this is an internal symmetry, independent of rotations.
4. Let's assume that $\epsilon(p) = \epsilon(-p)$, which is a consequence of *e.g.* parity invariance.

Now we enumerate all terms analytic in ψ and its momenta (since we are assuming that there are no other low-energy dofs integrating out which is the only way to get non-analytic terms in ψ) and consistent with the symmetries; we can order them by the number of fermion operators involved. Particle number symmetry means every ψ comes with a ψ^\dagger . The possible **quadratic terms** are:

$$\int dt \underbrace{d^{d-1}\vec{k} d\vec{\ell}}_{\sim b^0} \mu(k) \underbrace{\psi_\sigma^\dagger(p)\psi_\sigma(p)}_{\sim b^{-1}} \sim b^{-1}$$

is relevant. This is like a mass term. But don't panic: it just shifts the FS around. The *existence* of a Fermi surface is Wilson-natural (*i.e.* a stable assumption given generic coefficients of all possible terms in the action); any precise location or shape (modulo something enforced by symmetries, like roundness) is not.

Adding one extra ∂_t or factor of ℓ costs a b^1 and makes the operator marginal; those terms are already present in S_{free} . Adding more than one makes it irrelevant.

Quartic terms:

$$S_4 = \int dt \underbrace{\prod_{i=1}^4 d^{d-1}\vec{k}_i d\vec{\ell}_i}_{\sim b^{-1+4}} u(4 \cdot \dots \cdot 1) \underbrace{\psi_\sigma^\dagger(p_1)\psi_\sigma(p_3)\psi_{\sigma'}^\dagger(p_2)\psi_{\sigma'}(p_4)}_{\sim b^{-\frac{1}{2} \cdot 4}} \delta^d(\vec{p}_1 + \vec{p}_2 - \vec{p}_3 - \vec{p}_4)$$

The minus signs on $p_{3,4}$ is because $\psi(p)$ *removes* a particle with momentum p . We assume u depends only on k, σ , so does not scale – this will give the most relevant piece. How does the delta function scale?

$$\delta^d(\vec{p}_1 + \vec{p}_2 - \vec{p}_3 - \vec{p}_4) = \delta^d(k_1 + k_2 - k_3 - k_4 + \ell_1 + \ell_2 - \ell_3 - \ell_4) \stackrel{?}{\simeq} \delta^d(k_1 + k_2 - k_3 - k_4)$$

In the last (questioned) step, we used the fact that $\ell \ll k$ to ignore the contributions of the ℓ s. If this is correct then the delta function does not scale (since k s do not), and $S_4 \sim b^1$ is irrelevant (and quartic interactions with derivatives are moreso). If this were correct, the free-fixed point would be exactly stable.

There are two important subtleties: (1) the questioned equality above is questionable because of kinematics of the Fermi surface. (2) there exist phonons. We will address these two issues in order.



The **kinematic subtlety** in the treatment of the scaling of $\delta(p_1 + p_2 - p_3 - p_4)$ arises because of the geometry of the Fermi surface. Consider scattering between two points on the FS, where (in the labeling convention above)

$$p_3 = p_1 + \delta k_1 + \delta l_1, \quad p_4 = p_2 + \delta k_2 + \delta l_2,$$

in which case the momentum delta function is

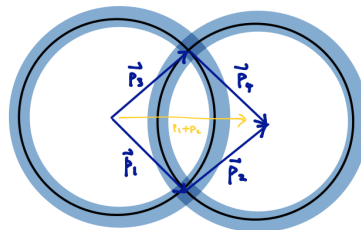
$$\delta^d(p_1 + p_2 - p_3 - p_4) = \delta^d(\delta k_1 + \delta l_1 + \delta k_2 + \delta l_2).$$



For generic choices of the two points $p_{1,2}$ (top figure at left), δk_1 and δk_2 are linearly independent and the δl s can indeed be ignored as we did above. However, for

two points with $p_1 = -p_2$ (they are called *nested*, as depicted in the bottom figure at left), then one component of $\delta k_1 + \delta k_2$ is automatically zero, revealing the tiny δl s to the force of (one component of) the delta function. In this case, $\delta(\ell)$ scales like b^{-1} , and for this particular kinematic configuration the four-fermion interaction is (classically) marginal. Classically marginal means quantum mechanics has a chance to make a big difference.

A useful visualization is at right ($d = 2$ with a round FS is shown; this is what's depicted on the cover of the famous book by Abrikosov-Gorkov-Dzyaloshinski): the blue circles have radius k_F ; the yellow vector is the sum of the two initial momenta $p_1 + p_2$, both of which are on the FS; the condition that $p_3 + p_4$, each also on the FS, add up to the same vector means that p_3 must lie on the intersection of the two circles (spheres in $d > 2$). But when $p_1 + p_2 = 0$, the two circles are on top of each other so they intersect everywhere! Comments:



1. We assumed that both p_1 and $-p_2$ were actually *on* the FS. This is automatic if $\epsilon(p) = \epsilon(-p)$, *i.e.* if ϵ is only a function of p^2 .

2. This discussion works for any $d > 1$.

3. **Forward scattering.** There is a similar phenomenon for the case where $p_1 = p_3$ (and hence $p_2 = p_4$). This is called *forward scattering* because the final momenta are the same as the initial momenta. (We could just as well take $p_1 = p_4$ (and hence $p_2 = p_3$.) In this case too the delta function will constrain the ℓ s and will therefore scale.

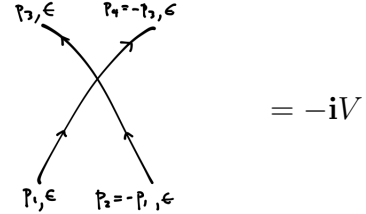
The tree-level-marginal 4-Fermi interactions at special kinematics leads to a *family* of fixed points labelled by ‘Landau parameters’. In fact there is whole *function’s* worth of fixed points. In 2d, the points on the FS are parametrized by an angle θ , and the fixed point manifold is parametrized by the forward-scattering function

$$F(\theta_1, \theta_2) \equiv u(\theta_4 = \theta_2, \theta_3 = \theta_1, \theta_2, \theta_1)$$

(Fermi statistics implies that $u(\theta_4 = \theta_1, \theta_3 = \theta_2, \theta_2, \theta_1) = -F(\theta_1, \theta_2)$) and the BCS-channel (nesting) interaction:

$$V(\theta_1, \theta_3) = u(\theta_4 = -\theta_3, \theta_3, \theta_2 = -\theta_1, \theta_1).$$

Now let’s think about what decision the fluctuations make about the fate of the nested interactions. The most interesting bit is the renormalization of the BCS interaction:



The electron propagator, obtained by inverting the kinetic operator in S_{free} , is

$$G(\epsilon, p = k + l) = \frac{\mathbf{i}}{\epsilon(1 + \mathbf{i}\eta) - v_F(k)\ell + \mathcal{O}(\ell)^2}$$

where I used $\eta \equiv 0^+$ for the infinitesimal specifying the contour prescription.

Let’s assume rotation invariance. Then $V(\theta_3, \theta_1) = V(\theta_3 - \theta_1)$, $V_l = \int d\theta e^{i\theta} V(\theta)$. Different angular momentum sectors decouple from each other at one loop.

We will focus on the *s*-wave bit of the interaction, so V is independent of momentum. We will integrate out just a shell in energy (depicted by the blue shaded shell in the Fermi surface figures). The interesting contribution comes from the following diagram:

$$-\mathbf{i}\delta^{(1)}V = \begin{array}{c} \begin{array}{c} p_3, \epsilon \quad p_4 = -p_3, \epsilon \\ \swarrow \quad \searrow \\ p_1, \epsilon \quad p_2 = -p_1, \epsilon \end{array} \\ \begin{array}{c} p', \epsilon + \epsilon' \quad -p', \epsilon - \epsilon' \\ \swarrow \quad \searrow \\ p_1, \epsilon \quad p_2 = -p_1, \epsilon \end{array} \end{array} = -n(-\mathbf{i}V)^2 \int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon' d^{d-1}k' d\ell'}{(2\pi)^{d+1}} \frac{\mathbf{i}^2}{(\epsilon + \epsilon' - v_F(k')\ell')(\epsilon - \epsilon' - v_F(k')\ell')}$$

$$\begin{aligned}
\text{do } \int d\ell' \text{ by residues} &= -nV^2 \int \frac{d\epsilon' d^{d-1}k'}{(2\pi)^{d+1}} \frac{2\pi\mathbf{i}}{v_F(k')} \left(\underbrace{\epsilon - \epsilon' - (\epsilon + \epsilon')}_{=-2\epsilon'} \right)^{-1} \\
&= +\mathbf{i} \frac{nV^2}{2} \underbrace{\int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon'}{\epsilon'}}_{=\log(1/b)} \underbrace{\int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}}_{\text{dos at FS}}.
\end{aligned} \tag{4.22}$$

Don't forget the fermion loop minus sign. Between the first and second lines, we did the ℓ' integral by residues. The crucial point is that we are interested in external energies $\epsilon \sim 0$, but we are integrating out a shell near the cutoff, so $|\epsilon'| > |\epsilon|$ and the sign of $\epsilon + \epsilon'$ is opposite that of $\epsilon - \epsilon'$; therefore there is a pole on either side of the real ℓ axis and we get the same answer by closing the contour either way. On one side the pole is at $\ell' = \frac{1}{v_F(k')} (\epsilon + \epsilon')$. (In the t-channel diagram (what Shankar calls ZS), the poles are on the *same* side and it therefore does not renormalize the four-fermion interaction.)

The result to one-loop is then

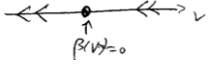
$$V(b) = V - V^2 N \log(1/b) + \mathcal{O}(V^3)$$

with $N \equiv \frac{n}{2} \int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}$ is the density of states at the Fermi surface (including the spin multiplicity). From this we derive the beta function (recall that $b \rightarrow 0$ in the IR in this section)

$$-b \frac{d}{db} V(b) = \beta_V = -NV^2(b) + \mathcal{O}(V^3)$$

and the solution of the flow equation at $E = bE_1$ is

$$V(E) = \frac{V_1}{1 + NV_1 \log(E_1/E)} \begin{cases} \rightarrow 0 & \text{in IR for } V_1 > 0 \text{ (repulsive)} \\ \rightarrow -\infty & \text{in IR for } V_1 < 0 \text{ (attractive)} \end{cases} \tag{4.23}$$

There is therefore a very significant dichotomy depending on the sign of the coupling at the microscopic scale E_1 , as in this phase diagram: 

The conclusion is that if the interaction starts attractive at some scale it flows to large attractive values. The thing that is decided by our perturbative analysis is that (if $V(E_1) < 0$) the decoupling we did with σ ('the BCS channel') wins over the decoupling with ρ ('the particle-hole channel'). What happens at $V \rightarrow -\infty$? Here we need non-perturbative physics.

The non-perturbative physics is in general hard, but we've already done what we can in §4.5.1.

The remaining question is: Who is V_1 and why would it be attractive (given that Coulomb interactions between electrons, while screened and therefore short-ranged, are repulsive)? The answer is:

Phonons. The lattice of positions taken by the ions making up a crystalline solid spontaneously break many spacetime symmetries of their governing Hamiltonian. This implies a collection of gapless Goldstone modes in any low-energy effective theory of such a solid⁴⁹. The Goldstone theorem is satisfied by including a field

$$\vec{D}(\vec{r}) \propto (\text{local}) \text{ displacement } \delta\vec{r} \text{ of ions near } \vec{r} \text{ from their equilibrium positions}$$

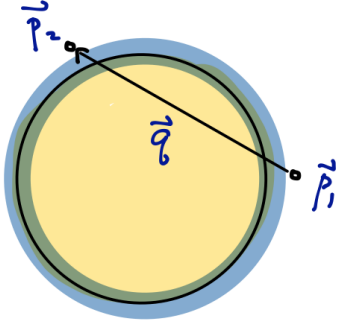
Most microscopically we have a bunch of coupled springs:

$$L_{\text{ions}} \sim \frac{1}{2}M \left(\dot{\delta\vec{r}} \right)^2 - k_{ij} \delta r^i \delta r^j + \dots$$

with spring constants k independent of the nuclear mass M . It is useful to introduce a canonically normalized field in terms of which the action is

$$S[\vec{D}] = (M)^{1/2} \int dt d^d q \left(\partial_t D_i(q) \partial_t D_i(-q) - \omega_{ij}^2(q) D_i(q) D_j(-q) \right).$$

Here $\omega^2 \propto M^{-1}$. Their status as Goldstones means that the eigenvalues of $\omega_{ij}^2(q) \sim |q|^2$ at small q : moving everyone by the same amount does not change the energy. This also constrains the coupling of these modes to the electrons: they can only couple through derivative interactions.



For purposes of their interactions with the electrons, a nonzero q which keeps the e^- on the FS must scale like $q \sim b^0$. Therefore

$$dt d^d q (\partial_t D)^2 \sim b^{+1+2[D]} \implies D \sim b^{-\frac{1}{2}}$$

and the restoring force $dt d^d q D^2 \omega^2(q) \sim b^{-2}$ is relevant, and dominates over the ∂_t^2 term for

$$E < E_D = \sqrt{\frac{m}{M}} E_0 \quad \text{the Debye energy.}$$

(For the more traditional derivation of the relation between E_D and E_0 , see *e.g.* DeGennes' *Superconductivity of Metals and Alloys*, pages 99-102.) This means that phonons mediate *static* interactions below E_D – we can ignore retardation effects, and their effects on the electrons can be fully incorporated by the four-fermion interaction we used above (with some \vec{k} dependence). How do they couple to the electrons?

$$S_{\text{int}}[D, \psi] = \int dt d^3 q d^2 k_1 d\ell_1 d^2 k_2 d\ell_2 M^{-\frac{1}{2}} g_i(q, k_1, k_2) D_i(q) \psi_\sigma^\dagger(p_1) \psi_\sigma(p_2) \delta^3(p_1 - p_2 - q)$$

⁴⁹Note that there is a subtlety in counting Goldstone modes from spontaneously broken spacetime symmetries: there are more symmetry generators than Goldstones. Basically it's because the associated currents differ only by functions of spacetime; but a localized Goldstone particle is anyway made by a current times a function of spacetime, so you can't sharply distinguish the resulting particles. Some useful references on this subject are [Low-Manohar](#) and most recently [Watanabe-Murayama](#).

$$\sim b^{-1+1+1-3/2} = b^{-1/2} \tag{4.24}$$

– here we took the delta function to scale like b^0 as above. This is relevant when we use the \dot{D}^2 scaling for the phonons; when the restoring force dominates we should scale D differently and this is irrelevant for generic kinematics. This is consistent with our previous analysis of the four-fermion interaction.

The summary of this discussion is: phonons do not destroy the Fermi surface, but they *do* produce an attractive contribution to the 4-fermion interaction, which is relevant in some range of scales (above the Debye energy). Below the Debye energy, it

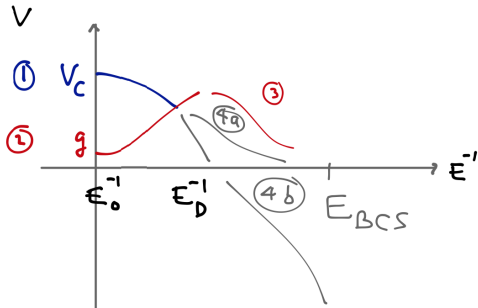
amounts to an addition to V that goes like $-g^2$: 

Notice that the scale at which the coupling V becomes strong ($V(E_{\text{BCS}}) \equiv 1$ in (4.23)) is

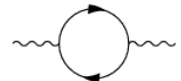
$$E_{\text{BCS}} \sim E_D e^{-\frac{1}{N V_D}}.$$

Two comments about this: First, it is non-perturbative in the interaction V_D . Second, it provides some verification of the role of phonons, since $E_D \sim M^{-1/2}$ can be varied by studying the same material with different isotopes and studying how the critical superconducting temperature ($\sim E_{\text{BCS}}$) scales with the nuclear mass.

Actually, we can make some headway towards understanding the result of this interaction going strong. Because the diagrams with the special kinematics are marginal and hence unsuppressed, while all other interactions flow to zero at low energy, certain diagrams dominate. In particular, bubble-chains dominate.



Here’s the narrative, proceeding as a function of decreasing energy scale, beginning at E_0 , the Planck scale of solids: (1) Electrons repel each other by the Coulomb interaction. However, in a metal, this interaction is

screened by processes like this: 

(the intermediate state is an electron-hole pair) and is short-ranged. It is still repulsive,

however. As we coarse-grain more and more, we see more and more electron-hole pairs and the force weakens. (2) While this is happening, the electron-phonon interaction is relevant and growing. This adds an attractive bit to V . This lasts until E_D . (3) At E_D the restoring force term in the phonon lagrangian dominates (for the purposes of their interactions with the electrons) and we can integrate them out. (4) What happens next depends on the sign of $V(E_D)$. If it’s positive, V flows harmlessly to zero. If

it's negative, it becomes moreso until we exit the perturbative analysis at E_{BCS} , and vindicate our choice of Hubbard-Stratonovich channel above.

Further brief comments, for which I refer you to Shankar:

1. Putting back the possible angular dependence of the BCS interaction, the result at one loop is

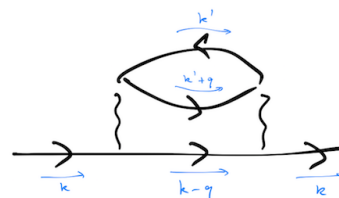
$$\frac{dV(\theta_1 - \theta_3)}{d\ell} = -\frac{1}{8\pi^2} \int_0^{2\pi} d\theta V(\theta_1 - \theta)V(\theta - \theta_3)$$

or in terms of angular momentum components,

$$\frac{dV_l}{d\ell} = -\frac{V_l^2}{4\pi}.$$

2. This example is interesting and novel in that it is a (family of) fixed point(s) characterized by a dimensionful quantity, namely k_F . This leads to a phenomenon called *hyperscaling violation* where thermodynamic quantities need not have their naive scaling with temperature.
3. The one loop analysis gives the right answer to all loops in the limit that $N \sim (k_F/\Lambda)^{d-1} \gg 1$, where Λ is the UV cutoff on the momentum.
4. The forward scattering interaction (for any choice of function $F(\theta_{13})$) is not renormalized at one loop. This means it is exactly marginal at leading order in N .
5. Like in ϕ^4 theory, the sunrise diagram at two loops is the first appearance of wavefunction renormalization. In the context of the Fermi liquid theory, this leads to the renormalization of the effective mass which is called m^* .

Another consequence of the FS kinematics which I should emphasize more: it allows the quasiparticle to be stable. The leading contribution to the decay rate of a one-quasiparticle state with momentum k can be obtained applying the optical theorem to the following process.



In the figure, the object  is the four-fermion vertex (the wiggly line is just

for clarity). The intermediate state is two electrons with momenta $k' + q$ and $k - q$, and one *hole* with momentum k' . The hole propagator has the opposite $i\eta$ prescription. (To understand the contour prescription for the *hole* propagator, it is useful to begin with

$$G(t, p) = \langle \text{gs} | \mathcal{T} c_p^\dagger(t) c_p(0) | \text{gs} \rangle, \quad c_p^\dagger(t) \equiv e^{-i\mathbf{H}t} c_p^\dagger e^{i\mathbf{H}t}$$

and use the free-fermion fact $[\mathbf{H}, c_p^\dagger] = \epsilon_p c_p^\dagger$. For more details, see the steps leading up to equation (7.7) of AGD (Abrikosov, Gorkov, Dzyaloshinski, *Methods of QFT in Statistical Physics*.)

After doing the frequency integrals by residues, we get

$$\Sigma(k, \epsilon) = \int \bar{d}q \bar{d}k' \frac{|u_q|^2}{D - \mathbf{i}\eta}$$

$$D \equiv \epsilon_k(1 + \mathbf{i}\eta) + \epsilon_{k'}(1 - \mathbf{i}\eta) - \epsilon_{k'+q}(1 + \mathbf{i}\eta) - \epsilon_{k-q}(1 + \mathbf{i}\eta)$$

(Notice that this is the eyeball diagram which gives the lowest-order contribution to the wavefunction renormalization of a field with quartic interactions.) By the optical theorem, its imaginary part is the (leading contribution to the) inverse-lifetime of the quasiparticle state with fixed k :

$$\tau^{-1}(k) = \text{Im} \Sigma(k, \epsilon) = \pi \int \bar{d}q \bar{d}k' \delta(D) |u_q|^2 f(-\epsilon_{k'}) f(\epsilon_{k'+q}) f(\epsilon_{k-q})$$

where

$$f(\epsilon) = \lim_{T \rightarrow 0} \frac{1}{e^{\frac{\epsilon - \epsilon_F}{T}} + 1} = \theta(\epsilon < \epsilon_F)$$

is the Fermi function. This is just the demand that a particle can only scatter into an empty state and a hole can only scatter into a filled state. These constraints imply that all the energies are near the Fermi energy: both $\epsilon_{k'+q}$ and $\epsilon_{k'}$ lie in a shell of radius ϵ about the FS; the answer is proportional to the density of possible final states, which is thus

$$\tau^{-1} \propto \left(\frac{\epsilon}{\epsilon_F} \right)^2.$$

So the width of the quasiparticle resonance is

$$\tau^{-1} \propto \epsilon^2 \ll \epsilon$$

much smaller than its frequency – it is a sharp resonance, a well-defined particle.

[End of Lecture 19]

4.7 The Standard Model as an EFT.

The Standard Model. [Schwartz, §29]

	$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	e_R	ν_R	$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	u_R	d_R	H
SU(3)	-	-	-	□	□	□	-
SU(2)	□	-	-	□	-	-	□
U(1) _Y	$-\frac{1}{2}$	-1	0	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

Table 1: The Standard Model fields and their quantum numbers under the gauge group. □ indicates fundamental representation, - indicates singlet. Except for the Higgs, each column is copied three times; each copy is called a *generation*. Except for the Higgs all the matter fields are Weyl fermions of the indicated handedness. Gauge fields as implied by the gauge groups. (Some people might leave out the right-handed neutrino, ν_R .)

Whence the values of the charges under the U(1) (“hypercharge”)? The condition $Y_L + 3Y_Q = 0$ (where Y is the hypercharge) is required by anomaly cancellation. This implies that electrons and protons $p = \epsilon_{ijk}u_iu_jd_k$ have exactly opposite charges of the same magnitude.

The Lagrangian is just all the terms which are invariant under the gauge group $SU(3) \times SU(2) \times U(1)$ with dimension less than or equal to four – all renormalizable terms. This includes a potential for the Higgs, $V(|H|) = m_H^2|H|^2 + \lambda|H|^4$, where it turns out that $m_H^2 \leq 0$. The resulting Higgs vacuum expectation value breaks the Electroweak part of the gauge group

$$SU(2) \times U(1)_Y \xrightarrow{\langle H \rangle} U(1)_{EM}.$$

The broken gauge bosons get masses from the Higgs kinetic term

$$|D_\mu H|^2 \Big|_{H=\begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}} \text{ with } D_\mu H = \left(\partial_\mu - \mathbf{i}gW_\mu^a\tau^a - \frac{1}{2}\mathbf{i}g'Y_\mu \right) H$$

where Y_μ is the hypercharge gauge boson, and $W^a, a = 1, 2, 3$ are the SU(2) gauge bosons. There are two massive W -bosons with electric charge ± 1 (as described in §4.3), with $M_W = \frac{vg}{2}$. The photon and Z boson are the linear combinations of Y and W^3 which diagonalize the remaining mass terms:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_w & \sin \theta_w \\ -\sin \theta_w & \cos \theta_w \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ Y_\mu \end{pmatrix}.$$

Here $\tan\theta_w \equiv \frac{g'}{g}$ defines the Weinberg angle. The masses are $M_\gamma = 0$ and $M_Z = \frac{M_W}{\cos\theta_w} < M_W$.

Fermion masses come from (dimension-four) Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^\ell \bar{L}_i H e_R^j - Y_{ij}^u \bar{Q}^i H d_R^j - Y_{ij}^d \bar{Q}^i (\mathbf{i}\tau^2 H^*) u_R^j + h.c.$$

The contortion with the τ^2 is required to make a hypercharge invariant. Plugging in the Higgs vev to *e.g.* the lepton terms gives $-m_e \bar{e}_L e_R + h.c.$ with $m_e = y_e v / \sqrt{2}$. There's lots of drama about the matrices Y which can mix the generations. The mass for the ν_R (which maybe could not exist – it doesn't have any charges at all) you can figure out on the homework later.

Here is a useful mnemonic for remembering the table of quantum numbers (possibly it is more than that): There are larger simple Lie groups that contain the SM gauge group as subgroups:

$$\begin{array}{ccccc} \text{SU}(3) \times \text{SU}(2) \times \text{U}(1)_Y & \subset & \text{SU}(5) & \subset & \text{SO}(10) \\ \text{one generation} & = & 10 \oplus \bar{5} \oplus 1 & = & 16 \end{array}$$

The singlet of $\text{SU}(5)$ is the right-handed neutrino, but if we include it, one generation is an irreducible (spinor) representation of $\text{SO}(10)$. This idea is called grand unification. It is easy to imagine that the gauge group is actually the larger groups on the right, and another instance of the Higgs mechanism accomplishes the breaking down to the Standard Model. (The running of the respective gauge couplings go in the right direction with approximately the right rate to unify to a single value at $M_{GUT} \sim 10^{16} \text{GeV}$.) Notice that this idea means leptons and quarks are in the same representations – they can turn into each other. This predicts that the proton should not be perfectly stable. Next we'll say more about this.

Beyond the Standard Model with EFT. At what energy does the Standard Model stop working? Because of the annoying feature of renormalizability, it doesn't tell us. However, we have experimental evidence against a cutoff on the Standard Model (SM) at energies less than something like 10 TeV. The evidence I have in mind is the absence of interactions of the form

$$\delta L = \frac{1}{M^2} (\bar{\psi} A \psi) \cdot (\bar{\psi} B \psi)$$

(where ψ represent various SM fermion fields and A, B can be various gamma and flavor matrices) with $M \lesssim 10 \text{ TeV}$. Notice that I am talking now about interactions *other* than the electroweak interactions, which as we've just discussed, for energies above $M_W \sim 80 \text{ GeV}$ cannot be treated as contact interactions – you can see the W s propagate!

If such operators were present, we would have found different answers for experiments at LEP. But such operators would be present if we consider new physics in addition to the Standard Model (in most ways of doing it) at energies less than 10 TeV. For example, many interesting ways of coupling in new particles with masses that make them accessible at the LHC would have generated such operators.

A little more explicitly: the Standard Model Lagrangian L_0 contains all the renormalizable (*i.e.* engineering dimension ≤ 4) operators that you can make from its fields (though the coefficients of the dimension 4 operators do vary through quite a large range, and the coefficients of the two relevant operators – namely the identity operator which has dimension zero, and the Higgs mass, which has engineering dimension two, are strangely small, and so is the QCD θ angle).

To understand what lies beyond the Standard Model, we can use our knowledge that whatever it is, it is probably heavy (it could also just be very weakly coupled, which is a different story), with some intrinsic scale Λ_{new} , so we can integrate it out and include its effects by corrections to the Standard Model:

$$L = L_0 + \frac{1}{\Lambda_{\text{new}}} \mathcal{O}^{(5)} + \frac{1}{\Lambda_{\text{new}}^2} \sum_i c_i \mathcal{O}_i^{(6)} + \dots$$

where the \mathcal{O} s are made of SM fields, and have the indicated engineering dimensions, and preserve the necessary symmetries of the SM (Lorentz symmetry and gauge invariance).

In fact there is only one kind of operator of dimension 5 meeting these demands:

$$\mathcal{O}^{(5)} = c_5 \epsilon_{ij} (\bar{L}^c)^i H^j \epsilon_{kl} L^k H^l$$

where $H^i = (h^+, h^0)^i$ is the $SU(2)_{EW}$ Higgs doublet and $L^i = (\nu_L, e_L)^i$ is an $SU(2)_{EW}$ doublet of left-handed leptons, and $\bar{L}^c \equiv L^T C$ where C is the charge conjugation matrix. (I say ‘kind of operator’ because we can have various flavor matrices in here.) On the problem set you get to see from whence such an operator might arise, and what it does if you plug in the higgs vev $\langle H \rangle = (0, v)$. This term violates lepton number symmetry ($L \rightarrow e^{i\alpha_L} L, Q \rightarrow Q, H \rightarrow H$).

At dimension 6, there are operators that directly violate baryon number, such as

$$\epsilon_{\alpha\beta\gamma} (\bar{u}_R)_\alpha^c (u_R)_\beta (\bar{u}_R)_\gamma^c e_R.$$

You should read the above tangle of symbols as ‘ $qqq\ell$ ’ – it turns three quarks into a lepton. The epsilon tensor makes a color $SU(3)$ singlet; this thing $\epsilon qq\ell$ has the quantum numbers of a baryon, such as the proton and neutron. The long lifetime of the proton (you can feel it in your bones – see Zee p. 413) then directly constrains the scale of new physics appearing in front of this operator.

Two more comments about this:

- If we didn't know about the Standard Model, (but after we knew about QM and GR and EFT (the last of which people didn't know before the SM for some reason)) we should have made the estimate that dimension-5 Planck-scale-suppressed operators like $\frac{1}{M_{\text{Planck}}}p\mathcal{O}$ would cause proton decay (into whatever \mathcal{O} makes). This predicts $\Gamma_p \sim \frac{m_p^3}{M_{\text{Planck}}^2} \sim 10^{-13} s^{-1}$ which is *not* consistent with our bodies not glowing. Actually it is a remarkable fact that there are no gauge-invariant operators made of SM fields of dimension less than 6 that violate baryon number symmetry ($L \rightarrow L, Q \rightarrow e^{i\alpha_B}Q, H \rightarrow H$). This is an *emergent* symmetry, expected to be violated by the UV completion.
- Surely nothing can prevent $\Delta L \sim \left(\frac{1}{M_{\text{Planck}}}\right)^2 qqql$. Happily, this is consistent with the observed proton lifetime.

There are $\sim 10^2$ dimension 6 operators that preserve baryon number, and therefore are not as tightly constrained⁵⁰. (Those that induce flavor-changing processes in the SM are more highly constrained and must have $\Lambda_{\text{new}} > 10^4$ TeV.) Two such operators are considered equivalent if they differ by something which vanishes by the tree-level SM equations of motion. This is the right thing to do, even for off-shell calculations (like green's functions and for fields running in loops). You know this from a previous problem set: the EOM are true as operator equations – Ward identities resulting from being free to change integration variables in the path integral⁵¹.

⁵⁰Recently, humans have gotten better at counting these operators. See [this paper](#).

⁵¹There are a few meaningful subtleties here, as you might expect if you recall that the Ward identity is only true up to contact terms. The measure in the path integral can produce a Jacobian which renormalizes some of the couplings; the changes in source terms will drop out of S-matrix elements (recall our discussion of changing field variables in the Consequences of Unitarity section.) but can change the form of Green's functions. For more information on the use of eom to eliminate redundant operators in EFT, see [Arzt, hep-ph/9304230](#) and [Georgi, "On-Shell EFT"](#).

5 Gauge theory